ACE

Applied and Computational Engineering

Proceedings of the 2023 International Conference
on Machine Learning and Automation

Adana, Turkey

October 18 - 25, 2023

EWA Publishing

Volume 32

Editor
Mustafa İSTANBULLU
Cukurova University

ISSN: 2755-2721
ISSN: 2755-273X (eBook)

ISBN: 978-1-83558-289-3
ISBN: 978-1-83558-290-9 (eBook)

Publication of record for individual papers is online:
https://ace.ewapublishing.org/

This imprint is published by **EWA Publishing**
Address: John Eccles House, Robert Robinson Avenue, Oxford, England, OX4 4GP
Email: info@ewapublishing.org

Committee Members

CONF-MLA 2023

General Chair

Marwan Omar, *Illinois Institute of Technology*

Organizing Chair

Roman Bauer, *University of Surrey*

Organizing Committee

Anil Fernando, *University of Strathclyde*
Ali Darejeh, *UNSW Sydney*
Alan Wang, *University of Auckland*
Faruk Aktaş, *Kocaeli University*
Sharidya Rahman, *Monash University*
Selda Kapan Ulusoy, *Erciyes University*
Abdullahi Arabo, *University of the West of England*
Richa Gupta, *Jamia Hamdard*

Technical Program Chair

Mustafa İSTANBULLU, *Cukurova University*

Technical Program Committee

Festus Adedoyin, *Bournemouth University*
Rudrendu Kumar Paul, *Boston University*
Ali Kashif Bashir, *Manchester Metropolitan University*
Moayad Aloqaily, *Mohamed Bin Zayed University of Artificial Intelligence*
Turgay Batbat, *Erciyes University*
Xinqing Xiao, *China Agricultural University*
Wei Li, *Chinese Academy of Sciences*
Ying Xing, *Beijing University of Posts and Telecommunications*
Chenkai Guo, *Nankai University*
Ce Li, *China University of Mining and Technology, Beijing*
Runyu Chen, *University of International Business and Economics*
Mukhtar Ullah, *FAST NUCES Islamabad*

Publicity Committee

Çiğdem Gülüzar Altıntop, *Erciyes University*
Toqeer Mahmood, *National Textile University Faisalabad*
Altaf Khan, *University of Narowal*
Sajid Khan, *Grand Asian University, Sialkot*

Preface

The 2023 International Conference on Machine Learning and Automation (CONF-MLA 2023) is an annual conference focusing on research areas including engineering and machine learning applications. It aims to establish a broad and interdisciplinary platform for experts, researchers, and students worldwide to present, exchange, and discuss the latest advance and development in engineering and machine learning applications.

This volume contains the papers of the 2023 International Conference on Machine Learning and Automation (CONF-MLA 2023). Each of these papers has gained a comprehensive review by the editorial team and professional reviewers. Each paper has been examined and evaluated for its theme, structure, method, content, language, and format.

Cooperating with prestigious universities, CONF-MLA 2023 organized five workshops in Chicago, Auckland, Oxford, Glasgow and Adana. Dr. Marwan Omar chaired the workshop "The Role of Machine Learning in Advancing Cyber Security", which was held at Illinois Institute of Technology. Dr. Alan Wang chaired the workshop "Workshop on Machine Learning for Medicine 2023 (WMLM2023)" at The University of Auckland. Prof. Hong Zhu chaired the workshop "Automation of Testing Machine Learning" at Oxford Brookes University. Prof. Anil Fernando chaired the workshop "Autoencoder Based Semantic Communications for Image Transmission on Error Prone Channels" at University of Strathclyde. Dr. Mustafa İSTANBULLU chaired the workshop "Process Design of Semiconductor Device Fabrication: From Concept to Production" at Cukurova University. Besides these workshops, CONF-MLA 2023 also held an online session. Eminent professors from top universities worldwide were invited to deliver keynote speeches in this online session, including Dr. Abdullahi Arabo from University of the West of England, Dr. Roman Bauer from The University of Surrey, Dr. Ali Darejeh from University of New South Wales (UNSW), etc. They have given keynote speeches on related topics of engineering and machine learning applications.

On behalf of the committee, we would like to give sincere gratitude to all authors and speakers who have made their contributions to CONF-MLA 2023, editors and reviewers who have guaranteed the quality of papers with their expertise, and the committee members who have devoted themselves to the success of CONF-MLA 2023.

Dr. Marwan Omar

General Chair of Conference Committee

Workshop

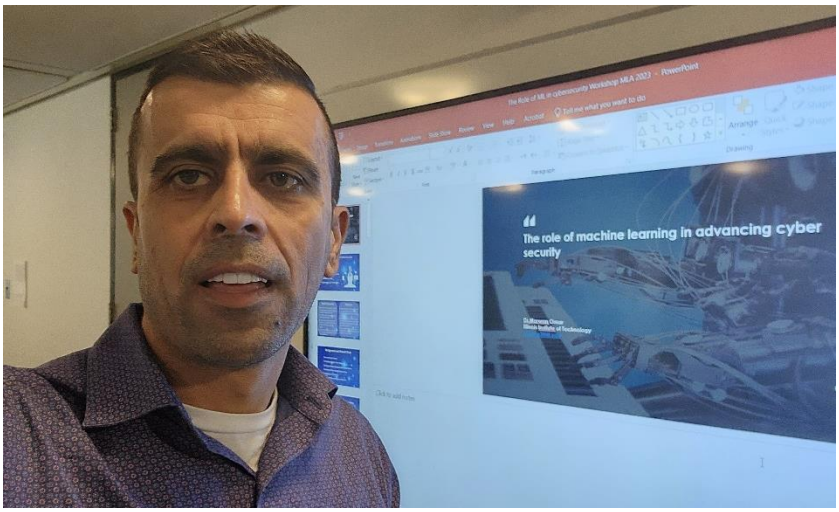
Workshop – Adana: Process Design of Semiconductor Device Fabrication: From Concept to Production

October 18th, 2023 (GMT+3)

Department of Biomedical Engineering, Faculty of Engineering, Cukurova University

Workshop Chair: Dr. Mustafa İSTANBULLU, Assistant Professor in Cukurova University

Workshop – Chicago: The Role of Machine Learning in Advancing Cyber Security



August 28th, 2023 (CDT)

ITM Department, Illinois Institute of Technology

Workshop Chair: Dr. Marwan Omar, Associate Professor in Illinois Institute of Technology

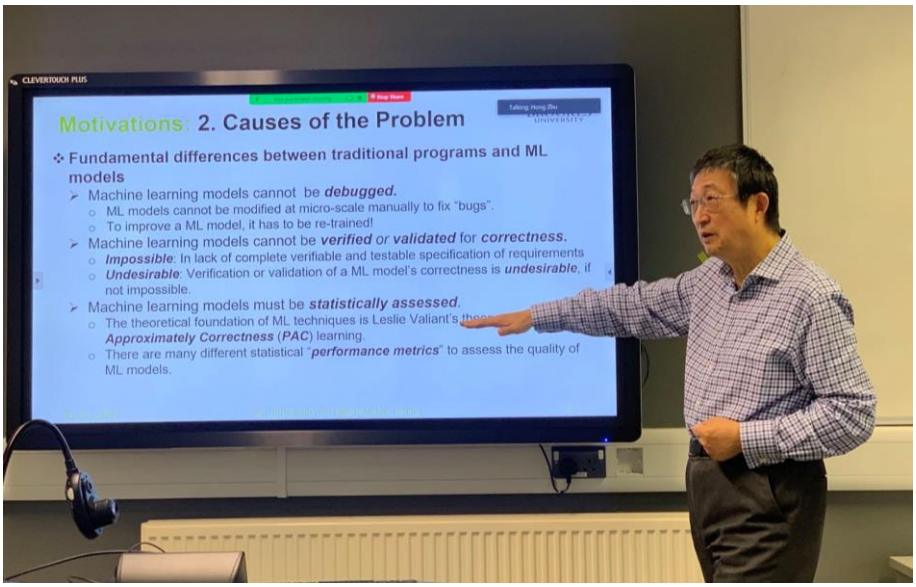
Workshop – Auckland: Workshop on Machine Learning for Medicine 2023 (WMLM2023)

October 13th, 2023 (GMT+13)

Faculty of Medical and Health Sciences and Bioengineering Institute, The University of Auckland

Workshop Chair: Dr. Alan Wang, Associate Professor in The University of Auckland

Workshop – Oxford: Automation of Testing Machine Learning

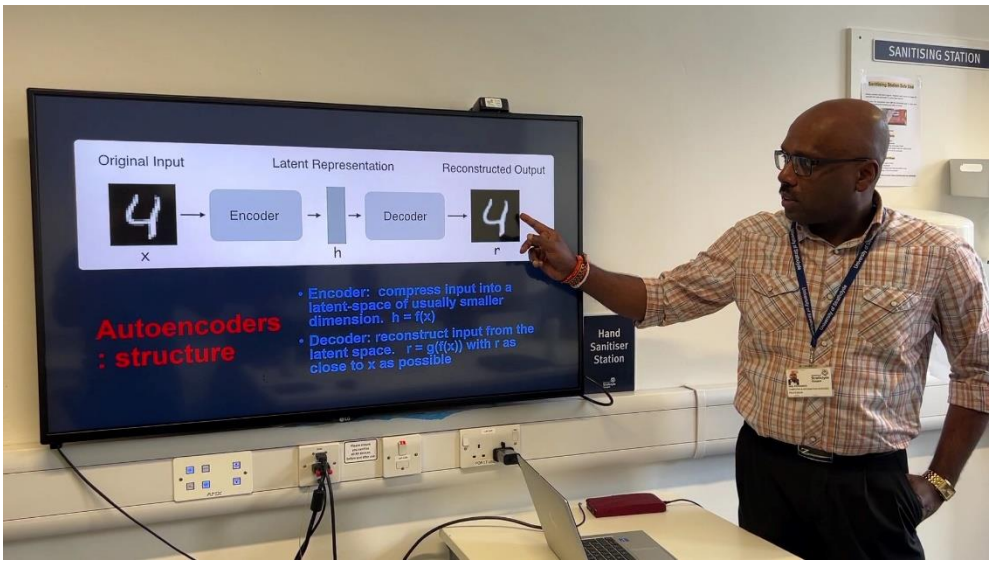


October 18th, 2023 (GMT+1)

School of Engineering, Computing and Mathematics, Oxford Brookes University

Workshop Chair: Prof. Hong Zhu, Professor of computer science in Oxford Brookes University

Workshop – Glasgow: Autoencoder Based Semantic Communications for Image Transmission on Error Prone Channels



May 2nd, 2023 (GMT+1)

Department of Computer and Information Sciences, University of Strathclyde

Workshop Chair: Prof. Anil Fernando, Professor in University of Strathclyde

The 2023 International Conference on Machine Learning and Automation

CONF-MLA 2023

Table of Contents

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230175

Comparative analysis of machine learning techniques for cryptocurrency price prediction

Siqi Yu

SWUFE-UD Institute of Data Science, Southwestern University of Finance, Chengdu, Sichuan, 611130, China.

yusiqi@udel.edu

Abstract. The emergence of cryptocurrencies has revolutionized the concept of digital currencies and attracted significant attention from financial markets. Predicting the price dynamics of cryptocurrencies is crucial but challenging due to their highly volatile and non-linear nature. This study compares the performance of various models in predicting cryptocurrency prices using three datasets: Bitcoin (BTC), Litecoin (LTC), and Ethereum (ETH). The models analyzed include Moving Average (MA), Logistic Regression (LR), Autoregressive Integrated Moving Average (ARIMA), Long Short-Term Memory (LSTM), and Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM). The objective is to uncover underlying patterns in cryptocurrency price movements and identify the most accurate and reliable approach for predicting future prices. Through the analysis, it could be observed that MA, LR, and ARIMA models struggle to capture the actual trend accurately. In contrast, LSTM and CNN-LSTM models demonstrate strong fit to the actual price trend, with CNN-LSTM exhibiting a higher level of granularity in its predictions. Results suggest that deep learning architectures, particularly CNN-LSTM, show promise in capturing the complex dynamics of cryptocurrency prices. These findings contribute to the development of improved methodologies for cryptocurrency price prediction.

Keywords: cryptocurrency, machine learning, ARIMA, neural network.

1. Introduction

The emergence of Bitcoin and other cryptocurrencies has revolutionized the concept of digital currency by introducing decentralized systems that operate without the need for a central authority. These cryptocurrencies rely on a peer-to-peer network and utilize blockchain technology to record and verify transactions. Among them, Bitcoin holds the largest market capitalization, followed by various altcoins such as Ripple, Litecoin, and Dash [1].

The price dynamics of Bitcoin and other cryptocurrencies can be viewed as time series data, making price prediction a crucial task in this domain. The limited supply and unique characteristics of Bitcoin contribute to its highly volatile nature and lack of correlation with traditional assets. This has attracted considerable attention from financial markets, positioning cryptocurrencies as assets with distinct features [2].

In recent years, deep learning techniques, especially those leveraging convolutional and long short-term memory (LSTM) layers, have gained prominence in time series prediction tasks, including

DOI: 10.54254/2755-2721/32/20230175

cryptocurrency market analysis [3,4]. Convolutional layers are effective in filtering noise and extracting meaningful features from complex time series data. They excel at capturing intricate patterns and relationships that may not be apparent at first glance.

By combining convolutional and LSTM layers in deep learning architectures, researchers have developed models that can effectively analyze and predict trends in the cryptocurrency market. This hybrid approach capitalizes on the feature extraction capabilities of convolutional layers and the ability of LSTM layers to capture complex temporal dependencies. As a result, it shows promise in enhancing the accuracy and reliability of time series predictions within the realm of cryptocurrency analysis [5].

In this article, the main objective is to compare the performance of various models in predicting cryptocurrency prices using three different datasets. Specifically, this work analyzes the effectiveness of the Moving Average, logistic regression, ARIMA, LSTM, and CNN-LSTM models. By conducting this comparative analysis, the author aims to uncover underlying patterns in cryptocurrency price movements and identify the most accurate and reliable approach for predicting future prices. Through the research, the author seeks to contribute to the development of improved methodologies for cryptocurrency price prediction.

2. Method

2.1. Dataset

The "Cryptocurrency Price Analysis Dataset: BTC, ETH, LTC (2018-2023)" is a comprehensive and valuable resource for researchers, analysts, and cryptocurrency enthusiasts. Covering a period of over five years, from January 1, 2018, to May 31, 2023, this dataset captures the daily price movements of six major cryptocurrencies: Bitcoin (BTC), Ethereum (ETH) and Litecoin (LTC).

With this dataset, the historical price behavior of these popular digital assets could be explored and analyzed. It enables the study of long-term trends, identification of volatility patterns, and gaining insights into the dynamics of the cryptocurrency market.

2.2. Preprocessing

In the research, the Min-Max normalization method is utilized to preprocess the figures. Referred to as feature scaling or data normalization, Min-Max normalization is a widely used data transformation method to ensure all data values are scaled proportionally within a specified range.

The concept of Min-Max normalization is straightforward: by identifying the minimum and maximum values in the dataset, the data is linearly mapped to a new range, typically between 0 and 1. To perform Min-Max normalization for a given feature or variable, following formula is leveraged:

$$x_{normalizaion} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

x represents an observation from the original data, minimum value is denoted as x_{min}, and maximum value is denoted as x_{max}.

One of the benefits of Min-Max normalization is its simplicity and ease of implementation. It does not change the shape of the data distribution, but rather linearly maps the data to a new range. This allows for the comparison and uniform treatment of features with different scales and ranges, eliminating any potential bias towards certain features due to scale differences. This preprocessing technique helped enhance the training effectiveness, stability, and prediction accuracy of the models.

DOI: 10.54254/2755-2721/32/20230175

2.3. Models

2.3.1. Moving average (MA). MA model is a commonly used technical analysis indicator in financial markets. It assists in identifying underlying patterns by averaging a security's price over a specified time range, effectively reducing the impact of short-term price fluctuations [6]. To make predictions using the MA model, historical data is used to calculate the moving average. The process involves summing the closing prices of the security for the chosen time period, as well as dividing it by the number of data points considered. This provides an average value that represents the current trend in the security's price.

By repeating this calculation for each time step in the validation set, a series of predicted values can be generated. These predictions can provide insights into the potential future direction of the security's price based on its historical behavior.

2.3.2. Logistic regression (LR). Considering the values of a given group of predictor variables, logistic regression (LR) is a widely utilized multivariate analysis model used to forecast whether there exists a property or consequence [7]. Across various domains, this method enjoys widespread popularity, such as corporate finance, banking, and investments. LR has been extensively applied in default-prediction models, where researchers utilize multivariate discriminant analysis (MDA) techniques [8].

2.3.3. ARIMA. Autoregressive Integrated Moving Average, is a widely used statistical regression model for time series forecasting, particularly in finance. It takes into account the previous values of a time series and adjusts for non-stationarity. ARIMA combines the autoregressive (AR) and moving average (MA) models, which are fundamental components of the model. ARIMA's ability to consider lagged values and handle non-stationarity makes it a popular choice for linear time series forecasting [9].

2.3.4. LSTM. RNN (Recurrent Neural Network) was initially introduced for learning sequential patterns in time series data. To solve the problem of vanishing gradients that RNN cannot handle, LSTM (Long Short-Term Memory) was developed. It incorporates three gate mechanisms within its structure, which belongs to recurrent neural network that effectively tackle the problem. Additionally, LSTM introduces a separate mechanism for memory cell transmission, allowing information to be propagated across different time intervals. This makes LSTM suitable for extracting temporal features from time series data and enables it to learn long-term dependencies within the sequence. The structure of LSTM consists of three types of gates: input gate, forget gate, and output gate [10]. These gates control the flow by selectively enabling or blocking the entry and exit of data in the neurons. The neuron's input gate regulates the data to be accepted, the output gate governs the data to be transmitted, and the forget gate determines the data to be disregarded. Furthermore, hidden state after computation also serves as the historical hidden state for the next neuron. The computation of current neuron's state after processing is different from that of represents the hidden state after computation, allowing for independent storage of memory data and long-term memory capabilities.

DOI: 10.54254/2755-2721/32/20230175

2.3.5. CNN-LSTM. This work developed a CNN-LSTM model tailored for time series forecasting in the cryptocurrency market. The architecture of the model combines Convolutional Neural Networks (CNNs) with Long Short-Term Memory (LSTM) networks to effectively capture both local patterns and temporal dependencies present in cryptocurrency price data. The CNN component of the model utilizes 1D convolutional layers to extract local features from the historical price data. By applying filters to the input sequence, the CNN identifies and captures important patterns, such as short-term fluctuations and local trends. The use of multiple convolutional layers with batch normalization helps enhance the model's feature extraction capabilities. The output from the CNN layers is then fed into the LSTM component of the model. The LSTM layers are capable of grasping and understanding long-term relationships within the time series data. This includes capturing recurring patterns and seasonality present in the data [10]. By incorporating LSTM layers, the model can effectively capture the complex relationships and dependencies present in cryptocurrency price data, which is crucial for accurate prediction.

The loss function of the model utilizes the mean squared error (MSE) to quantify the discrepancy of its predicted cryptocurrency prices from the actual prices during the training process. To update the model's parameters and minimize the loss, this work considers the Adam optimizer. The model is trained using a historical dataset of cryptocurrency prices, iteratively optimizing the model over multiple epochs.

By leveraging the combined power of CNNs and LSTMs, the proposed CNN-LSTM model can effectively analyze and forecast cryptocurrency prices.

2.4. Evaluation matrixes

Widely used MASE, RMSE, and RMAE are chosen to be model evaluation metrics. MASE measures the relative accuracy of a model by comparing it to a naive or baseline model, with values less than 1 indicating better performance. RMSE, irrespective of the scale of the values, is a useful method to uncover relatively large prediction errors. It calculates the average magnitude of residuals between predicted and actual values, penalizing larger errors more than MAE. RMAE, similar to RMSE, uses absolute errors and provides a measure of average magnitude. These metrics assist in assessing the accuracy and quality of predictions, enabling model comparison, selection, and performance monitoring. These equations are shown below, where R_i represents the true price and $\widehat{R_i}$ represents the predicted ones.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{R_i} - R_i)^2} \tag{2}$$

$$\mathrm{RMAE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|\widehat{R_i} - \mathrm{R}_i|} \tag{3}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\widehat{R_i} - \mathrm{R_i}}{R_i}\right| \tag{4}$$

3. Result and discussion

3.1. Performance on Bitcoin (BTC)

The experimental results of five models on the BTC dataset are demonstrated in Figure 1.

The MA, Logistic Regression, and ARIMA models consistently exhibit a similar pattern in their predictions, showing an upward trend, while the actual prices experience a downward trend followed by an upward trend.

However, it is important to note that these models have limitations and may struggle to capture complex trends and non-linear patterns in the data. Their inability to accurately reflect the actual trend indicates the need for more sophisticated and flexible models that can capture the nuances of the underlying data dynamics. Both the LSTM and CNN-LSTM models demonstrate a strong fit to the actual price trend. They are able to capture the complex patterns and dynamics present in the data more effectively compared to other models.

DOI: 10.54254/2755-2721/32/20230175

The time units on the x-axis are decreased for both the LSTM and CNN-LSTM models to observe more intricate trends as shown in Figure 2. Results show the CNN-LSTM model outperforms the LSTM model in capturing price trends with greater detail. The CNN-LSTM model offers more refined predictions, offering a deeper understanding of the patterns in price movement. However, both models exhibit lagging effects in their predictions.

Moreover, It appears that the LSTM model tends to produce predictions that are generally lower than the actual results, indicating a potential underestimation of the target variable and leading to relatively conservative predictions. Alternatively, the opposite trend was shown in the CNN-LSTM mode, suggesting that it tends to overestimate the target variable with its predictions, which may tend to be more optimistic.

It is crucial to acknowledge that the conclusions drawn from these observations are specific to the dataset and model performance mentioned in the statement. The performance of models can vary significantly depending on the characteristics and peculiarities of the dataset at hand.

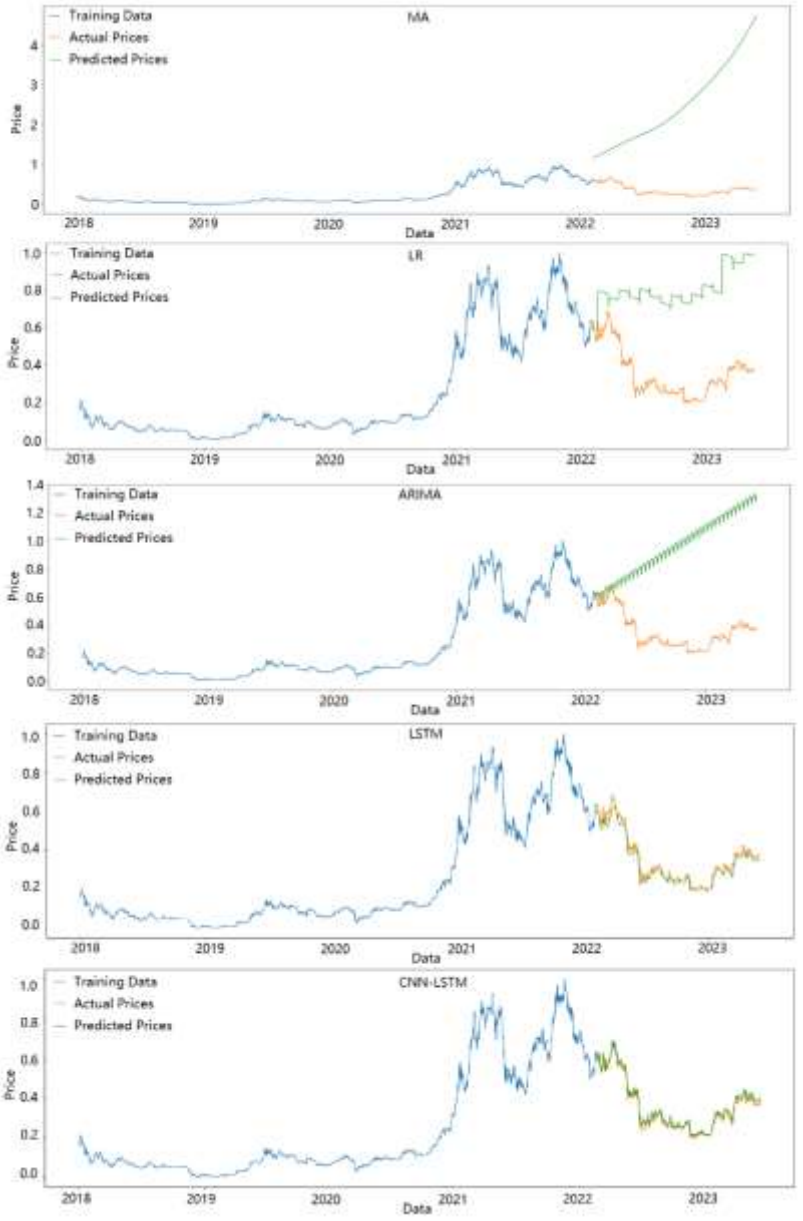


Figure 1. Prediction visualization on BTC data (Figure credit: Original).

DOI: 10.54254/2755-2721/32/20230175

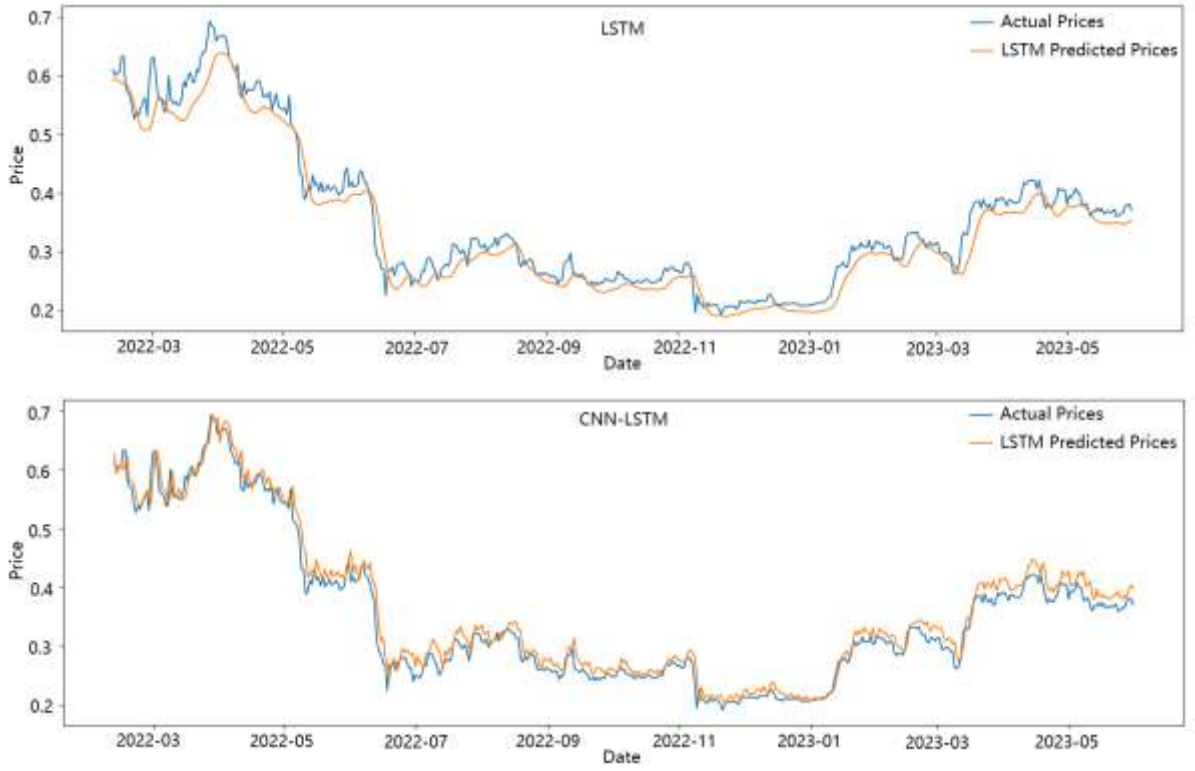


Figure 2. Actual Price and Predicted Price of LSTM and CNN-LSTM models of Bitcoin (BTC) on 11 February 2022 to 31 May 2023 (Figure credit: Original).

3.2. Performance on Litecoin (LTC)

Experimental results of a model on a single dataset can be subject to randomness. Furthermore, predictions are performed also on LTC and others. These insights are crucial for developing more accurate and reliable stock price prediction models and improving their applicability in real-world scenarios.

The experimental results of five models on the LTC dataset are shown in Figure 3.

Indeed, MA models can struggle to capture the underlying trends and dynamics of non-periodic data, and ARIMA models may have limitations when applied to cryptocurrencies due to their non-periodic nature. A variety of elements such as market sentiment, news events, and technological advancements influence Cryptocurrency prices, which may not adhere to the assumptions of stationarity and periodicity in ARIMA modeling.

An interesting observation is that the Logistic Regression (LR) model, despite its higher predicted values compared to the actual situation, provides a general prediction that aligns more closely with the true trend, which can still be valuable in certain scenarios. Similar to BTC, Both the CNN and CNN-LSTM models demonstrate accurate predictions in the given context.

The time units is decreased ton the x-axis for both the LSTM and CNN-LSTM models to observe more intricate trends as show in Figure 4. The analysis indicates that the CNN-LSTM model outperforms the LSTM model in capturing price trends with a higher level of detail. In contrast, the result from LSTM appears to struggle in predicting local peak values and tends to generate more conservative and lower predictions compared to the true values. This suggests that the LSTM model may not fully capture the extreme fluctuations or sudden spikes in cryptocurrency prices.

DOI: 10.54254/2755-2721/32/20230175

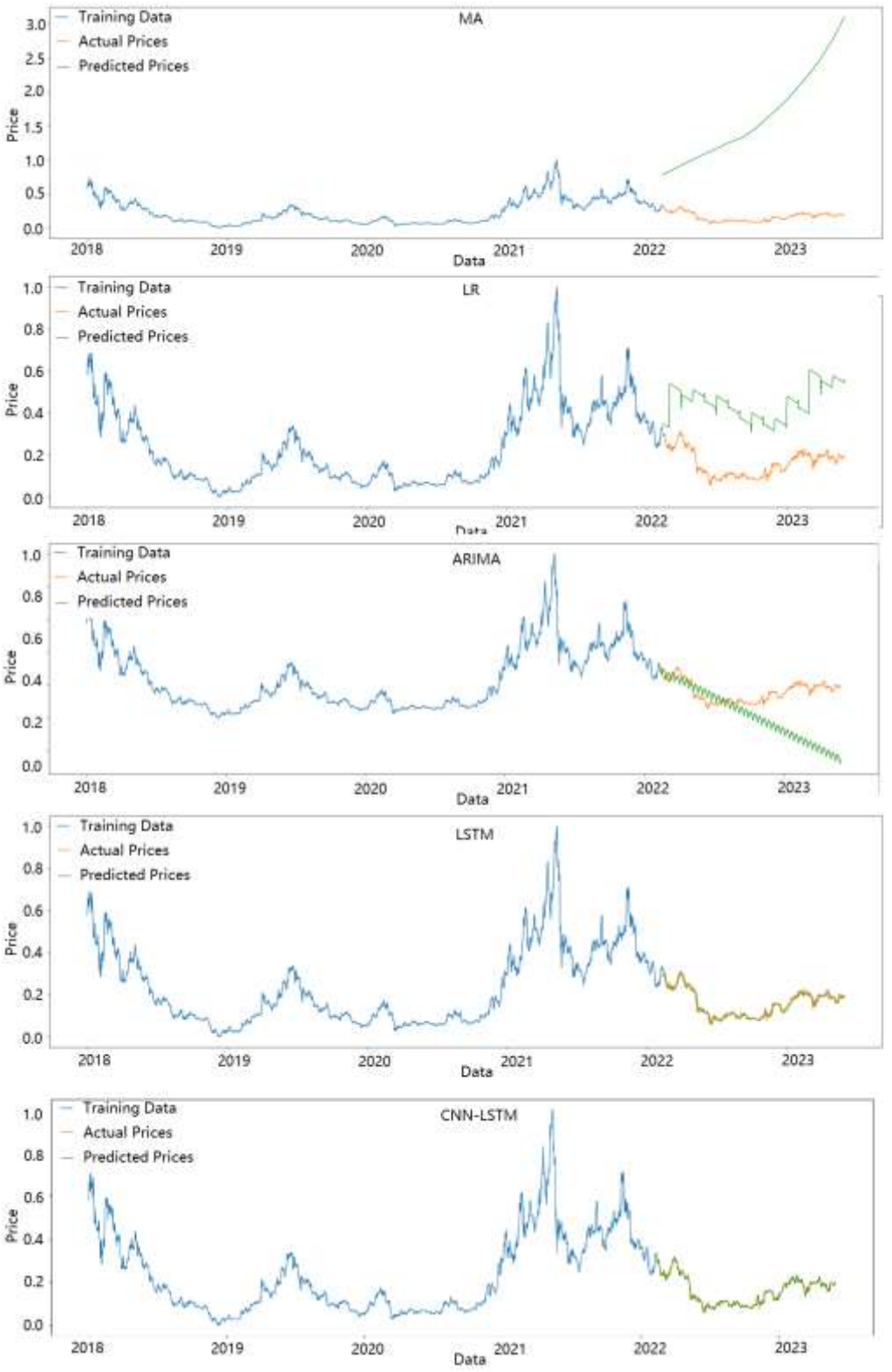


Figure 3. Prediction visualization on LTC data (Figure credit: Original).

DOI: 10.54254/2755-2721/32/20230175

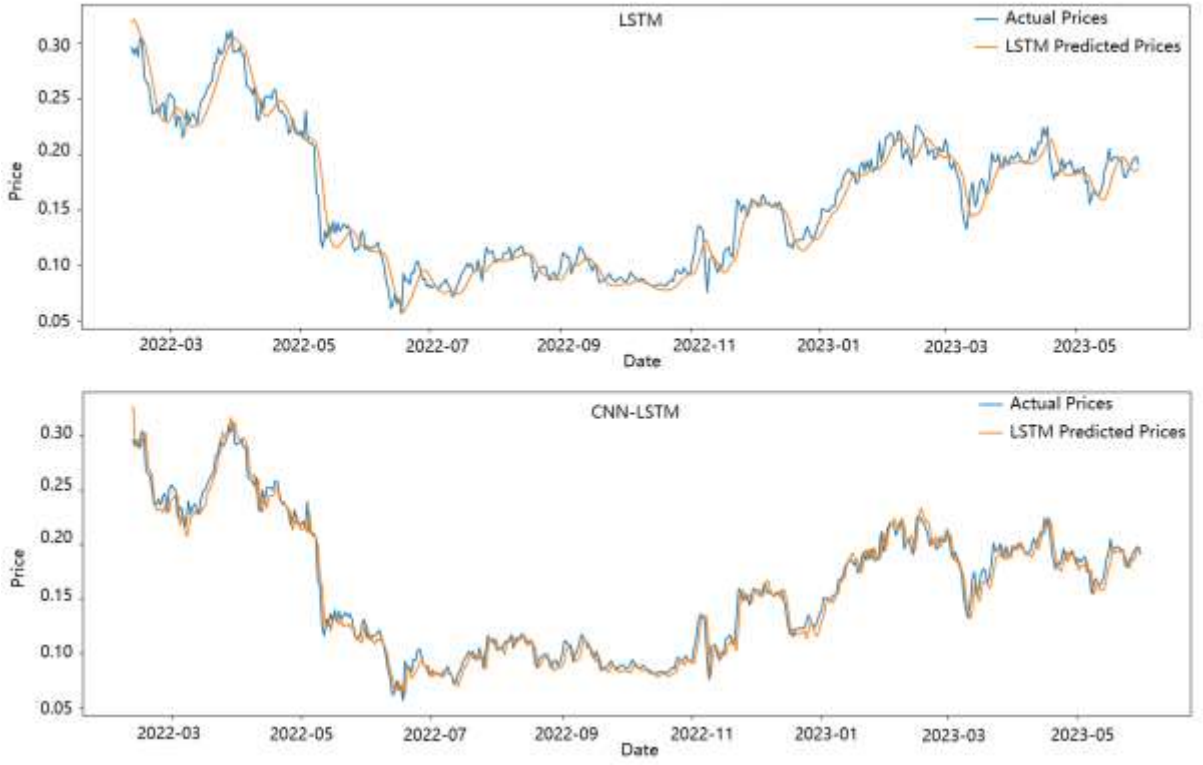


Figure 4. Actual Price and Predicted Price of LSTM and CNN-LSTM models of Litecoin (LTC) on 11 February 2022 to 31 May 2023 (Figure credit: Original).

3.3. Performance on Litecoin (ETH)

The experimental results of five models on the ETH dataset are demonstrated in Figure 5

The predictions of the five models for ETH display similar patterns to the observations discussed earlier. By examining the performance of the models across multiple datasets, a broader perspective is achieved and can assess their general predictive capabilities. As a result, it could show that the stronger predictive capabilities across the three datasets was demonstrated in CNN-LSTM model. Its ability to capture nuanced patterns and provide accurate predictions makes it a favorable choice for cryptocurrency price forecasting tasks.

It is evident from the analysis in Figure 6 that the CNN-LSTM model is in position to reflect price trends in a more detailed manner compared to LSTM. The CNN-LSTM model exhibits a higher level of granularity in its predictions, providing more nuanced insights into the price movement patterns. However, it is important to note that the predicted prices from both models tend to lag behind the actual prices, where LSTM model is more lagging, indicating a potential deviation from accuracy.

DOI: 10.54254/2755-2721/32/20230175

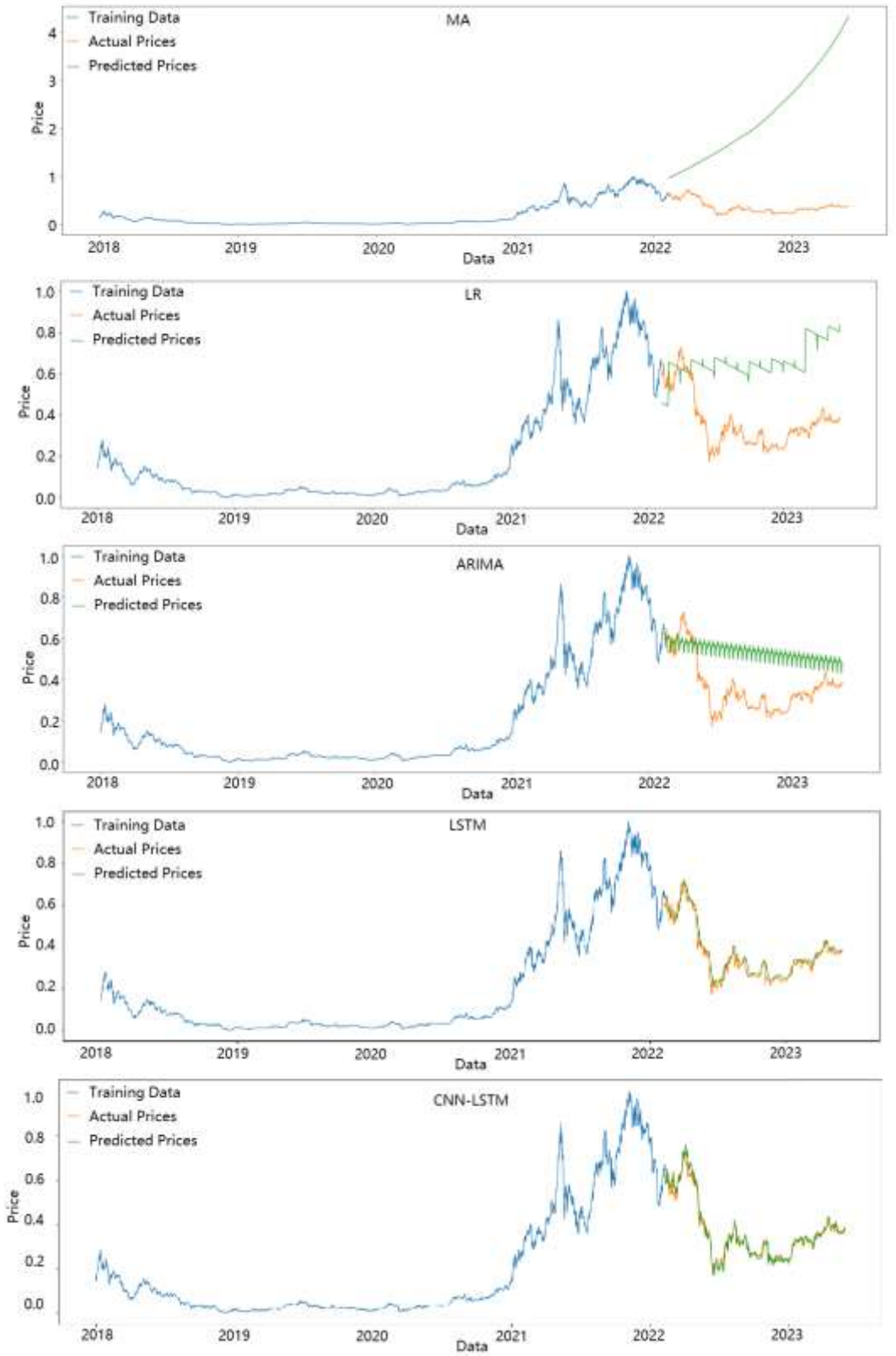


Figure 5. Prediction visualization on ETH data (Figure credit: Original).

DOI: 10.54254/2755-2721/32/20230175

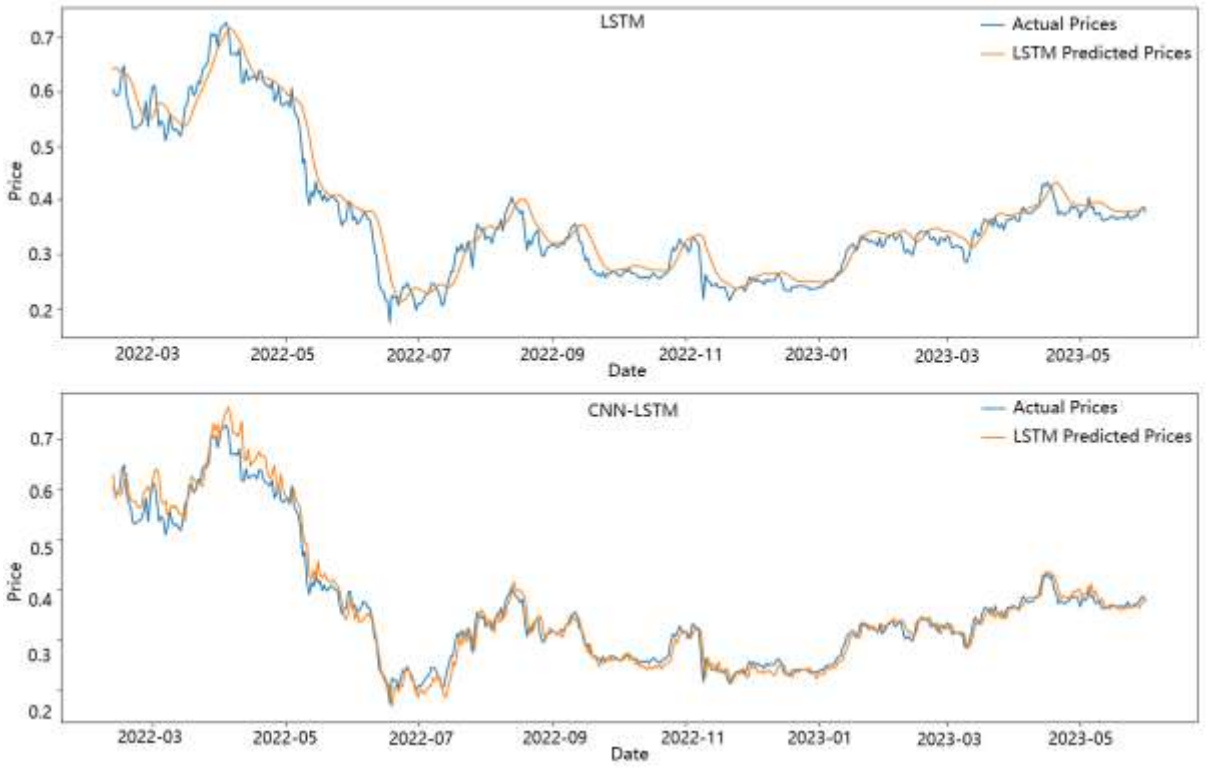


Figure 6. Actual Price and Predicted Price of LSTM and CNN-LSTM models of Ethereum (ETH) on 11 February 2022 to 31 May 2023 (Figure credit: Original)

3.4. Quantitative evaluation

By demonstrating the RMSE, RAME and MAPE values in Table 1, this work can better assess the performance of each model and compare their strengths and weaknesses.

Table 1. Comparative analysis of predictive models on cryptocurrency datasets.

Dataset	Model	RMSE	RAME	MAPE
BTC	MA	2.376	1.462	705.95%
	Linear Regression	0.478	0.671	152.53%
	ARIMA	0.675	0.779	210.76%
	LSTM	0.028	0.147	6.198%
	CNNLSTM	0.020	0.125	4.664%
LTC	MA	1.593	1.210	1041.965%
	Linear Regression	0.299	0.537	218.125%
	ARIMA	0.215	0.103	98.404%
	LSTM	0.014	0.092	7.976%
	CNNLSTM	0.011	0.086	5.004%
ETH	MA	2.148	1.381	603.958%
	Linear Regression	0.336	0.557	102.086%
	ARIMA	0.203	0.426	62.145%
	LSTM	0.030	0.148	6.425%
	CNNLSTM	0.021	0.124	4.230%

After applying different models to three kinds of cryptocurrency, Table 1 specify their corresponding performance. Among the models compared, CNN-LSTM stands out as the best performer in terms of

DOI: 10.54254/2755-2721/32/20230175

three metrics, exhibiting the minimum RMSE (<0.021), RAME (<0.125), and MAPE values (<10%). It effectively captures the price trends in the BTC dataset.

The LSTM and CNNLSTM models consistently outperform the other models across all three datasets in terms of RMSE, RAME, and MAPE, indicating their superior predictive performance. The LSTM model demonstrated the low RMSE values for BTC (0.028), LTC (0.014), and ETH (0.030), indicating its superior predictive accuracy. And the CNN-LSTM model demonstrated significant optimization over the LSTM model, resulting in a notable decrease in RMSE by 28.6% for BTC, 21.4% for LTC, and 30% for ETH, respectively.

The MA model generally exhibits the highest errors and the highest MAPE values, suggesting it may not capture the underlying patterns and dynamics effectively. ARIMA also performs well, demonstrating competitive results with significantly lower errors compared to the MA and Linear Regression models.

4. Conclusion

The analysis of the MA, LR, and ARIMA models' predictions reveals limitations in capturing the actual changes and trends in the data. The MA model exhibits a general upward trend in its predictions; however, the predicted values deviate significantly from the actual values. It tends to overestimate the actual values, and it performs poorly in capturing the peaks and fluctuations in the data. The LR model provides a relatively close approximation to the overall trend in the predictions. However, the LR model yields high values in model evaluation metrics, indicating poor performance in terms of accuracy. The ARIMA model shows relatively poor performance in capturing the overall trend of the actual data. Nevertheless, it performs relatively well in terms of evaluation metrics.

Considering the unique attributes of cryptocurrencies, approaches such as deep learning architectures like LSTM and CNN-LSTM, have shown promise in capturing and predicting the intricate dynamics of cryptocurrency prices. These models are capable of capturing non-linear relationships, long-term dependencies, and complex patterns, which can be particularly advantageous in the context of cryptocurrencies. However, it is worth noting that this work has observed consistent trends in the prices of the three cryptocurrencies during the same time period. This suggests the possibility of some degree of consistency in the results across the datasets. Further research is required to investigate the correlation among these cryptocurrencies and explore if there are underlying factors that contribute to the observed similarities. It is important to conduct more extensive studies to gain a deeper understanding of their predictive behaviors.

References

[1] Hameed, S., & Farooq, S. (2017). The art of crypto currencies: A comprehensive analysis of popular crypto currencies. arXiv preprint arXiv:1711.11073.

[2] Rebane, J., Karlsson, I., Papapetrou, P., & Denic, S. (2018). Seq2Seq RNNs and ARIMA models for cryptocurrency prediction: A comparative study. In SIGKDD Fintech'18, 19-23, 2018.

[3] Dyhrberg, A. H. (2016). Bitcoin, gold and the dollar–A GARCH volatility analysis. Finance Research Letters, 16, 85-92.

[4] Vidal, A., & Kristjanpoller, W. (2020). Gold volatility prediction using a CNN-LSTM approach. Expert Systems with Applications, 157, 113481.

[5] Livieris, I. E., Pintelas, E., & Pintelas, P. (2020). A CNN–LSTM model for gold price time-series forecasting. Neural computing and applications, 32, 17351-17360.

[6] Naved, M., & Srivastava, P. (2015). The profitability of five popular variations of moving averages on Indian market Index S&P CNX Nifty 50 during January 2004-December 2014. SSRN. 1-6.

[7] Lee, S. (2004). Application of likelihood ratio and logistic regression models to landslide susceptibility mapping using GIS. Environmental Management, 34, 223-232.

[8] Dutta, A., Bandopadhyay, G., & Sengupta, S. (2012). Prediction of stock performance in the Indian stock market using logistic regression. International Journal of Business and Information, 7(1), 105.

DOI: 10.54254/2755-2721/32/20230175

[9] Ghaderpour, E., Pagiatakis, S. D., & Hassan, Q. K. (2021). A survey on change detection and time series analysis with applications. Applied Sciences, 11(13), 6141.

[10] Thakkar, A., & Chaudhari, K. (2021). A comprehensive survey on deep neural networks for stock market: The need, challenges, and future directions. Expert Systems with Applications, 177, 114800.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230177

Exchange rate prediction research based on LSTM-ELM hybrid model

Xintong Cao

Boda College, Jilin Normal University, 1999 East Feng Road, Tiexi District, Siping City

m18844160940@163.com

Abstract. The fluctuation of exchange rates holds paramount importance for a country's economic and trade activities. Due to the non-stationary and nonlinear structural characteristics of exchange rate time series, accurately predicting exchange rate movements is a challenging task. Single machine learning models often exhibit lower precision in exchange rate prediction compared to combined machine learning models. Hence, employing a combined model approach aims to enhance the predictive performance of exchange rate models. Both Long Short-Term Memory (LSTM) and Extreme Learning Machine (ELM) exhibit intricate structures, making their direct integration challenging. To address this issue, an innovative weighted approach is adopted in this study, combining LSTM and ELM models and further refining the combination weights using an improved Marine Predators Algorithm. This paper encompasses both univariate and multivariate prediction scenarios, employing two distinct allocation strategies for training and testing datasets. This is done to investigate the influence of different dataset allocations on exchange rate prediction. Finally, the proposed LSTM-ELM weighted combination exchange rate prediction model is compared with SVM, Random Forest, ELM, LSTM, and LSTM-ELM average combination models. Experimental results demonstrate that the LSTM-ELM weighted combination exchange rate prediction model outperforms the others in both univariate and multivariate prediction settings, yielding higher predictive accuracy and superior fitting performance. Consequently, the LSTM-ELM weighted combination prediction model proves to be effective in exchange rate forecasting.

Keywords: Exchange rate prediction; Long Short-Term Memory neural network; Extreme Learning Machine

1. Introduction

In recent years, China has continuously pushed forward with the reform of its exchange rate marketization. As the status of the Renminbi (RMB) has risen in the international market, its exchange rate fluctuations have become more pronounced than before. These fluctuations not only affect investors' investment decisions but also have significant implications for enterprises' cross-border investments, arbitrage hedging, risk management, and other crucial determinations. Furthermore, they are factors that demand particular consideration when the government formulates economic policies and manages exchange rate risks. Particularly, the escalation of China-US trade tensions since 2018 and the outbreak of the COVID-19 pandemic in 2019, followed by its global spread, have further intensified the risks in

DOI: 10.54254/2755-2721/32/20230177

the RMB foreign exchange market [1]. As a result, research related to the prediction of exchange rate fluctuations has garnered extensive attention from various sectors.

Currently, scholars have conducted substantial research on the RMB exchange rate, and the methods for predicting RMB exchange rates are continuously being updated and optimized. Due to the influence of numerous intricate factors on exchange rates, predicting them remains a challenging issue. An analysis of existing literature reveals that exchange rate predictions often focus on research related to the driving forces of economic fundamentals [2], as well as technical studies based on the temporal characteristics of exchange rates themselves [3][4][5][6].

2. Theoretical Foundations

2.1. LSTM Network

The Long Short-Term Memory (LSTM) network is a specialized type of recurrent neural network that relies on three "gates" to selectively process input information. The structure of a single LSTM neuron is depicted in Figure 2.1.

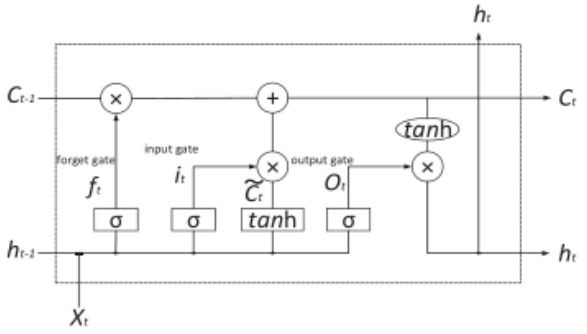


Figure 2.1. the structure of LSTM

In this figure, X_t represents input data entering the LSTM unit from the external environment, h_t denotes the output of this LSTM unit, C_{t-1} signifies the state of the previous LSTM unit at the preceding time step, h_{t-1} represents the output of the previous LSTM unit, and i_t, O_t, f_t denote the input gate, output gate, and forget gate respectively. The LSTM unit computes the current state and output based on these input data.

The specific calculation formulas are as follows:

$$i_t = s(W_{iX}X_t + W_{iM}M_{t-1} + W_{ic}C_{t\text{-}1} + b_i), \tag{2.1}$$

$$f_t = s\big(W_{fX}X_t + W_{fM}M_{t-1} + W_{fC}C_{t-1} + b_f\big), \tag{2.2}$$

$$C_t = f \otimes C_{t-1} + i_t \otimes g(W_{CX}X_t + W_{CM}M_{t-1} + b_C), \tag{2.3}$$

$$O_t = \sigma(W_{OX}X_t + W_{OM}M_{t-1} + W_{OC}C_t + b_O), \tag{2.4}$$

$$M_t = O_t \otimes h(C_t), \tag{2.5}$$

$$h_t = W_{yM}M_t + b_y. \tag{2.6}$$

Where σ is a $sigmoid$ function, $W_{iX}, W_{iM}, W_{iC}, W_{fX}, W_{fM}, W_{fC}, W_{CX}, W_{CM}, W_{OX}, W_{OM}, W_{OC}, W_{yM}$ are the weight coefficients for the forget gate, and b_i, b_f, b_C, b_O, b_y are biases term in the calculations.

2.2. Extreme Learning Machine (ELM)

Extreme Learning Machine (ELM) is a novel neural network with unique characteristics and excellent performance. It generates all hidden layer parameters randomly and balances recognition accuracy with

DOI: 10.54254/2755-2721/32/20230177

algorithm extensibility. It has found widespread applications in various research fields. Figure 2.2 illustrates the structure of ELM.

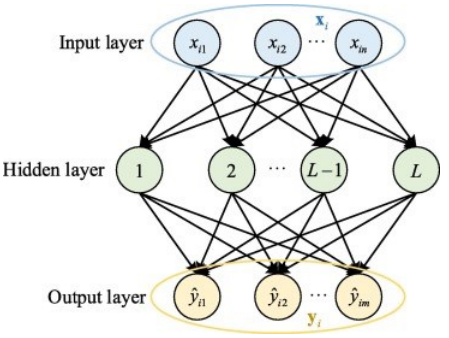


Figure 2.2. the structure of ELM

Considering the sample matrix $\{x_i, t_i\}$, where $i = 1, \cdots, N$ and N is the number of samples, $x_i = (x_{i1}.x_{i2}, \cdots, x_{in})^T \in R_n{}^m$ is the network input vector, $t_i = (t_{i1}.t_{i2}, \cdots, t_{im})^T \in R^m$ is the network output vector, $n, \overset{\Lambda}{N}, m$ are the dimensions of the input layer, hidden layer, and output layer respectively, and the activation function $g(x)$ is typically a $Sigmoid$ type. Then, the mathematical expression of ELM is given by:

$$o_j = \sum_{i=1}^{\overset{\Lambda}{N}} \beta_i g\left(w_i \cdot x_j + b_i\right), j = 1, \cdots, N. \tag{2.7}$$

Where β_i represents the connection weights between the ith hidden layer node and the output layer, w_i denotes the connection weights between the input layer and the i -th hidden layer node, and b_i is the bias of the i -th hidden layer node.

The loss function [7] of ELM is as follows:

$$E = \sum_{j=1}^{L}(\varepsilon_j), \ \ \varepsilon_j = \sum_{j=1}^{\overset{\Lambda}{N}} \beta_i g\left(w_i \cdot x_j + b_i\right) - o_j. \tag{2.8}$$

Where $\varepsilon_j = \left[\varepsilon_{j1}, \varepsilon_{j2}, \cdots, \varepsilon_{jm}\right]$ is the error for the j -th sample. Achieving zero error approximation to t_i leads to the ideal expectation: $\sum_{i=1}^{N} \|o_i - t_i\| = 0$, which means that there exists β_i, w_i, and b_i that makes $\sum_{j=1}^{\overset{\Lambda}{N}} \beta_i g\left(w_i \cdot x_j + b_i\right) = t_i$.

2.3. Marine Predators Algorithm (MPA) and Optimization

The Marine Predators Algorithm (MPA) is a new type of intelligent optimization algorithm proposed by Faramarzi et al. [8]. In MPA, each predator acts as a searching individual, and its position represents a candidate solution. Predators update their positions using predation and individual dispersion operators to ultimately obtain prey (optimal solutions). Compared to existing intelligent optimization algorithms, MPA possesses a unique search mechanism and demonstrates significant advantages in solving various classical optimization problems. However, during the optimization process involving alternating $Brownian$ and $Lévy$ motions, large step lengths may lead to the intersection of optimal solutions. To address this, an adaptive parameter controlling step length, originally expressed as Equation (2.9):

$$CF = (1 - \frac{t}{t_{\max}})^{(2\frac{t}{t_{\max}})} \tag{2.9}$$

is replaced with Equation (2.10):

DOI: 10.54254/2755-2721/32/20230177

$$CF = \frac{(1+\cos(\frac{\pi \times t}{t_{\max}}))}{2} \tag{2.10}$$

Furthermore, ideas are presented for addressing the issues of a limited initial population and bypassing local optima, as well as providing extensive exploration of the search space by introducing the Opposite-Based Learning strategy (OBL) [9]. OBL mitigates the shortcomings of a random population and enhances the convergence of the Marine Predators Algorithm. For OBL, assuming $Opp = (X_{\min} + X_{\max}) - X$ is the inverse function of a real number $X \in [X_{\min}, X_{\max}]$, with Opp being the inverse variable, the above formula can be written as:

$$\vec{Opp}_i = \left(\vec{X}_{\min} + \vec{X}_{\max} \right) - \vec{X}_i \tag{2.11}$$

Where $\vec{X}_i$ is the component of the i -th solution, and $\vec{Opp}_i$ is the inverse solution corresponding to $\vec{X}_i$.

3. Empirical Analysis

3.1. Data Source

The daily average price data used in this study is sourced from the S&P Capital IQ database. All other data, including daily trading data for USD/CNY exchange rates and indices such as NASDAQ Composite Index, Dow Jones Industrial Average, Shanghai Composite Index, and Hang Seng Index, are obtained from the Wind database. The daily average price data is used to predict the next trading day's USD/CNY exchange rate, while the eight aforementioned variables are used to predict the USD/CNY closing price for the following day. Specific data details are as follows:

(1) USD/CNY Price Data

This study selects the daily trading data for the USD/CNY exchange rate between January 1, 2015, and January 1, 2020. The data includes daily average price, opening price, highest price, lowest price, and closing price.

(2) Stock Price Data

Stock price data covers the daily trading data of the NASDAQ Composite Index (Code: IXIC.GI), Dow Jones Industrial Average (Code: DJI.GI), Shanghai Composite Index (Code: 000001.SH), and Hang Seng Index (Code: HSI.HI) between January 1, 2015, and January 1, 2020.

Excluding weekends, a total of 1305 daily average price data points are used for univariate predictions of the USD/CNY exchange rate. For multivariate predictions, a total of 1221 daily data points including opening price, highest price, lowest price, closing price, NASDAQ Composite Index, Dow Jones Industrial Average, Shanghai Composite Index, and Hang Seng Index are used for the same period.

3.2. Evaluation Metrics

Commonly used evaluation metrics to assess the performance of prediction models are R^2、MAPE、MSE, introduced as follows:

(1)R^2

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 / n}{\sum_{i=1}^{n} (\hat{y}_i - \bar{y}_i)^2 / n}, \tag{3.1}$$

DOI: 10.54254/2755-2721/32/20230177

where $\hat{y}_i$ is the predicted value, y_i is the true value, $\bar{y}_i$ is the mean of y, and R^2 ranges between 0 and 1.

(2)MSE

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2, \tag{3.2}$$

where $\hat{y}_i$ is the predicted value, y_i is the true value, and a lower MSE indicates better predictive performance.

(3)MAPE

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right|. \tag{3.3}$$

where $\hat{y}_i$ is the predicted value, y_i is the true value, and a lower MAPE indicates more accurate predictions.

3.3. LSTM-ELM Weighted Combination Exchange Rate Prediction Model

3.3.1. Model Construction

In this study, the predicted value from the LSTM model is denoted as Y_1, and the predicted value from the ELM model is denoted as Y_2. A combined prediction model is established by multiplying the two prediction results by their respective weights and then adding them together to obtain the final result of the combined method:

$$Y = W_1 Y_1 + W_2 Y_2. \tag{3.4}$$

Where W_1, W_2 represents the weighting coefficients ($W_1 + W_2 = 1$).

The improved Marine Predators Algorithm (MPA) is employed to determine the optimal ratio of the two models in the combination model. Using MPA, with MAPE as the fitness function during the optimization process, the weights of the combination are optimized. These optimized weights are assigned to the prediction values of each model to obtain the prediction values of the combined model. The optimization process is detailed in Figure 3.1.

DOI: 10.54254/2755-2721/32/20230177

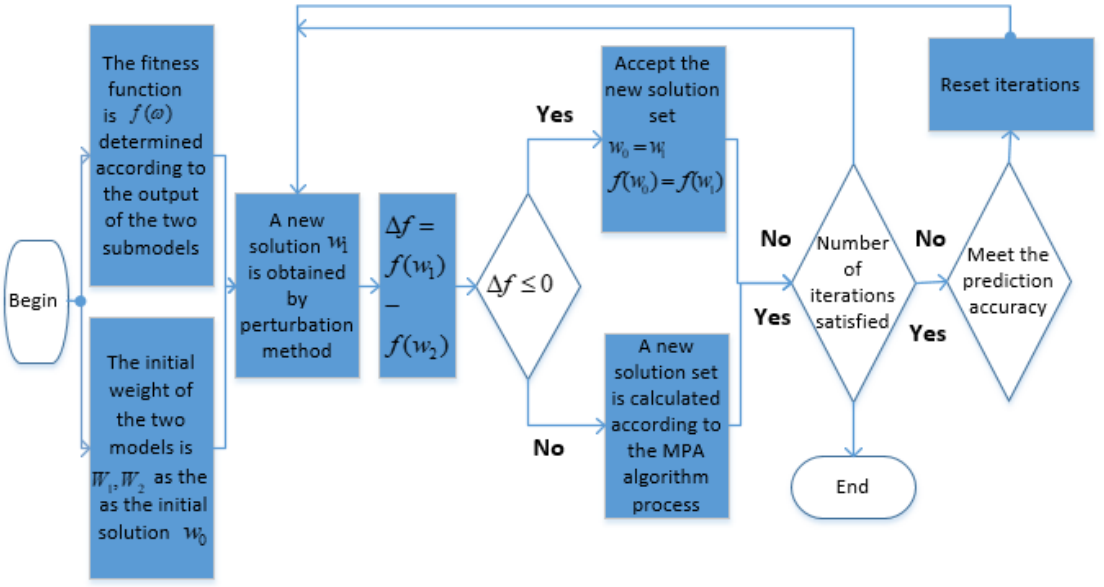


Figure 3.1. Flowchart of the Marine Predators Algorithm (MPA) for Determining Optimal Weights of Two Exchange Rate Prediction Models

3.3.2. Experimental Results

Table 3-1 presents the evaluation metrics of the univariate LSTM-ELM weighted combination exchange rate prediction model using different training datasets (80% and 90% of the data).

Table 3-1 Evaluation Metrics of Univariate LSTM-ELM Weighted Combination Exchange Rate Prediction Model with Different Training Sets

Evaluation Metric	80% Training Set Value	90% Training Set Value
MAPE	0.00164	0.00174
MSE	0.00035	0.00042
R^2	0.99898	0.99687

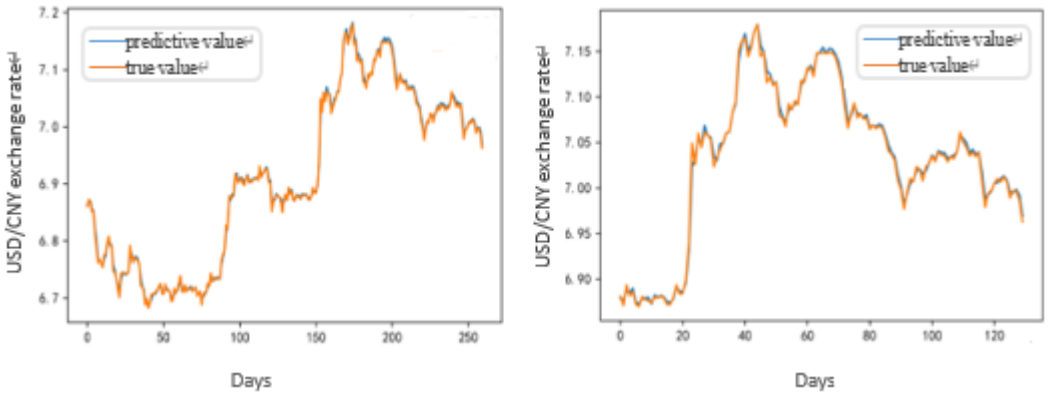


Figure 3.2. Comparison Chart of Real and Predicted Values of USD/CNY Exchange Rate by the LSTM-ELM Weighted Combination Model for 80% and 90% Training Sets

From the tables and figures, it is evident that in both training dataset scenarios, the 80% training dataset performs better in terms of prediction accuracy, as indicated by the lower MAPE and MSE values.

DOI: 10.54254/2755-2721/32/20230177

Additionally, the R^2 value is closer to 1 in the 80% training dataset scenario, indicating better fitting performance.

Table 3-2 and Figure 3.3 show the evaluation metrics and comparison of real and predicted values for the multivariate LSTM-ELM weighted combination exchange rate prediction model using different training datasets (80% and 90% of the data). Similar to the univariate scenario, the 80% training dataset outperforms the 90% training dataset in terms of prediction accuracy, as evidenced by the lower MAPE and MSE values. The R^2 value is also closer to 1 in the 80% training dataset scenario, indicating better fitting performance.

Table 3-2. Evaluation Metrics of Multivariate LSTM-ELM Weighted Combination Exchange Rate Prediction Model with Different Training Sets

Evaluation Metric	80% Training Set Value	90% Training Set Value
MAPE	0.00226	0.00262
MSE	0.00022	0.00031
R^2	0.98939	0.98897

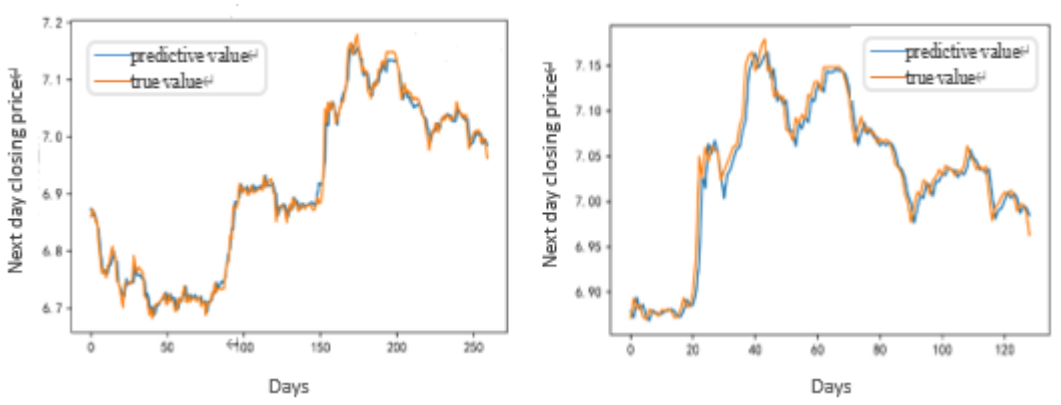


Figure 3.3. Comparison Chart of Real and Predicted Values of Multivariate LSTM-ELM Weighted Combination Exchange Rate Prediction Model for 80% and 90% Training Sets

3.4. Comparison of Prediction Results for Different Models

Based on the predicted values of the six methods, the evaluation metrics for each method are calculated for both univariate and multivariate predictions using both 80% training and 20% testing datasets, as well as 90% training and 10% testing datasets. The evaluation results are summarized in Tables 3-3, 3-4, 3-5, and 3-6.

Table 3-3. Comparison of Evaluation Metrics for Univariate Six Different Prediction Models with 80% Training Set and 20% Testing Set

Evaluation Metric \ Model	LSTM-ELM Weighted	LSTM-ELM Average	LSTM	ELM	Random Forest	SVM (Linear Kernel)
MAPE	0.00164	0.11264	0.00494	0.00377	0.00214	0.00273
MSE	0.00035	0.01023	0.00041	0.00101	0.00061	0.00039
R^2	0.99898	0.99491	0.97997	0.98592	0.99323	0.98595

DOI: 10.54254/2755-2721/32/20230177

Table 3-4. Comparison of Evaluation Metrics for Univariate Six Different Prediction Models with 90% Training Set and 10% Testing Set

Evaluation Metric \ Model	LSTM-ELM Weighted	LSTM-ELM Average	LSTM	ELM	Random Forest	SVM (Linear Kernel)
MAPE	0.00174	0.12679	0.00683	0.00232	0.00182	0.00273
MSE	0.00042	0.01155	0.00076	0.00054	0.00051	0.00047
R^2	0.99687	0.97998	0.89034	0.99219	0.99503	0.95400

Table 3-5. Comparison of Evaluation Metrics for Multivariate Six Different Prediction Models with 80% Training Set and 20% Testing Set

Evaluation Metric \ Model	LSTM-ELM Weighted	LSTM-ELM Average	LSTM	ELM	Random Forest	SVM (Linear Kernel)
MAPE	0.00226	0.11634	0.00342	0.00167	0.00263	0.00281
MSE	0.00022	0.00038	0.00056	0.00029	0.00061	0.00032
R^2	0.98939	0.98179	0.97441	0.98629	0.98144	0.98464

Table 3-6. Comparison of Evaluation Metrics for Multivariate Six Different Prediction Models with 90% Training Set and 10% Testing Set

Evaluation Metric \ Model	LSTM-ELM Weighted	LSTM-ELM Average	LSTM	ELM	Random Forest	SVM (Linear Kernel)
MAPE	0.00262	0.12992	0.00297	0.00169	0.00284	0.00272
MSE	0.00031	0.00057	0.00046	0.00032	0.00091	0.00034
R^2	0.98897	0.90936	0.86368	0.95337	0.98356	0.94890

From the tables, it is clear that among the six methods, the LSTM-ELM weighted combination exchange rate prediction model outperforms the others. This model exhibits the lowest MAPE and MSE values, indicating superior prediction accuracy. In terms of predictive performance, the R^2 value for the LSTM-ELM weighted combination model is closest to 1. Thus, the proposed LSTM-ELM weighted combination exchange rate prediction model, whether in terms of fitting performance or error values, outperforms the other five comparison models. It demonstrates excellent capability in predicting both the average USD/CNY exchange rate and the closing price for the following day.

4. Conclusion

In this study, a weighted combination exchange rate prediction model using LSTM and ELM was proposed. Through comparison with five other models, it was found that the proposed prediction model exhibited superior predictive performance and achieved favorable results in exchange rate forecasting. The primary focus of this paper was to investigate the LSTM-ELM weighted combination exchange rate prediction model in the context of both univariate and multivariate exchange rate predictions. Two testing set allocation schemes were adopted: 90% as training set and 10% as testing set, and 80% as training set and 20% as testing set, to explore the impact of different training and testing set distributions on exchange rate prediction. Concerning the LSTM-ELM weighted combination exchange rate prediction model, the allocation scheme of 80% training set and 20% testing set yielded higher prediction accuracy and better fitting results. In the future, the model proposed in this study could also be applied to address other complex forecasting problems, such as crude oil price prediction, traffic flow prediction, stock index prediction, among others.

DOI: 10.54254/2755-2721/32/20230177

References

[1] Wang, P. P. (2021). Sino-U.S. Trade Frictions, U.S. Economic Policy Uncertainty, and RMB Exchange Rate Fluctuations. *World Economy Studies*, (329), 75-92.

[2] Tatsuma, W. (2022). Out-of-sample forecasting of foreign exchange rates: The band spectral regression and LASSO. *Journal of International Money and Finance*, 128.

[3] Pengfei, L., Ze, W., Daoqun, L., et al. (2023). A CNN-STLSTM-AM model for forecasting USD/RMB exchange rate. *Journal of Engineering Research*, 11(2).

[4] Yujie, Z., Qian, Y., Zhicai, L., et al. (2023). A Prediction Model Based on Gated Nonlinear Spiking Neural Systems. *International Journal of Neural Systems*.

[5] Xueling, L., Xiong, X., Baojun, G. (2023). Increasing the prediction performance of temporal convolution network using multimodal combination input: Evidence from the study on exchange rates. *Frontiers in Physics*.

[6] Zhang, L., Sun, S., Wang, Y. (2021). Exchange Rate Forecasting Based on Deep Learning LSTM Model. *Statistics & Decision*, 37(13), 158-162.

[7] Mi, X., Liu, H., Li, Y. (2017). Wind speed forecasting method using wavelet, extreme learning machine and outlier correction algorithm. *Energy Conversion and Management*, 151, 709-722.

[8] Faramarzi, A., Heidarinejad, M., et al. (2020). Marine predators algorithm: A nature-inspired metaheuristic. *Expert Systems with Applications*, 152, 113377.

[9] Aarts, E., Aarts, E. H., Lenstra, J. K. (2003). Local search in combinatorial optimization. Princeton University Press.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230178

Review of object tracking algorithms in computer vision based on deep learning

Xiao Luo

College of Computer and Network Security (Oxford Brookes College), Chengdu University of Technology Chengdu Sichuan Province 610059 China

luo.xiao@student.zy.cdut.edu.cn

Abstract. This paper is a survey of object tracking algorithms in computer vision based on deep learning. The author first introduces the importance and application of computer vision in the field of artificial intelligence, and describes the research background and definition of computer vision, and Outlines its broad role in fields such as autonomous driving. It then discusses various supporting techniques for computer vision, including correcting linear unit nonlinearities, overlap pooling, image recognition based on semi-naive Bayesian classification, human action recognition and tracking based on S-D model, and object tracking algorithms based on convolutional neural networks and particle filters. It also addresses computer vision challenges such as building deeper convolutional neural networks and handling large datasets. We discuss solutions to these challenges, including the use of activation functions, regularization, and data preprocessing, among others. Finally, we discuss the future directions of computer vision, such as deep learning, reinforcement learning, 3D vision and scene understanding. Overall, this paper highlights the importance of computer vision in artificial intelligence and its potential applications in various fields.

Keyword: Computer Vision, Target Tracking Algorithm, Convolutional Neural Networ.

1. Introduction

Computer vision is an important research direction in artificial intelligence field. Its goal is to enable computers to perceive and understand visual information, enabling automated analysis and understanding of highlight and video data. Through computer vision technology, computer can extract meaningful information from images and videos and perform advanced visual tasks, such as target detection, image segmentation, and pose estimation. Computer vision has gone through multiple stages of research and development, form edge detection and object recognition algorithms to deep learning and convolutional neural networks, and its capabilities have been significantly improved, enabling computer to automatically process image and video data, reducing labor costs and improving work efficiency. Compute vision is widely used in many fields. For example, in the field of autonomous driving, computer vision can identify and track vehicles and pedestrians on the road to assist the realization of intelligent driving systems. It plays an important role in face recognition and medical imaging fields.

DOI: 10.54254/2755-2721/32/20230178

2. Supporting technology

2.1. Rectified Linear Unit Nonlinearity

In machine learning and neural networks, an activation function is a function that maps an input to an output to introduce nonlinear properties. Modified linear unit (ReLU) is a commonly used activation function. ReLU:$f(x) = \max(0, x)$. Where x is the input and f(x) are the output, the function has the following characteristics: when x is greater than or equal to zero, the output is equal to the input f(x)=x, that is, the function remains linear. When x is less than zero, the output is zero. Moreover, ReLU function has three characteristics: fast training, avoiding gradient disappearance problem and sparse activation [1]. In deep convolutional networks, ReLU functions are usually used as nonlinear activation functions after convolutional layers, which can help networks better learn features and improve network performance. ReLU can also be used in combination with other technologies, such as dropout, to further improve the performance of deep neural networks [1].

2.2. Overlapping pooling

Overlapping pooling is a pooling operation used to reduce the spatial dimension in the convolutional neural network [1]. Different from common pooling operations, overlapping pooling allows pooling Windows to be partially overlapped with fixed steps, thus increasing sampling density. This allows the input to be sampled at a finer granularity, capturing more characteristic information.

2.3. Image recognition based on semi-naive Bayes classification

Here, the One-Dependence Estimator is usually adopted. As shown in the formula:

$$\mathrm{P}(\mathrm{Y}|\mathrm{X}) \propto \mathrm{P}(\mathrm{Y} = y_k) \prod_{j=1}^{n} P\ (X^{(j)} = x^{(j)}|\mathrm{Y} = c_k, \mathrm{p}x^{(j)}) \quad [2]$$

Consider random variables in the input space X and the output space Y. Where y is some value in the class tag collection y = $\{ y_1, y_2, \ldots . y_k \}$. $x^{(j)}$ is j^{th} first j a characteristic of samples. P (X|Y) represents the probability distribution of x given y.

2.4. Image human behavior recognition and tracking based on S-D model

Aiming at the problem that SNBC is not accurate enough to recognize human motion in moving images, an S-D algorithm combining DT and SNBC is proposed, and the corresponding S-D model is established. The basic structure of the S-D model is shown in the figure below.

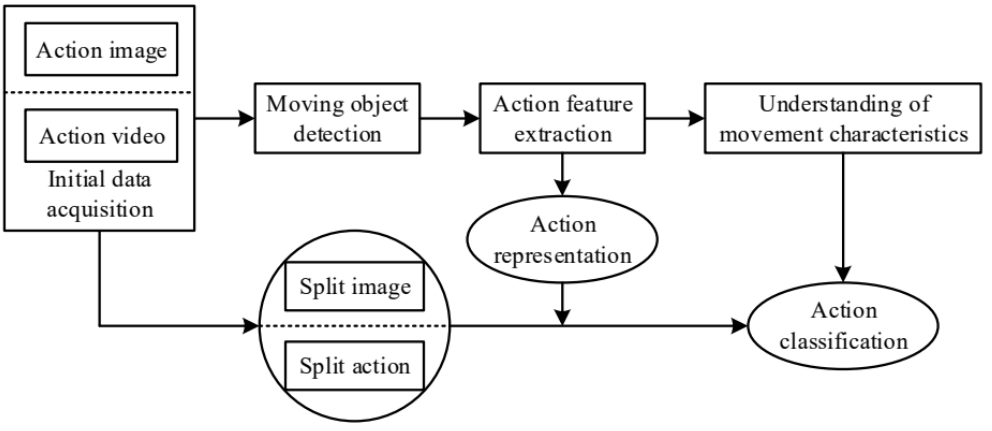


Figure 1. S-D model structure [2].

The S-D algorithm is used to calculate the optical flow information in the moving image and extract the movement trajectory of the athlete. Through feature extraction and processing of trajectory, SNBC is used for training and classification to realize recognition and tracking of human motion. For each

DOI: 10.54254/2755-2721/32/20230178

point in the image sequence, sampling points with smaller eigenvalues are deleted according to the eigenvalue threshold. The motion coordinates of the next frame are calculated by the median filter and the optical flow portion. The following formula can be used to obtain the human movement trajectory, so as to realize the movement recognition and tracking.

$$T = 0.001 * \max_{i \in I} min(\lambda_i^1, \lambda_i^2) \quad [2]$$

$$P_{t+1} = (x_{t+1}, y_{t+1}) = (x_t, y_t) + (M * \omega_t)|x_t, y_t \quad [2]$$

$$\omega_t = (u_t, v_t) \quad [2]$$

2.5. Object tracking algorithm based on convolutional neural network

Convolutional neural network is a layered model that can learn features directly from image original pixels [3]. Firstly, in the first stage, the 32*32 pre-processed black and white image is input into the convolution layer composed of 6 5*5 filters, and the feature map size is 28*28. Then, the nonlinear transformation is carried out by ReLU function. Next, use the maximum under sampling layer to select the maximum value in each 2*2 field. This network structure makes the network robust to micro translation. In the second stage, convolution and ReLU transformation are similarly carried out, and the advanced features are transformed into one-dimensional vectors by maximum subsampling feature mapping. Through these two stages of operation, the network is able to extract higher-level features.

2.6. Object tracking algorithm based on particle filter and convolutional neural network

Particle filter is a recursive Bayesian filtering algorithm, which uses sequential Monte Carlo important sampling method to represent the posterior probability. The core idea is to approximate a posterior probability distribution using a series of random particles. A particle filter has two main components: a state transition model that generates candidate samples based on previous particle samples. Observe the model and calculate the similarity between the candidate samples and the model of this objective view. A given observation sequence: $y_{1:t} = [y_1, \ldots\ldots, y_t]$, target tracking system's goal is to estimate the posterior probability density function of the target at time t $p(x_t|y_{1:t})$. According to Bayesian theory, the posterior probability density can be expressed as:

$$p(x_t|y_{1:t}) \propto p(y_t|x_t) \int p\,(x_t|x_{t-1}) p(x_{t-1}|y_{1:t-1}) dx_{t-1} \quad [3]$$

In the above formula, $p(x_t x_{t-1}), p(y_t x_t)$ are the dynamic model and the observation model, respectively. ssThe optimal target state x at the last time t can be obtained from the maximum posterior probability:

$$x_t^* = arg \max_{x_t} p\,(x_t y_{1:t}) = x_t^i = arg \max_{x_t^i} w_t^i \quad [3]$$

In order to improve the computational efficiency, the algorithm to choose only track the target position and size, $x_t = (p_t^x, p_t^y, w_t, h_t)$, in order to target the abscissa and ordinate, width and length. It is assumed that the dynamic model of two consecutive frames follows Gaussian distribution.

$$p(x_t x_{t-1}) = M(x_t; x_{t-1},) \quad [3]$$

2.7. Residual learning framework

The residual learning framework solves the problem of gradient disappearance and gradient explosion in deep neural network training [5]. It introduces a residual term so that the output of each layer includes residuals, that is, the difference between the current layer output and the input. This design makes the current layer output, and improves the model performance and generalization ability. Residuals are the key components, including convolution layer, batch normalization layer and activation function layer. Residuals are generated by transformation and added to input, which deepens the learning ability of the network [4].

DOI: 10.54254/2755-2721/32/20230178

3. Challenges and solutions

3.1. Construct a deeper convolutional neural network

Increasing network depth can improve accuracy, but building deeper overpasses requires consideration of training time, computational resources, and overfitting challenges [5]. Gradient disappearance and gradient explosion: In deep networks, gradients may gradually decrease or increase, resulting in problems such as gradient disappearance or gradient explosion, making the network difficult to train and converge. Increased computing and storage resource requirements: Deeper networks require more computing resources and storage space to handle more parameters and intermediate computation results, increasing the pressure on computing and storage. Overfitting problem: Deeper networks have stronger representation capabilities, but also tend to overfit training data and are difficult to generalize to new data samples. Activation functions (ReLU) are used to mitigate the gradient disappearance problem [1], regularization and random deactivation are used to reduce overfitting problems, and normalization is used to accelerate training and improve the stability of the network. As well as the use of residual links and attention mechanisms to help gradient propagation and optimize network structure.

3.2. Processing of large data sets

Large image databases such as ImageNet provide rich data resources, but processing these data is still challenging. Efficient algorithms and processing methods are needed to improve speed and accuracy [5]. Labeling data is expensive and time-consuming, and the dataset may be affected by class imbalance and data bias. In addition, processing big data requires powerful computing resources and storage space. In order to cope with these problems, it is necessary to effectively process data, solve class imbalance and bias, and use distributed computing to meet the processing requirements. Data storage management can be used in distributed storage systems or cloud services to increase reliability and access speed. Data preprocessing and cleaning can improve quality and reduce noise effects. Parallel computing breaks down tasks and leverages distributed computing frameworks to accelerate them. Virtual datasets can simulate real datasets by generating models, which can be used for data analysis, model training and testing. Large image databases such as ImageNet provide rich data resources, but processing these data is still challenging. Efficient algorithms and processing methods are needed to improve speed and accuracy [5]. Labeling data is expensive and time-consuming, and the dataset may be affected by class imbalance and data bias. In addition, processing big data requires powerful computing resources and storage space. In order to cope with these problems, it is necessary to effectively process data, solve class imbalance and bias, and use distributed computing to meet the processing requirements. Data storage management can be used in distributed storage systems or cloud services to increase reliability and access speed. Data preprocessing and cleaning can improve quality and reduce noise effects. Parallel computing breaks down tasks and leverages distributed computing frameworks to accelerate them. Virtual datasets can simulate real datasets by generating models, which can be used for data analysis, model training and testing [6].

4. Future development direction

The future of computer vision will focus on several key directions. In the future, computer vision will focus on several key directions: deep learning and neural network evolution, designing more complex structures to improve expression and generalization; Reinforcement learning and autonomous decision making through interaction with the environment; Realize adaptive learning under limited annotated data to quickly adapt to new fields or tasks [7].

The field of 3D vision and scene understanding will be further developed to enable accurate understanding and simulation of objects, people and actions in complex scenes. At the same time, more emphasis will be placed on multimodal vision, blending visual data from different sensors and information sources to achieve more comprehensive visual understanding and interaction. With the popularity of 3D vision systems, the design and optimization of deep learning models that process 3D data becomes important [8]. The article mentions the development of 3D CNNs (three-dimensional

DOI: 10.54254/2755-2721/32/20230178

Convolutional neural networks), and the application of geometric deep learning in computer graphics, robotics, video classification and other fields. Future research may focus on how to design more efficient 3D CNNs to apply deep learning to more complex 3D computer vision tasks. Robustness and privacy protection are becoming increasingly important, and research will focus on developing models and algorithms that can withstand adversarial attacks, as well as designing privacy-secure data processing and information transmission methods. However, differential privacy protection mechanisms can negatively affect model accuracy for unusual data or long-tail distributions [9]. Future research could delve into the impact of differential privacy on individual samples and propose solutions to improve the accuracy of the model on these data. Finally, the concept of lifelong learning will play an important role in the above direction. Models will continue to learn from new data and adapt to changing environments [10]. These common advances will advance the application of computer vision in areas such as autonomous driving, medical imaging, and intelligent safety. Therefore, computer vision will contribute strong visual perception and understanding capabilities to the development of artificial intelligence.

5. Conclusion

Computer vision is an important research direction in the field of artificial intelligence, which aims to enable computers to perceive and understand visual information. Through the introduction of technologies such as deep learning and neural networks, computer vision has made significant progress in image and video processing, and is widely used in multiple fields such as autonomous driving, face recognition, and medical imaging. In the future, the development of computer vision will focus on deep learning, reinforcement learning, three-dimensional vision and scene understanding. These developments will promote the application of computer vision technology in various fields, providing more powerful visual perception and understanding capabilities for the development of artificial intelligence.

References

[1] Krizhevsky, A, Sutskever, I and Hinton, G (2017) ImageNet Classification with Deep Convolutional Neural Networks Available at: ImageNet classification with deep convolutional neural networks | Communications of the ACM Access date: July 16, 2023

[2] Song, Y (2021) Research on Sports Image Recognition and Tracking Based on Computer Vision Technology Available at: Research on Sports Image Recognition and Tracking Based on Computer Vision Technology | IEEE Conference Publication | IEEE Xplore Access date: July 16, 2023

[3] Tian, Y and Cao, D (2022) Computer vision recognition and tracking algorithm based on convolutional neural network Available at: (PDF) Computer vision recognition and tracking algorithm based on convolutional neural network (researchgate.net) Access date: July 16, 2023

[4] He, K et.al (2015) Deep Residual Learning for Image Recognition Available at: [1512.03385] Deep Residual Learning for Image Recognition (arxiv.org) Access date: August 15, 2023

[5] Simonyan, K and Zisserman, A (2015) VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION Available at: [1409.1556] Very Deep Convolutional Networks for Large-Scale Image Recognition (arxiv.org) Access date: July 16, 2023

[6] Sun, S et.al (2020) The virtual training platform for computer vision Available at: The virtual training platform for computer vision | IEEE Conference Publication | IEEE Xplore Access date: August 16, 2023

[7] Zhang, Y , Wu, Y and Chen, H (2023) Research progress of visual simultaneous localization and mapping based on Deep learning Available at: Research progress of visual simultaneous localization and Mapping based on deep learning - CNKI (cdut.edu.cn) Access date: August 15, 2023

[8] Chen, X and Guo, H (2023) A Futures Quantitative Trading Strategy Based on a Deep Reinforcement Learning Algorithm Available at: A Futures Quantitative Trading Strategy

DOI: 10.54254/2755-2721/32/20230178

Based on a Deep Reinforcement Learning Algorithm | IEEE Conference Publication | IEEE Xplore Access date: August 15, 2023

[9] Shaqwi, F et. Al (2021) A Concise Review of Deep Learning Deployment in 3D Computer Vision Systems Available at: A Concise Review of Deep Learning Deployment in 3D Computer Vision Systems | IEEE Conference Publication | IEEE Xplore Access date: August 15, 2023

[10] Golatkar, A (2022) Mixed Differential Privacy in Computer Vision Available at: Mixed Differential Privacy in Computer Vision | IEEE Conference Publication | IEEE Xplore Access date: August 15, 2023

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230179

Investigation of medical image segmentation techniques and analysis of key applications

Hao Dong

College of Letters and Science, University of California, Los Angeles, 90024, United States

haodong@g.ucla.edu

Abstract. This research examines the application of the UNet convolutional neural network model, specifically for semantic segmentation tasks in the field of medical imaging, juxtaposing its efficacy with Fully Convolutional Networks (FCNs). The primary focus of this comparative analysis rests on the performance of the UNet model on the dataset employed for this study. Surpassing our initial expectations, the UNet model demonstrated remarkable performance superiority over the FCN model on the curated dataset, thereby suggesting its potential applicability and utility for analogous tasks within the realm of medical imaging. In a surprising turn of events, our trials revealed that data augmentation techniques did not usher in a notable enhancement in segmentation accuracy. This observation was especially striking given the substantial size of the dataset employed for the experiments, encompassing as many as 1000 images. This outcome suggests that the merits of data augmentation may not always come to the fore when dealing with considerably large datasets. This intriguing discovery prompts further exploration and investigation to uncover the underlying reasons behind this observed phenomenon. Moreover, it brings to light an open-ended research query - the quest for alternative methodologies that could potentially amplify segmentation accuracy when operating on large scale datasets in the sphere of medical imaging. As the field continues to evolve and mature, it is these open questions that will continue to push the boundaries of what is possible in medical image analysis.

Keywords: semantic segmentation, UNet, fully convolutional networks, medical imaging, data augmentation.

1. Introduction

Semantic segmentation, a critical task that partitions an image into meaningful regions by assigning a corresponding label to each pixel, has garnered considerable attention due to its extensive applications. These include autonomous driving, image comprehension, augmented reality, and notably, medical imaging. Currently, semantic segmentation is integral to medical image segmentation, supporting processes like organ segmentation, tumor delineation, lesion segmentation, and anatomical structure localization. Medical image semantic segmentation plays a pivotal role in the domain of medical imaging and computer-aided diagnosis. It involves segmenting different anatomical structures or regions of interest within medical images, such as MRI, CT scans, or ultrasound images, to aid in accurate diagnosis and treatment planning. Through semantic segmentation, precise classification and

DOI: 10.54254/2755-2721/32/20230179

segmentation of different tissue structures or pathological regions within an image are possible, thereby enabling exact localization and segmentation.

Furthermore, semantic segmentation automates the processing and analysis of images, vastly enhancing diagnostic efficiency and accuracy. This contrasts with traditional diagnostic methods, which often require manual analysis and interpretation by physicians, a process that can be time-consuming and subjective. Additionally, semantic segmentation visually represents the segmentation results, allowing physicians and patients to intuitively comprehend the distribution and morphology of pathological regions or organs. This not only facilitates effective communication but also improves patient understanding and acceptance of disease diagnosis and treatment.

Finally, semantic segmentation can be coupled with other medical imaging analysis techniques like computer-aided diagnosis and surgical navigation. This enables automation of analysis and decision-making assistance in cases such as cataract and knee joint diseases, thereby enhancing diagnostic accuracy, surgical outcomes, and overall patient care.

2. Research on key technologies of medical semantic segmentation

2.1. U-Net

To address localization in visual tasks like biomedical image processing, Ciresan et al. developed a sliding-window approach [1]. Their network predicted class labels for each pixel by using local patches as input. This approach overcame limited training data availability and enabled accurate pixel-level classification in biomedical tasks. As shown in Figure 1.

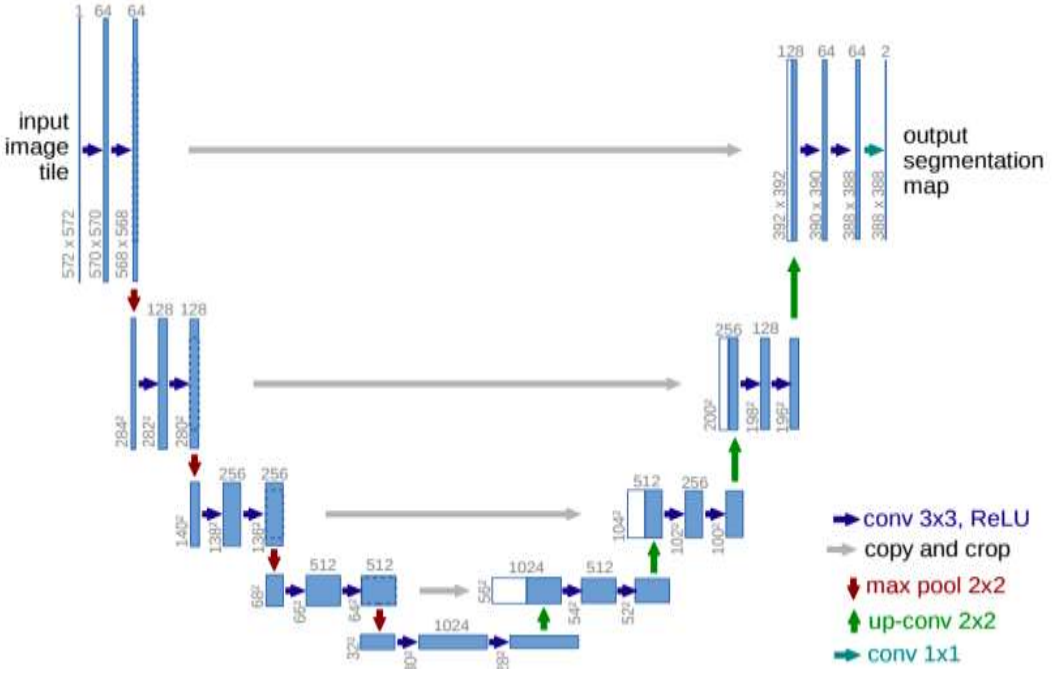


Figure 1. Unet network Structure [2].

The networks comprise a contracting path and an expansive path. To achieve downsampling and feature extraction, the contracting path utilizes a series of downsampling operations, which involve iteratively applying two unpadded 3x3 convolutions followed by ReLU activation. Additionally, 2x2 max pooling with stride 2 is employed in this process. At each downsampling step, the network doubles the number of feature channels. Conversely, the expansive path in U-Net focuses on upsampling, reconstructing the image segmentation map with higher resolution. This is achieved by applying a 2x2 convolution that reduces the number of feature channels. To ensure detailed information from earlier

DOI: 10.54254/2755-2721/32/20230179

layers is preserved, the up-convolved feature map is concatenated with the corresponding cropped feature map from the contracting path. This fusion of information facilitates the precise localization of objects in the segmentation process.

To further refine the feature map, two 3x3 convolutions with Rectified Linear Unit (ReLU) activation functions are applied to the concatenated feature map. These convolutions help enhance the discriminative power of the network and capture more intricate details present in the input image. One challenge encountered during the U-Net architecture's implementation is the loss of border pixels due to convolution operations. To address this issue, an Overlap-tile strategy is employed, which extrapolates the missing pixels by mirroring. This technique ensures that the segmentation map covers the entire input image, even at the borders. Finally, the U-Net architecture culminates in a 1x1 convolutional layer, which transforms each 64-component feature vector into the desired number of classes. This final layer plays a crucial role in generating the segmentation map, where each pixel is assigned to a specific class, enabling precise object localization and delineation. In conclusion, the U-Net architecture has become a cornerstone in image segmentation tasks, showcasing its effectiveness in various domains. Its unique combination of contracting and expansive paths, coupled with the concatenation and convolution operations, allows for accurate and detailed segmentation of objects in images. With its robust design and impressive performance, U-Net continues to be a prominent choice for researchers and practitioners in the field of computer vision.

2.2. Common network structures such as FCN

Long et al. aimed to developed a more precise and context-aware segmentation method [3]. The structure of Fully Convolutional Networks (FCNs) incorporates position attention and channel attention modules to capture semantic dependencies in both spatial and channel dimensions, resulting in improved feature representation and accurate segmentation results. FCNs consist of multiple layers of convolutional operations, which extract hierarchical features from input images. The position attention module aggregates features at each position, taking into account the relationship between similar features regardless of their spatial distance. On the other hand, the channel attention module emphasizes interdependent channel maps by integrating associated features from all channel maps. As shown in Figure 2.

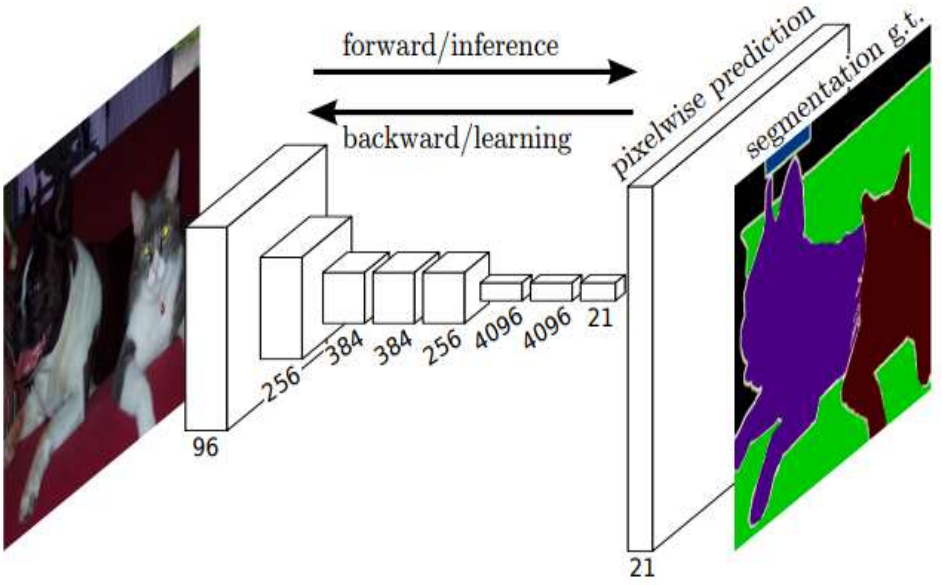


Figure 2. Fully convolutional networks [4].

Additionally, Fully Convolutional Networks (FCNs) provide flexibility in handling input image sizes. FCNs incorporate deconvolution layers to upsample the feature map from the last convolutional layer,

DOI: 10.54254/2755-2721/32/20230179

aligning it with the original image size. This enables pixel-level predictions while preserving spatial information.

2.3. Data enhancement and regularization

Data enhancement is a method to generate new training samples by performing a series of random transformations on the original training data. These transformations can include image rotation, scaling, translation, flipping and other operations, and for text data, lexical replacement, insertion or deletion can also be performed. By applying these random transformations, data enhancement expands the diversity of the training data and helps the model better learn the invariance and generalization of the data. Regularization is a technique used to control model complexity and reduce overfitting. Common regularization methods include L1 regularization and L2 regularization, which penalize the weight of the model by adding regularization terms to the model's loss function. Regularization prevents the model from over-relying on the details and noise of the training data, thereby improving its performance on previously unseen test data.

3. Analysis of typical medical semantic segmentation applications

3.1. Cataract image segmentation

To improve the performance of models on cataracts segmentation task, Grammatikopoulou et al. focus on four prominent architectures: UNet, DeepLabV3+, UPerNet, and HRNetV2 [5]. Three distinct tasks are considered, with an emphasis on different groups of instrument classes. It is used to analyze the impact of simultaneous instrument classification

The study's findings indicate that when dealing with a small number of classes and comparable pixel representation, all four networks perform similarly and successfully tackle the class imbalance. However, HRNetV2 and UPerNet exhibit superior performance compared to DeepLabV3+ and UNet, specifically in simultaneous anatomy segmentation and instrument classification with moreclasses. This advantage can be attributed to the larger receptive fields and enhanced capability of HRNetV2 and UPerNet in segmenting finer features. DeepLabV3+, UPerNet, and HRNetV2 all have a better performance in instrument segmentation and classification across all tasks than Une while UNet has a better performance in segmenting large areas. Notably, achieving spatially consistent instrument classification remains a significant challenge, as different parts of the same instrument may be classified as different types. Additionally, accurately segmenting instruments, particularly when they are inserted into anatomical structures, poses a persistent challenge. Quellec et, al proposed a solution for real-time segmentation and categorization of surgical tasks in ophthalmology using video recordings [6]. The goal was to provide timely information and recommendations to surgeons, particularly those with less experience. The proposed system utilized the content-based video retrieval paradigm and employed analogy reasoning to analyze the ongoing surgery.

The videos were segmented into idle phases, which indicated periods of no clinically-relevant motions, and action phases, representing active surgical tasks. Whenever an idle phase was detected, the preceding action phase was categorized, and the subsequent action phase was predicted using a conditional random field. To evaluate the system's performance, a dataset comprising 186 cataract surgeries performed by ten different surgeons was used. The dataset was manually annotated with up to ten possibly overlapping surgical tasks per surgery. The proposed system achieved an average recognition performance, measured by the Az metric, of 0.832 ±0.070. This experiment demonstrates the effectiveness of the proposed solution in accurately segmenting and categorizing surgical tasks during ophthalmic surgeries. The system's high recognition performance indicates its potential to provide valuable real-time information and recommendations to surgeons, aiding in the improvement of surgical outcomes, especially for less experienced practitioners.

Fox et. al. researchers present a novel application of the Mask R-CNN deep-learning segmentation method for the automatic detection and localization of surgical tools in ophthalmic cataract surgery videos [7]. Their approach involved annotating datasets for multi-class instance segmentation, and they

DOI: 10.54254/2755-2721/32/20230179

achieved promising results with a mean average precision of 61% for instance segmentation. Furthermore, the approach performed well in bounding box detection and binary segmentation tasks. To enhance the robustness of their model, they conducted an in-depth analysis of the segmentation performance for each instrument class and experimented with various data augmentation techniques. This research significantly contributes to the field of content-based video analysis in ophthalmology, offering valuable resources for training and further investigation of medical research questions in ophthalmic surgery. The successful application of deep learning in surgical tool detection and localization opens up new possibilities for improving surgical workflow and enhancing surgical outcomes in ophthalmic procedures.

3.2. Knee tissue image segmentation

Kessler et al. use a cGAN with U-Net generator and PatchGAN discriminator to segment knee joint tissues in MR images [8]. A cGAN model was successfully trained with a small dataset of only eight subjects, yielding promising results. The segmentation accuracy of bone structures surpassed expectations, with a Dice similarity coefficient (DSC) exceeding 0.95, indicating high precision. Cartilage and muscle tissues also displayed good segmentation performance, achieving a DSC above 0.83. However, the specific area of cruciate ligament segmentations revealed room for improvement, as the DSC was around 0.66. Overall, despite the limited training data, the cGAN model showcased its potential for accurate segmentation, while identifying the need for further enhancements in the cruciate ligament segmentations. It is worth noting that these segmentation results were achieved despite the limited training dataset, which consisted of only eight subjects.

To summarize, this study successfully demonstrated the application of a cGAN with a U-Net generator for knee joint tissue segmentation on MR images. While achieving high segmentation performance for several tissues, further improvements are necessary for cruciate ligament segmentations. Nevertheless, this study lays the groundwork for future technical developments and the utilization of cGANs in segmentation tasks, offering potential benefits for evaluating joint health in osteoarthritis.

Zhou, Z et al. developed new segmentation method for knee joint tissue segmentation which combining CNN, 3D fully connected CRF, and 3D simplex deformable modelling [9]. The evaluation of the method demonstrated excellent performance, with Dice coefficients exceeding 0.9 for four tissue types and mean coefficients between 0.8 and 0.9 for seven other tissue types. Joint effusion and Baker's cyst achieved mean Dice coefficients between 0.7 and 0.8. Most musculoskeletal tissues showed an average symmetric surface distance lower than 1 mm, indicating high accuracy. The method exhibited rapid segmentation, making it promising for musculoskeletal imaging applications. The method achieved accurate segmentation of knee joint tissues, offering potential benefits for clinical use and further research.

Khan, S et al. presented a deep learning framework for achieving precise knee tissue segmentation [10]. The framework merges encoder-decoder segmentation with low-rank tensor-reconstructed segmentation network. To model tissue boundaries and utilize superimposed regions, trimap generation is employed. The method achieves a segmentation dice score of 0.8925 on Osteoarthritis Initiative datasets, with cartilage segments exceeding a dice score of 0.9. The framework's performance highlights significant advancements in knee tissue segmentation, offering promising prospects for improved diagnosis and treatment of musculoskeletal disorders.

4. Experiments and analysis

4.1. Dataset description and preprocessing

Kvasir-SEG is an open-access dataset specifically developed to address the challenging task of pixel-wise image segmentation in medical image analysis. It comprises 1000 gastrointestinal polyp images and their corresponding segmentation masks and all of the data has been verified by a professional gastroenterologist. The Kvasir-SEG dataset has 1000 polyp images and their segmentation masks sourced from the Kvasir Dataset v2.

DOI: 10.54254/2755-2721/32/20230179

4.2. Experimental setups and evaluation index selection

The original dataset used in this study was divided into two subsets: testing data with 200 images and training data with 800 images. The testing data served as an independent evaluation set to measure the performance of the U-net models with different types of data augmentation.

To establish a control group, the training data without undergoing any data augmentation was used to train the U-net model and FCN through cross-validation. Both validation pixel accuracy and validation IoU were used as the evaluation metrics. Then we will select the method that has a better performace to test the impact of data enhancement. This control group allowed for a direct evaluation of the impact of data augmentation on the performance of the U-net models. Data augmentation techniques, including cropping, brightness adjustment, perspective transformation, and combinations of these techniques, were exclusively applied to the training data subset. This process involved generating augmented versions of the original training images. The augmented training data was then used for training the U-net models using cross-validation as the experimental groups. Once the U-net model was trained using the training and validation data, it was applied to the test data, and evaluation metrics, such as IoU, were calculated to measure the performance of the models. To evaluate the significance of data augmentation on the accuracy, Two Sample T-tests were employed. These tests were used to compare the performance of models trained with and without data augmentation.

4.3. Experimental results and performance comparison

In the comparison between Unet and FCN for segmentation performance, Unet demonstrated superior results. It achieved a pixel accuracy of 0.913 and an IoU (Intersection over Union) score of 0.5, indicating its excellent segmentation capabilities. On the other hand, FCN attained a pixel accuracy of 0.847 and an IoU of 0.218, indicating comparatively lower performance. These findings suggest that Unet outperforms FCN, particularly in terms of IoU, which serves as a crucial evaluation metric for segmentation tasks. Unet's higher IoU score highlights its ability to accurately delineate object boundaries, making it a more reliable choice for this dataset. As shown in Table 1.

Table 1. Results and Analysis.

	Validation Pixel Accuracy	Validation IoU
U-net	0.913	0.5
FCN	0.847	0.218

In the comparison between Unet and FCN for segmentation performance, Unet demonstrated superior results. It achieved a pixel accuracy of 0.913 and an IoU (Intersection over Union) score of 0.5, indicating its excellent segmentation capabilities. On the other hand, FCN attained a pixel accuracy of 0.847 and an IoU of 0.218, indicating comparatively lower performance. These findings suggest that Unet outperforms FCN, particularly in terms of IoU, which serves as a crucial evaluation metric for segmentation tasks. Unet's higher IoU score highlights its ability to accurately delineate object boundaries, making it a more reliable choice for this dataset. As shown in Table 1.

DOI: 10.54254/2755-2721/32/20230179

Table 2. Comparison of Segmentation Accuracy Across Different Data Augmentation Techniques.

	Mean	Sample Size	P_value
Control Group	0.51148	5	
Cropping	0.47562	5	0.6941
Brightness	0.39240	5	0.2622
Perspective	0.41924	5	0.3424
Cropping, Brightness and Perspective	0.47836	5	0.7424

In the data augmentation experiment, we analyzed a sample of size 5 for each group. Notably, the control group, which did not undergo validation, exhibited the highest average IoU score, indicating the best overall performance. Upon calculating the p-values between the experimental groups and the control group, we observed that all p-values exceeded 0.025. This suggests that, at a significance level of 0.05, data augmentation does not have a significant impact on the segmentation accuracy. Consequently, the experiment results imply that the use of data augmentation techniques does not lead to substantial improvements in the accuracy of segmentation tasks based on the evaluation of the average IoU scores. As shown in Table 2.

5. Discussion and challenges

5.1. Limitations and challenges of medical semantic segmentation

Annotated medical images with pixel-level segmentation masks are scarce and time-consuming to create. Besides, different body parts or different diseases require different kinds of training sets. The limited availability of such data hinders the training and evaluation of segmentation models. Segmentation that can be applied to medicine requires extreme precision. However, Medical images often contain intricate structures and fine-grained details that can be challenging to segment accurately. Ambiguous boundaries and subtle variations in textures and appearances make it difficult for segmentation models to precisely classify different regions.

5.2. Possible improvements and future development directions

The transformation function and loss function in the model involve several hyperparameters, such as cropping size, rotating angle, and weight. These parameters were optimized specifically for the dataset at hand but might require manual adjustment when building a new model. Alternatively, embedding these parameters into the neural network could enable automatic optimization. In the future, adversarial learning techniques, like generative adversarial networks (GANs), can be explored to enhance the robustness of the UNet model in medical image segmentation. Through adversarial learning, the U-net models can be trained to generate more accurate and visually realistic segmentations, thereby improving their overall performance.

DOI: 10.54254/2755-2721/32/20230179

6. Conclusion

The chosen UNet model, a variant of the convolutional neural network, has demonstrated superior efficiency in medical semantic segmentation as compared to FCN, according to this image dataset. Interestingly, in an experiment involving data augmentation on a dataset consisting of 1000 images, there was no observed increase in accuracy. This could potentially suggest that in the context of a fairly substantial dataset, such as one with 1000 images, data augmentation might not contribute significantly towards improving segmentation accuracy. It is imperative to further investigate and experiment to understand the underlying reasons for this lack of impact, and to explore alternative strategies that may enhance segmentation performance for datasets of this magnitude.

References

[1] Ciresan, D. C., Gambardella, L. M., Giusti, A., & Schmidhuber, J. (2012). Deep neural networks segment neuronal membranes in electron microscopy images. In Proceedings of the NIPS (pp. 2852-2860).

[2] Long, J., Evan, S., & Trevor, D. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[3] Grammatikopoulou, M., Flouty, E., Kadkhodamohammadi, A., Quellec, G., Chow, A., Nehme, J., Luengo, I., & Stoyanov, D. (2021). CaDIS: Cataract dataset for surgical RGB-image segmentation. Medical Image Analysis, 71, 102053.

[4] Quellec, G., Lamard, M., Cochener, B., & Cazuguel, G. (2014). Real-time segmentation and recognition of surgical tasks in cataract surgery videos. IEEE Transactions on Medical Imaging, 33(12), 2352-2360. doi: 10.1109/TMI.2014.2340473.

[5] Fox, M., Taschwer, M., & Schoeffmann, K. (2020). Pixel-based tool segmentation in cataract surgery videos with Mask R-CNN. In IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS) (pp. 565-568). Rochester, MN, USA.

[6] Sharma, A., et al. (2019). The optimisation of deep neural networks for segmenting multiple knee joint tissues from MRIs. Journal of Medical Systems, 43(7), 210.

[7] Zhou, Z., Zhao, G., Kijowski, R., & Liu, F. (2018). Deep convolutional neural network for segmentation of knee joint anatomy. Magnetic Resonance in Medicine, 80(6), 2759-2770.

[8] Khan, S., Azam, B., Yao, Y., & Chen, W. (2022). Deep collaborative network with alpha matte for precise knee tissue segmentation from MRI. Computer Methods and Programs in Biomedicine, 222, 106963.

[9] Jha, D., Smedsrud, P. H., Riegler, M. A., Halvorsen, P., Lange, T. d., Johansen, D., & Johansen, H. D. (2020). Kvasir-seg: A segmented polyp dataset. In International Conference on Multimedia Modeling.

[10] Siddique, N., Paheding, S., Elkin, C. P., et al. (2021). U-net and its variants for medical image segmentation: A review of theory and applications. IEEE Access, 9, 82031-82057.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230180

Utilizing stable diffusion and fine-tuning models in advertising production and logo creation: An application of text-to-image technology

Chenyang Wang

Jeme Tienyow Honors College, Beijing Jiaotong University, Beijing, 100091, China

20221020@bjtu.edu.cn

Abstract. This article delves into the implementation of text-to-image technology, taking advantage of stable diffusion and fine-tuning models, in the realms of advertising production and logo design. The conventional methods of production often encounter difficulties concerning cost, time constraints, and the task of locating suitable imagery. The solution suggested herein offers a more efficient and cost-effective alternative, enabling the generation of superior images and logos. The applied methodology is built around stable diffusion techniques, which employ variational autoencoders alongside diffusion models, yielding images based on textual prompts. In addition, the process is further refined by the application of fine-tuning models and adaptation processes using a Low-Rank Adaptation approach, which enhances the image generation procedure significantly. The Stable Diffusion Web User Interface offers an intuitive platform for users to navigate through various modes and settings. This strategy not only simplifies the production processes, but also decreases resource requirements, while providing ample flexibility and versatility in terms of image and logo creation. Results clearly illustrate the efficacy of the technique in producing appealing advertisements and logos. However, it is important to note some practical considerations, such as the quality of the final output and limitations inherent in text generation. Despite these potential hurdles, the use of artificial intelligence-generated content presents vast potential for transforming the advertising sector and digital content creation as a whole.

Keywords: stable diffusion, text-to-image technology, advertisement production, latent diffusion models.

1. Introduction

As the time-honored adage professes, "a picture is worth a thousand words." Indeed, images have an essential role in our daily lives. Ranging from the small scale of a book cover to the large scale of videos we consume, many elements of life comprise combinations of images. Particularly in the field of advertising, images possess a noteworthy capacity to seize people's attention far more effectively than words [1]. They consistently exhibit vibrant colors and compelling compositions, which command individuals' focus [2]. Moreover, a well-chosen image can directly communicate the features, functions, and advantages of a product or brand to the audience. The usage of symbols, icons, and visual elements can transmit messages, provoke emotions, and stimulate the desire to purchase [3].

DOI: 10.54254/2755-2721/32/20230180

Images possess the capacity to communicate information directly through visual means, thereby not necessitating the reliance on text [4]. Nonetheless, while there are considerable advantages to using images in advertising, creating an image is not always a convenient or cost-effective process. If you aim to incorporate models into your images to enhance their appeal, this entails significant expenses, such as initial conceptualization, preparation of photography equipment, and hiring of professionals like models, photographers, and lighting technicians. Alternatively, sourcing images from stock photo websites can be a time-intensive process and may not always yield suitable results. Even if one opts for software-generated images over real-world photographs, a professional designer is often still indispensable.

Thus, this article proposes an easier solution: generating suitable images through text-to-image technology and fine-tuning models. The images produced by this method can be effectively utilized in advertisements. Moreover, these fine-tuned models can be employed to create logos or generate creative ideas and inspiration. This approach can drastically reduce the need for human resources and material costs while producing images that align more closely with expectations.

2. Methodology

2.1. Stable diffusion technique

Stable diffusion is a technique widely used in the field of text-to-image and image-to-image recently. It is basically based on the Latent Diffusion Models, which are probabilistic generative models that achieve impressive synthesis results on high-resolution and complex scenes by decomposing the image formation process into a sequential application of denoising autoencoders [5].

It can be divided into two basic stages. The first stage is the compression learning stage, it uses variational autoencoders (VAE) to encode the input into a lower dimension. Through the compression learning stage, the model learns a low-dimensional latent representation space that is perceptually equivalent to the image space. This allows the subsequent generative learning stage to efficiently generate images in this low-dimensional space, without the need for complex computations in the high-dimensional image space [6]. In another word, the model pre-trained in this stage allows the users can do the text-to-image mission based on it.

The second stage is the generative learning. In this stage, the model utilizes the pre-trained low-dimensional latent representation space from the compression learning stage to generate images. This stage takes advantage of diffusion models (DMs), which are built from a hierarchy of denoising autoencoders. And the model architecture is enhanced with cross-attention layers, which enable the conditioning of the image generation process with inputs such as text or bounding boxes [7]. This means that users can provide additional information to guide the image synthesis, making it more flexible and versatile.

2.2. Fine-tuning and adaptation using LoRA

Low-Rank Adaption (LoRA) model was firstly proposed as a natural language processing model, it was used in the pre-trained large language models for fine-tuning in order to lead it into some specific tasks [8]. It allows the injection of trainable rank decomposition matrices into each layer of the Transformer architecture, effectively reducing the number of trainable parameters for downstream tasks.

The key principle behind LoRA is to freeze the pre-trained model weights and optimize the rank decomposition matrices instead. By doing so, LoRA achieves a significant reduction in the number of trainable parameters, making it more computationally efficient and memory-friendly. This addresses the challenge of deploying large language models, such as GPT-3, which have an enormous number of parameters and are prohibitively expensive to fine-tune independently, but now it is widely used in the process of the image generation [9].

Before the accomplishment of LoRA in the text-to-image, if the user wants to generate a specific element in an image, an anime character for example, it will be complicated and inaccurate.

DOI: 10.54254/2755-2721/32/20230180

However, the LoRA model offers a useful way to do such jobs, also with several advantages. Firstly, it allows for the sharing of a pre-trained model across multiple tasks. By freezing the shared model and only updating the rank decomposition matrices, the storage requirement and task-switching overhead are greatly reduced. Secondly, LoRA improves training efficiency by eliminating the need to calculate gradients or maintain optimizer states for most parameters. This lowers the hardware barrier to entry and enables faster training, especially when using adaptive optimizers [10].

Most Important, in the practical jobs it's twice as fast than Dreambooth method and sometimes even better performance than full fine-tuning, and the end result is also very small, which makes it easier to share and download.

The following four sets of images show the generated images of the target character, the generated images of the character without using LoRA, and the generated images of the character using LoRA under the same prompt. By comparison, it is evident that the use of LoRA produces better results.

Figure 1. The comparation of using Lora to generate images.

2.3. The Web-UI of stable diffusion

Stable Diffusion Web-UI can provide users a simple and powerful interface, which allows users to easily explore different modes and settings, and realize your creativity.

The users can utilize the text-to-image mode, allowing them to input any scene or object they can envision and have the model generate corresponding images for them. Many plugins, such as LoRA, can also be added to facilitate users in fine-tuning the generated images from various perspectives. Additionally, stable upscale mode can be used to enlarge and enhance low-resolution images, resulting in improved clarity and detail. In the following text, the Web-UI will be employed to ease the loading of various models, adjustments of prompt words, and utilization of plugins.

3. Approach

3.1. Streamlining production processes with stable diffusion and LoRA

3.1.1. Advertisement. The process of designing an outstanding advertisement graphic commences with the critical step of setting goals and identifying the intended audience. This initial phase lays the groundwork for the message the advertisement aims to convey and the audience it seeks to engage. Subsequently, market research is conducted to gain insights into the preferences, needs, and behavioral patterns of the target audience. This research also encompasses the examination of the advertising strategies and design styles employed by competitors. Such information is pivotal in better positioning the advertisement and crafting content that resonates with the audience.

Following the research, the creative conceptualization process begins. This involves brainstorming the theme of the advertisement, the method of information delivery, visual elements, and layout. The fundamental elements that are to be featured in the advertisement are identified and presented as prompts

DOI: 10.54254/2755-2721/32/20230180

in the Web-UI. The selection of suitable images and text is the next crucial step. The images, generated in bulk, are sifted through to find those that align with the advertisement's theme and appeal to the target audience. These images should be clear, attractive, and relevant. The diversity in style is achieved by adjusting different model parameters or prompt weights, based on the generated images. For content with distinct characteristics or styles, fine-tuning with the LoRA model can lead to optimal results.

Once a satisfactory image is obtained, it is further optimized and adjusted using tools such as Photoshop. This stage involves refining the details, and designing the font placement and layout to enhance the advertisement's visual appeal. The goal is to ensure the layout is clear, easily legible, and effectively guides the audience's attention.

Upon completion of the final design, the advertisement is exported in an appropriate file format, ready for publication across various media channels. Adjustments in size and format may be required to cater to the specific requirements of each advertising channel. This comprehensive process ensures the creation of an engaging and visually appealing advertisement graphic.

3.1.2. Logo. The process of designing a compelling logo commences with clearly defining the target client. Primarily, this includes understanding the brand - its products or services, target audience, and the key message they aim to convey through the logo. The inherent meaning that the logo needs to express is consequently outlined.

The next stage involves creative ideation, generating preliminary design elements and prompts corresponding to the brand's unique attributes. Here, the LoRA model is instrumental in forming the logo style based on these components. The primary elements are incorporated in the form of words or phrases separated by commas and form the basis of prompts added in the relevant areas of the Web-UI.

Subsequently, the crux of appropriate logo selection takes place. The process requires curating images from an array of generated options that cater to the client's needs. The image should be crisp, aesthetically pleasing, simplistic, and engaging. Additionally, diversification of style can be achieved by altering models or prompt weights based on the qualities of the generated images. The multitude of images can also provide inspiration, aiding in creating a more polished logo.

After obtaining a satisfactory logo, optimization and adjustments follow. Tools akin to Photoshop can be utilized for refining details and adding limited textual content to the logo to align it with the ideal version. This assures that the final product surpasses the initial concept, thereby producing an impactful and high-quality logo which conveys the brand's identity and values succinctly. This comprehensive approach to logo design ensures that the end product is not just visually appealing but encapsulates the essence of the brand successfully.

3.2. Aims and benefits of the approach

In the abstract of generating images and logos for advertisements, using the StableDiffusion AI model has numerous aims and benefits, specifically when incorporating the text-to-image technique and fine-tuning of the model. This approach fundamentally aims to streamline the image creation process, minimizing the need for extensive resources and materially intensive methods.

The main objective of this approach is to offer a convenient, cost-effective, and time-efficient solution. The utilization of the Stable Diffusion AI model eradicates the need for expensive processes, such as conceptualization, preparatory steps for photography, and outsourcing professionals like models, photographers, and lighting technicians. It also reduces dependence on stock photo websites, which can be time-consuming and might not yield suitable images.

The benefits of this model implement a value-added system that ensures quality and adaptability. It can generate images that closely align with the user's expectations while simulating the capabilities of a professional designer, providing high-quality, versatile image outputs and logos that can be customized to suit varying requirements. The fine-tuned models can be used not just for generating creative images but also for developing logos and deriving inspirations.

Furthermore, it provides flexibility and opens a realm of possibilities for advertising agencies and designers alike. By just inputting prompts, they can receive an array of image and logo options. Each

DOI: 10.54254/2755-2721/32/20230180

option can offer a unique perspective that reflects the desired message or feeling. This not only fast-tracks the creative process but also allows for exploration of diverse options without the constraints of traditional design methods.

Thus, the application of the Stable Diffusion AI model for generating advertisement images and logos through the text-to-image technique and fine-tuning strategy can revolutionize the creative designing process. It simplifies, economizes, and optimizes image creation while guaranteeing delivery of compelling visuals that encapsulate the intended messaging effectively.

4. Results

4.1. Real-world applications

4.1.1. Advertisement. Considering beer is the world-wide popular drinking, it is used as an example of the advertisement in this case. In the Results section of this study, the practical application of the Stable Diffusion AI model in generating images for a beer advertisement is explored. The primary objective is to create an appealing image that not only represents the product - beer, but also illustrates its refreshing quality. Certain elements such as condensation on the beer bottle and beer foam are included in the image to enhance its visual appeal and make it more enticing to potential customers. Consequently, the keywords for the prompt have been determined.

The selected keywords are 'beer', 'condensation', and 'beer foam'. These are fed into the AI model as prompts for image generation. The goal is to produce an image that effectively showcases a cold, refreshing beer with foam. Moreover, there is another key prompt creatively chosen to be added, 'snow'. It makes the visual elements of the image more diverse and rich, thereby creating an enticing visual representation that would appeal to the target audience's senses, and ultimately, their desire to purchase and consume the product. As shown in Figure 1.

Figure 2. Advertisements (Photo/Picture credit: Original).

4.1.2. Logo. Retaining the context from the previous sections where beer was used as the example, the same approach is applied in the creation of a logo. For this purpose, it is paramount for the logo to exhibit simplicity and clarity. In pursuit of such stable output, the 'NeverEnding Dream' checkpoint is adopted. This large model, known for generating 3D-style imagery, ensures the elimination of excessive and unnecessary details, thereby giving rise to a smoother and more cohesive image.

Moreover, in order to produce images exhibiting distinct logo traits such as uncluttered backgrounds primarily in solid colors and a prominent subject boasting unique geometric features, the LoRA model 'Anylogo' is engaged. The prompts “logo” and “beer” have been utilized here. Examining the resultant images, the following can be observed. As shown in Figure 2.

DOI: 10.54254/2755-2721/32/20230180

Figure 3. Logos (Photo/Picture credit: Original).

The results clearly demonstrate that the combination of these two models brings forth aesthetically appealing logo images. However, the text appearing in most logos does not carry explicit meaning. This occurrence is attributed to the underlying technology's inability to recognize or generate words. It can only learn to produce shapes resembling letters. Therefore, post AI generation, further adjustments to the images are necessitated, generally via software like Photoshop. The process is illustrated using one of the generated images as an instance. As shown in Figure 3 and 4.

Figure 4. Original logo (Photo/Picture credit: Original).

Figure 5. Logo without wrong texts (Photo/Picture credit: Original).

4.2. Implementation considerations

The capabilities of Stable Diffusion's AIGC technology have been instrumental in generating commendable advertisements and logos. However, a rigorous in-depth analysis has uncovered several areas that require further attention and consideration to refine the process.

One such area is the quality of the output generated by AI models. While AI models can produce a wide variety of images, the inconsistency in quality often presents significant challenges. To maximize the effectiveness of these generated images, it is crucial to ensure that they meet the standards and requirements of specific use-cases. Therefore, the development of an objective evaluation scheme, potentially drawing upon principles from mathematics, is essential for comparing the generated content with predetermined quality standards.

Another important aspect to refine is the selection of prompts in the generation process itself. The model should be sensitive enough to choose precise prompts and adjust them promptly based on the output of each iteration. When executed meticulously, this can have a dramatic impact on the quality of

DOI: 10.54254/2755-2721/32/20230180

subsequent rounds. By closely examining the shortcomings of each iterative step, it becomes possible to fine-tune the succeeding prompts, thereby ensuring a continuously improving narrative.

Turning attention to text generation, it is evident that our AI models currently have limitations. As observed in the logo generation process, the ability to generate meaningful text is somewhat restricted and often requires additional editing and adjustments once the generation is complete. This additional step in the process highlights the need for improvement and should be approached with an initiative to further evolve the models. By recognizing and addressing these concerns, users can facilitate the maturation process of our AI models, ensuring they continue to produce outputs of exceptional quality.

5. Conclusion

In conclusion, while the realm of Artificial Intelligence Generated Content (AIGC) is experiencing rapid growth, it is still in its infancy, with many areas yet to be fully explored and developed. One of the chief concerns revolves around the lack of comprehensive legal regulations pertaining to AIGC. The copyright issues associated with AIGC content are ambiguously defined across various countries and regions, creating a potential breeding ground for legal disputes.

Moreover, the credibility of AIGC content raises additional concerns. The authenticity of AI-generated images cannot be assured, potentially undermining their effectiveness in certain applications. For example, in the fashion industry, substituting real models with AI might reduce costs, but the resultant images could arouse skepticism regarding their authenticity. This could inadvertently counteract the original intention of helping customers visualize the actual impact of a product. Nevertheless, despite these challenges, the incredible potential of this technology is undeniable. When harnessed correctly, AI models like Stable Diffusion used for generating images and logos can trigger a revolution in the advertising industry, drastically reducing costs and human resource requirements. Furthermore, it can offer a more efficient and flexible approach to creating visually captivating content that aligns closely with a brand's identity and message. As this technology continues to evolve, it is anticipated that the issues surrounding its application will be addressed, and more robust legal frameworks will be established. Without a doubt, with cautious and ethical usage, AIGC technology is poised to generate substantial economic benefits in the future, revolutionizing various industries. This makes it a promising instrument in the arena of digital content creation.

References

[1] Vahid, H., & Esmae'li, S. (2012). The power behind images: Advertisement discourse in focus. International Journal of Linguistics, 4(4), 36-51.

[2] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10 684-10 695).

[3] Hu, E. J., Shen, Y., Wallis, P., et al. (2021). Lora: Low-rank adaptation of large language models. arXiv preprint. arXiv:2106.09685.

[4] Valipour, M., Rezagholizadeh, M., Kobyzev, I., & Ghodsi, A. (2022). Dylora: Parameter efficient tuning of pre-trained models using dynamic search-free low-rank adaptation. arXiv preprint arXiv:2210.07558.

[5] Rost, M., & Andreasson, S. (2023). Stable Walk: An interactive environment for exploring Stable Diffusion outputs.

[6] Pieters, R., Wedel, M., & Zhang, J. (2007). Optimal feature advertising design under competitive clutter. Management Science, 53(11), 1815-1828.

[7] Malik, M. E., Ghafoor, M. M., Iqbal, H. K., et al. (2013). Impact of brand image and advertisement on consumer buying behavior. World Applied Sciences Journal, 23(1), 117-122.

[8] Wu, J., Gan, W., Chen, Z., et al. (2023). AI-generated content (aigc): A survey. arXiv preprint. arXiv:2304.06632.

[9] Harrer, S. (2023). Attention is not all you need: The complicated case of ethically using large language models in healthcare and medicine. EBioMedicine, 90.

DOI: 10.54254/2755-2721/32/20230180

[10] Gozalo-Brizuela, R., & Garrido-Merchán, E. C. (2023). A survey of Generative AI Applications. arXiv preprint. arXiv:2306.02781.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230181

The analysis of different authors' views on recommendation systems based on convolutional neural networks

Xinyi Jiang

Brunel London School, North China University of Technology, Beijing, China, 100144

2975055252@qq.com

Abstract. Previous research revealed that the recommendation system could be based on convolutional neural networks to offer users some information which they liked to search for in the future. It is already known that the recommendation system can learn by itself, so this paper assumed that there may be other methods which can be applied to the computer program based on convolutional neural networks. This paper finds and summarizes some authors' opinions on recommendation systems based on convolutional neural networks and summarizes their skills which are used to improve the accuracy. The findings indicated that the recommendation system is feasible and is used in many fields, and it has many functions, like analyzing emotions and summarizing users' features, in addition to that, it can make proper judgements on users' preferences. And the link between users and products is very worthy of being paid attention to, and there is a need to add more reference information to the testing module to make it more accurate, and to recommendation system should not be restricted by the current data set, so there should be other analysis on information such as potential emotions to improve the independence of the recommendation system.

Keywords: recommendation system, CNN, recommendation modules, users' emotions, recommendation accuracy.

1. Introduction

The recommendation system has been promoted a lot in some fields, such as learning recommendations based on convolutional neural networks (CNN) [1], which can make individualized information recommendations for students and recommend relevant knowledge points and catch the important information from users' comments. Some of them took advantage of natural language processing to combine the arts, science, and engineering, and use computer language to illustrate users' feelings, while there are still some questions which should be addressed, including the fact that building a module of users' behaviors in a short term should focus on the recent past behaviors like shortening the data in several-year history traits of what the user looked through into one-month or several-week history trait. This paper analyzes different categories of recommendation systems used by different authors in different domains and the feature of their various applications, explores the performance of the

DOI: 10.54254/2755-2721/32/20230181

recommendation system module set up by those authors individually, and points out the advantage of those various recommendation systems with a certain function and the author's comments. This research can make different applications of recommendation systems clearer, making people know which kind of method they should use to design their recommendation systems in a specific field which can have a better impact on users' experience.

2. Classifying different recommendation modules

Regarding setting up different recommendation modules [2], one author believes that the off-line recommendation module consists of a neural matrix completion module and an off-line recommendation service module, and the recommendation algorithm uses the CNN-A algorithm. In terms of the online recommendation module, the author thinks that this module can collect the record of users' online blogs and analyze their behaviors and then makes the online recommendation list and send it to the business database, and last, illustrate it to users through the user interaction module. And this module uses the CNN-TA algorithm as its recommendation algorithm.

The author's comments: a non-functional demand which is not mentioned straightforwardly is easily ignored, but it is the key point. There should be five demands, including the interactive demand used to serve users making them receive information in a better way, the safety demand which protects users' privacy, and the accuracy demand because the accurate information can make users more willing to use the recommendation system, and the expandable demand which can adapt to users' new demand every period of time, and the stability demand which shows the system running very smoothly even in the case of the high throughput.

3. The use of a recommendation system

3.1. The application to news

As for the application of news, one author's opinion is to improve K-means [3] and two news recommendation algorithms based on CNN. In comparison, the two algorithms have an apparent effect on recommendation, but there are still some drawbacks. On the one hand, as for the K-means algorithm, it is unilateral that build users' module only by digging into the browsing history, so there is a need to add some features about the user and make users' similarity be calculated with weighted fusion to improve the users' preference module. On the other hand, the author uses the text categorisation module for the news recommendation system based on CNN to improve the training effect.

The author's comments: the media becomes more and more popular, and the chances are that people are exposed to many kinds of news, so the news recommendation system has practical significance. In addition to that, the time factor should be added to the recommendation system because users' interest varies with time passing, and this solution can catch users' real-time hobbies.

As for the application on news, the other author thinks that it is feasible to create a Point-of-Interests recommendation system [4] about the user's hobby based on detecting current events, combining with the current event, proper time opportunity and the feature of users' hobbies, and get inherent characteristics of POI and information about current events from different time periods of data on social media. Creating a tree convolutional neural network can easily and effectively deal with the information about passage meanings in those grammar trees and that information in a sequence of words. At the same time, make a standard about the POI embedding equipped with time sense to reflect those statistics in different modes into a unified embedded zone through multimode embedded modules to trail and catch information of current events on those social media.

The author's comments: combining current POI embedded vectors with a matrix factorization algorithm to create a current POI recommendation system can make the recommendation system learn about the trend of users' preferences in different time periods. Mixing embedded features of POI current events and POI auxiliary information together to help collaborative filtering recommendation modules know about the connection between each user and the POI can help the recommendation system find the potential hobby of the user. It is significant for deriving features of zones and meanings to unify the

DOI: 10.54254/2755-2721/32/20230181

position of POI, the keyword and the time unit, which can calculate the similarity between different classes and objects. The POI recommendation module based on detecting real-time events can re-sort the POI recommendation list when the event happens, which is good for exploring the real-time potential connection between different POIs to offer a more accurate real-time recommendation to users.

3.2. The application of medical care

As for the application of medical care [5], one author's opinion is to compare three control groups and one experimental group. The formers are named Moudle1, Moudle2, and Moudle3. Module 1 is based on CNN with collaborative filtering, Moudle2 is the typical 6-layer CNN, and Moudle3 fills the dimension of CNN. The latter is the CNN which is optimized. And then, add some information about the hospital level and users' demands on doctor's age and other context information to Moudle4. What is more, compare their curves about the loss and accuracy, and the fewer the statistic of loss is, the more accurate the module is. After choosing the best module, optimize it and add a convolutional layer and a pooling layer.

The author's comments: Moudle4 is apparently better than the other 3 modules. Its accuracy, recall and F1 score improved a lot, increasing by more than 20%. And when the iteration approaches 10000, the loss error amount can decrease within 0.119, and the function is nearly convergent, which demonstrates that this module has relatively high recommendation accuracy. The research is mainly about theoretical research on an online professional medical recommendation system based on CNN, and this method transfers the link between the expert's advantages and the patient's problem to a feature label, which is novel.

3.3. Hotel recommendation system with analyzing emotions

As for the hotel recommendation system with analyzing emotions [6], one author's opinion is to make a module illustrating the comment text with a vector matrix and put it into the trained analyzing emotion module. First, divide those words and use the word2vec module producing word vectors and input them to a short-text emotion analyzing module which combines Bi-LSTM with the feature of the output of CNN. And let it classify those comments into a positive kind and a negative kind. LDA extracts topic words from the positive comment group and the negative comment group. Finally, the rate of positive comments, topic words, the positive comment group and the negative comment group are put into the list of hotels.

The author's comments: as for short texts, it is feasible to combine with Bi-LSTM and the feature of the output of CNN, of which the accuracy is higher than other modules. As for long texts, multi-head attention can replace Bi-LSTM, and combining it with CNN can pay attention to different aspects of various words. Besides, the length of a sentence does not have an impact on receiving the meaning of the sentence.

On the other hand, the other author thinks that designing a module to classify emotions. First, make a BERT-CNN-Bi GRU emotion classifying module [7] to mainly solve the lack of the ability of static word vectors to express and the lack of ability of traditional emotion classifying modules to derive emotional features from texts. Using the BERT pre-training module to get dynamic word vectors can not only mix information about positions and parts of speech and so on into word vectors, but also get deeper information about word features, and alleviate the difference about various meanings of words, to make the word vector have more variable features of language meanings. Secondly, the author designs a CNN-Bi GRU-FM(CBF) individualized recommendation system, digging further into users' comments to get that information about features, mixing different vectors through the feature mixing layer, and eventually output the score about the user's forecasting goods.

The author's comments: in comparison, there is a gap of 5% between the eventual effects of the Word2Vec word vector module and Chinese BERT pre-training, therefore, the effect of dynamic word vector is improved more than that of static word vector. In terms of short text classifying, the relationship between the above passage and the following passage is quite important, and although many layers in the BERT pre-training module can get features of that deeper information, it is effective to mix with a

DOI: 10.54254/2755-2721/32/20230181

combination of neural network and downstream tasks to derive the meaning of the content and the feature to improve the quality.

3.4. The application for purchasing

As for the application on purchasing [8], one author's opinion is that the recent recommendation system relies too much on the total data set while analyzing the interactive information about users and goods to suppose those goods which users do not like or buy can improve the data set, to build an interactor matrix including users' preference and goods. And based on the product list knowledge graph and graph CNN, add the user knowledge graph. And input information through a knowledge graph and combine it with CNN to raise the accuracy of goods recommendations. In addition to that, the information between users and goods is easily caught, and the user's cold-start problem can be addressed.

The author's comments: a data form with a well-structured knowledge graph can effectively improve the accuracy, and graph CNN fully digs high-order language information of users and products, which is significantly better than only using graph CNN of the product knowledge graph, with the data of AUC of module KGCN increasing up 6.2% and the data of F1 of module KGCN increasing up 7.9%, which proves that the users' information is very essential to the recommendation system.

3.5. The application on IPTV

As for the application on IPTV [9], one author's opinion is that with the rapid development of IPTV users and the internet television industry, there is a common demand for recommending interesting TV programs for cable TV users. To address problems that how to catch the information about the feature of users' interests and the character of programs, and how to make sure the practice of deep module simplifies the time complexity, the IPTV recommendation system based on graph learning can make graph convolution operations come true to get collaborative information, and simplifying the module and another method can raise the efficiency.

The author's comments: it is feasible to put collaborative filtering based on light GCN into the IPTV intelligent recommendation field. And as for the recommendation accuracy, the data of LGN is much higher than other baseline recommendation modules. In the practical application, among 5 lists recommended for users, the top 3 are what users are interested in, which shows that using the graphical approach to build a module can learn more detailed features about data and relational learning so that the accuracy can increase a lot.

3.6. The application for shilling attacks detection

As for the protection of the safety of recommendation systems, authors think that it is important to detect shilling attacks accurately and efficiently [10], and the existing detection method usually relies on some features derived from a certain perspective of deep learning or some professional knowledge, and they think that considering automatically derive features from different perspectives, at the same time introducing the fuzzy decision, a detection method can take effect. First, make three behavior matrices from the rating value, rating preference and rating time of each user, using bicubic interpolation to scale those three matrices to get corresponding dense rating matrices, dense preference matrices and dense time matrices. Second, regard those scale matrices of any perspective from each user as an image, train CNN at three different perspectives and calculate the membership of attacking users from every perspective. At last, introducing a group of hesitant fuzzy to make a comprehensive decision about the testing result from those perspectives, according to which to recognize the attacking user.

The author's comments: use SVM-TIA, CoDetector, CNN-SAD, SDAEs-PCA, CNN-R, CNN-P and CNN-T as comparing methods to assess three evaluating indicators, the accuracy, recall and F1-measure of the database of MovieLens 1M and Amazon, the former contains 1000209 scores and their time of 3952 movie projects from 6040 users, and the latter collects 645072 users' scores and the time on 136785 projects. The threshold of hesitant fuzzy distance decisions depends on the membership of attacking users, and deep learning can distinguish normal users and malicious users according to those derived

DOI: 10.54254/2755-2721/32/20230181

feature vectors. And the result illustrates that the three evaluating indicators are better than the other seven comparing methods, getting performances of a higher quality.

3.7. The application of intelligent home theater on demand

As for those drawbacks like cold start, the lack of data, the ambiguity of potential features and the lack of interaction between users and the projects, the author think that improving the ConvMF recommendation algorithm [11] and in the field of intelligent home theater on demand, use intelligent voice interactions and web UI creatively as two interactive interfaces to design a recommendation system about a movie list to satisfy users' demands. First, create a score forecast and create a basic information feature matrix about the user and the movie, and the MFF-CF can mix multi-feature information. Second, when creating the word vector modules, add word vectors in the same corpus as an embedding layer of Text-CNN to increase the accuracy of recommendation systems. Thirdly, given that the difference of initial scores of users affects the procedure of decomposing PMF matrices, introduce the SD-ConvCMF recommendation algorithm based on deep learning, and mix it and the MFF-CF. Lastly, design two interaction modules on voice and web UI to create multi-interaction movie list recommendation systems, at the same time, create modules on collecting movie information, downloading data, the engine dealing with data, and making a list recommendation system through the distributed deployment.

The author's comments: with the complexity of recommendation systems increasing, although average each time of iterating during the procedure of Text-CNN training, the accuracy of forecasting users' scores increases up 2.11% and 3.71% respectively after the embedding layer of ConvMF being improved and the decomposing rating matrices being improved, while the accuracy of improved SD-ConvCMF forecasting increases up 5.43%. Therefore, the improved two embedding layers of Text-CNN networks will derive information which can represent potential features of movies, making some additional improvements in the accuracy of forecasting, and improving the experience of using the intelligent home theater on demand.

4. Conclusion

With information becoming more and more varied, people are exposed to much unnecessary news, the recommendation can choose corresponding information for users. Also, the chances are that users leave their views after looking through the information. Given the diversity of users' comments, the relatively more logical computer program should focus on some potential emotions, and CNN should combine with another method like the attention mechanism to raise the accuracy of the recommendation system. The recommendation system has a profound significance, in that it can be used in many fields and help people filtrate [12] some information that is unnecessary.

References

[1] Yu Bing, Yuan Beibei, Zhuang Kexin, Qian Jing, Ni Xiaoyan, Zhang Meiren. Learning Recommendation System Based on Convolutional Neural Network[J]. Journal of Fujian Computer,2020,36(04):55-57.

[2] Zhang Yifan. Research and practice of recommendation algorithm based on convolutional neural network[D]. Sichuan Normal University Press,2022, pp.50-58.

[3] Sun Chang. News recommendation system based on improved K-means and convolutional neural network[D]. Fuyang Normal University Press,2022, pp.60-62.

[4] Li Zhi, Sun Rui, Yao Yuxuan, Li Xiaohuan. Recommending Point-of-Interests with Real-Time Event Detection [J]. Data Analysis and Knowledge Discovery,2022,6(10):114-127, pp.118-131.

[5] Tang Yingjie. An online health community medical expert recommendation study based on CNN module[D]. Shanghai University of Finance and Economics Press,2021, pp.41-78.

[6] Huang Jun. Design and Implementation of Hotel Recommendation System Based on Sentiment Analysis [D]. Jiangsu University Press,2022, pp.51-58.

DOI: 10.54254/2755-2721/32/20230181

[7] Dai Yucong. Personalized Recommendation Research Based on Sentiment Analysis [D]. Donghua University Press,2022, pp.41-48.
[8] Li Xianzong. Research on Commodity Recommendation based on Double Knowledge Graph and Graph Convolution Neural Network [D]. Jilin University Press,2022, pp.12-12,57.
[9] Wei Feng, HU Fei, WANG Chenzi, WANG Liping, YANG Jiajia, YANG Zhengyi. IPTV Recommendation System Based on Lightweight Graph Convolutional Neural Network [J]. SOFTWARE,2022,43(06):6-8+25, pp.6-8,25.
[10] Cai Hongyun, YUAN Shilin, WEN Yu, REN Jichao, MENG Jie. Shilling Attacks Detection Based on CNN and Hesitant Fuzzy Sets [J]. ADVANCED ENGINEERING SCIENCES,2022,54(03), pp.80-90.
[11] Zhou Xin. Research and Application of Deep Learning Recommendation Algorithm Based on ConvMF [D]. Chongqing University Press,2020, pp.62-65.
[12] Su Yong, Xie Yu-qing, SUN Hua-cheng, WANG Yuan, WANG Yong-li. User Comment Experience Extraction Method for Recommendation System [J]. Computer Technology and Development,2022,32(06):52-56, pp.52-56.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230183

An enhanced single-disk fast recovery algorithm based on EVENODD encoding: Research and improvements

Jiangbo Luo

College of Electronics and Information Engineering, Shenzhen University, Shenzhen, 518060, China

2019282103@email.szu.edu.cn

Abstract. In the wake of rapid advancements in information technology, the need for reliable and efficient data transmission continues to escalate in importance. Channel coding, as a pivotal technology, holds significant influence over data communication. This paper delves into the fundamental technologies of channel coding and their prominent applications. Initially, the study introduces the current research status and the significance of channel coding. Following this, a comprehensive illustration and introduction to the classical coding methods of channel coding are provided. Concluding this exploration, the paper elucidates on the prevalent applications of different channel coding methodologies in scenarios such as the Internet of Things, 5G, and satellite communication, using real-world examples for clarity. Through this comprehensive research, readers gain an understanding of the key technologies underpinning channel coding, as well as the diverse applications that typify its use. By casting light on the practical implications of channel coding in contemporary technological contexts, the paper serves as a valuable resource for those seeking to deepen their knowledge and understanding of this pivotal field.

Keywords: channel coding, internet of things, 5G, satellite communications.

1. Introduction

In the contemporary digital era, the transmission and exchange of information have become essential across various industries and departments. The demand for data transmission in modern society continues to grow. Whether considering the internet, mobile communication, television broadcasting, the Internet of Things, or intelligent transportation, reliable and efficient communication systems are of vital importance to their normal operation and development. During the process of data transmission, the transfer is often affected by channel noise and interference, which leads to an increased data transmission error rate [1]. To achieve reliable data transmission within such a complex environment, it becomes necessary to utilize channel coding technology to correct and detect transmission errors. In practice, communication systems in different scenarios often have diverse transmission requirements and restrictions. Appropriate channel coding can be optimized according to the characteristics of specific channels, fully leveraging the potential of the channels. For instance, in wireless communication, channels often face multipath effects and fading phenomena. The appropriate coding can enhance the reliability of data transmission and reduce resource consumption, such as bandwidth and power, through anti-fading measures and suppressing multipath interference. In order to meet channel coding requirements in different scenarios, this paper explores the feasibility and effectiveness of various

DOI: 10.54254/2755-2721/32/20230183

encodings in diverse scenarios by understanding the use requirements of various encodings and scenarios.

2. Relevant theories

2.1. Channel coding

Channel coding stands as a pivotal technique designed to augment the reliability of data transmission and correct errors that could manifest during the process. In a communication system, data undergoes transmission through a channel susceptible to a variety of influences such as noise, interference, and fading. These elements can introduce errors into the transmitted data. The purpose of channel coding is to metamorphose the original data by inserting redundant information, culminating in encoded data. The utilization of channel coding notably enhances the reliability of data transmission as errors can be identified and rectified, ensuring the integrity of the transmitted information.

In 1948, Bell Labs' C.E. Shannon published "A Mathematical Theory of Communication" [1], a defining paper in modern information theory that established the foundation for the field of information and coding theory. This pioneering work signified the inception of a new discipline. Shannon's paper presented a key insight: the transmission rate of information is dictated by the channel capacity. As long as the transmission rate remains beneath the channel capacity, reliable communication can be achieved through error control techniques. This fundamental principle transformed the understanding of communication systems and facilitated further advancements in the field. Consequently, error-proof coding technology has attracted widespread interest and value. Prominent error-resistant coding techniques can be bifurcated into Automatic re Quest-for-Repeat and Forward-Error-Correction. In the Automatic re Quest-for-Repeat scheme, the channel code only detects whether the received code word contains error bits, and requests the sender to resend the code word in the event of an error. In the Forward-Error-Correction scheme, the coding task is not just to identify errors, but also to rectify them. This necessitates a particular design of the encoder, such coding is called error detection and error correction code. In the Automatic re Quest-for-Repeat scheme, the whole packet requires retransmission when the packet is sent incorrectly. This process is inefficient and cumbersome in real-time mobile communication. Modern wireless communication often employs Forward-Error-Correction schemes that can correct error bits in real-time. In 1950, R. Hamming proposed the Hamming code, the first practical error coding scheme capable of detecting and correcting one error element. At the transmitting end, the encoding scheme divides the input bit information sequence into groups, and each group of bit information obtains the corresponding check bit code word through the Exclusive OR operation. At the receiving end, the decoding result is also verified by the Exclusive OR operation, and the error check group and the error bit are inverted and corrected. Following this, M. Golay proposed the Golay code, referring to the idea of Hamming code, which can correct three random error bits. In 1957, Eugene Prange proposed cyclic code, whose generating matrix and check matrix possess cyclic characteristics, strict algebraic structure and simple implementation of the coding circuit. In 1960, Bose and Ray-Chaudhuri improved on the basis of the cyclic code and developed a more robust error correction Bose-Chaudhuri-Hocquenghem code. This code has exceptional performance and inherited the simple structure of cyclic code, and was widely adopted in various communication systems of that era. A parallel concatenated convolutional code referred to as Turbo code was proposed. Turbo code was the first to incorporate the concept of feedback in the circuit into the decoding structure, and its ingenious coding and decoding structure greatly elevated the performance of channel coding and approached the Shannon limit. Owing to its low coding complexity, Turbo code has evident advantages in the performance of short and medium codes, and the encoding and decoding technology is relatively mature, making Turbo code widely employed in practical mobile communication systems.

Low-Density Parity-Check Code was first suggested by Dr. Robert G. Gallager at the Massachusetts Institute of Technology in 1963. The performance of Low-Density Parity-Check codes is quite remarkable, almost approaching the Shannon limit, and can be applied to any channel. However, its decoding algorithm is exceedingly complex, and the research technical conditions at that time were

DOI: 10.54254/2755-2721/32/20230183

limited, and the Low-Density Parity-Check code did not attract the attention of most scholars after it was proposed. It was not until 1993 when Berrou and others made the breakthrough discovery of Turbo codes. Building upon this advancement, MacKay, Neal, and others revisited Low-Density Parity-Check codes around 1995. They proposed a decoding algorithm that gained widespread acceptance. In the subsequent decade, researchers have made breakthroughs in the study of Low-Density Parity-Check codes, bringing the performance of Low-Density Parity-Check codes closer to the Shannon limit, and practical applications have become feasible. Until now, the research on Low-Density Parity-Check code has been very mature, and has entered the standard of wireless communication and other related fields.

2.2. Specific coding

2.2.1. Hamming code. Hamming code is a commonly used error detection and correction coding scheme, proposed by Richard W. Hamming in 1950 [2].

The basic principle of Hamming code is to combine information bits and check bits into a coded word by adding check bits. Specifically, for a given k bits of information, the number of redundant bits needs to be selected to satisfy the relation: $2^r \geq k + r + 1$. Thus, the total length of the coded word is n = k + r. In the construction of Hamming codes, check bits are arranged at specific locations in the coded word in order to detect and correct errors during transmission. The positions of these check bits are the power positions of 2 in the binary representation (1, 2, 4, 8, etc.). Each check bit is responsible for checking a specific set of information bits.

In terms of error detection, by calculating the parity of the check bits and comparing them with the actual received check bits, it is possible to determine whether there is an error. If an error occurs, the location information of the check bit can be used to locate and correct the error.

The advantage of Hamming code is that it can detect and correct multi-bit errors and has low codec complexity. However, its main disadvantage is that it has high redundancy and certain limitations in error correction ability.

Hamming codes have gained significant popularity and find extensive utilization in computer memory and communication systems. Particularly in computer memory, these codes play a crucial role in detecting and correcting bit errors that may occur within unit memory components like RAM. In the field of communications, Hamming codes are used for error control in data transmission to ensure reliable transmission of data across radio, fiber optic and wired communication links.

In summary, Hamming code is a classic error detection and correction coding scheme, which realizes error detection and correction in the process of data transmission by adding check bits. It has low codec complexity and is widely used in computer memory and communication systems.

2.2.2. BCH code. The BCH Code (Bose-Chaudhuri-Hocquenghem Code) is a widely used linear block code scheme for error-correcting coding, independently proposed by R. C. Bose and D. K. Ray-Chaudhuri in the early 1960s [3]. It was further developed by A. Hocquenghem.

BCH codes can detect and correct a certain number of errors in the transmission process by using primitive polynomials to construct code words. It has good error correction ability and decoding efficiency, and is widely used in data storage, communication system and other fields. The generating polynomial of a BCH code is an indecomposable polynomial of low degree where the roots satisfy specific mathematical properties. By combining these primitive elements, a series of BCH codes of different lengths can be generated. During decoding, the BCH code is decoded using an error-correcting algorithm whose ability is related to the length of the code word. Common decoding algorithms include Berlekamp-Massey algorithm, Forney algorithm and so on. By decoding and correcting the received code words, the original data block can be recovered. The characteristics of BCH code are as follows:

1. powerful error correction ability: BCH code can detect and correct the transmission error of multiple error bits, with high error correction ability.

2. simple decoding algorithm: the decoding algorithm of BCH code is relatively simple, with low decoding complexity.

DOI: 10.54254/2755-2721/32/20230183

Codeword length variability: By adjusting the number of polynomials generated, BCH codes of different lengths can be generated to meet different application requirements.

3. BCH code is widely used in a variety of storage media and communication systems, such as disk drives, CD-ROMs, radio communications, etc. It plays an important role in the protection of error correction in the process of data transmission and improves the reliability and integrity of data.

In summary, BCH code is a linear block code error correction coding scheme, with strong error correction ability and low decoding complexity. It is widely used in data storage and communication systems, providing reliable protection and error correction capabilities for data transmission.

2.2.3. Urbo code. Turbo Code is a high-performance error-correcting coding scheme proposed by Claude Berrou, Alain Glavieux and Pierre Thitimajshima in 1993 [4]. Turbo code has excellent error correction performance while approaching the channel capacity through iterative decoding.

The basic principle of Turbo code is to use two or more encoders and an iterative decoder. During the encoding process, the data to be transmitted is fed into two separate encoders and produces two distinct sequences of code words. When passing through the channel, these two code word sequences are interwoven according to certain rules.

In the process of decoding, the method of iterative decoding is used to compare the received code word with the previous decoding result, and make corrections according to the comparison result. The iteration process is repeated several times until a preset number of iterations is reached or specific stopping criteria are met. The core of Turbo code is an iterative decoding algorithm, which uses a technique called iterative decoding. Iterative decoding is modified continuously by feeding the previous decoding results into the next round of decoding as feedback. This iterative correction process can effectively improve the performance of error correction. The Turbo code has the following important features:

- achieve coding efficiency close to the channel capacity, and can transmit more information under limited signal-to-noise ratio.
- has a high error correction ability, can effectively detect and correct the error generated in the transmission process.
- through iterative decoding, error correction performance can be further improved, close to the limit of channel capacity.

Because of its excellent performance, Turbo codes are widely used in many communication systems, such as mobile communications, satellite communications, digital television, wireless local area network (WLAN) and so on. In 4G mobile communication standard, Turbo code is adopted as the error correction coding scheme of downlink. In summary, Turbo code is a coding scheme with strong error correction ability, and uses iterative decoding technology to approach the performance of channel capacity. It is widely used in many communication fields and provides reliability and efficiency for data transmission.

2.2.4. LDPC code. LDPC is an error-correcting coding scheme proposed by Robert G. Gallager in 1962. LDPC codes have excellent error correction performance and low decoding complexity by means of sparse check matrix [5].

The core idea of LDPC code is to combine information bit and check bit to form code word by introducing sparse check matrix in the process of coding. In the check matrix, most of the elements are zero, and the distribution of non-zero elements is sparse. This structure allows LDPC codes to be decoded using efficient iterative Decoding algorithms, such as Message Passing Decoding.

In the encoding process, the LDPC code uses the check matrix to perform linear operations on the transmitted data and generate check bits. The generated code word consists of information bit and check bit. When decoding, the iterative decoding algorithm is used to backinfer the received code words, calculate and correct the possible errors. The iterative decoding algorithm iteratively corrects the message between the check bit and the information bit until the stop criterion is satisfied. The commonly used iterative decoding algorithms include Belief Propagation (BP) algorithm, Sum-Product algorithm, etc.

DOI: 10.54254/2755-2721/32/20230183

LDPC codes have the following features:

- low decoding complexity: due to its sparse check matrix structure, LDPC code decoding algorithm has a low complexity, suitable for real-time applications.
- excellent error correction performance: LDPC code performs well in the performance of close to the channel capacity, and can effectively detect and correct errors generated in transmission.
- flexibility: the LDPC code check matrix can be designed and adjusted according to different requirements to adapt to different channel conditions and performance requirements.

Due to its excellent performance and flexibility, LDPC codes are widely used in many communication systems, such as satellite communications, optical fiber communications, wireless local area network (WLAN), Blu-ray disc and so on. In the 5G mobile communication standard, LDPC code is adopted as the error correction coding scheme of data channel.

In a word, LDPC code is an error-correcting coding scheme implemented by sparse check matrix, which has low decoding complexity and excellent error-correcting performance. It is widely used in communication systems and plays an important role in high-speed data transmission and wireless communication.

2.2.5. Polar code. Polar Code is a coding method used in communication systems, proposed by Arikan in 2008 [6]. Polarization code transforms a group of independent sub-channels with the same channel conditions into a stable group of high performance channels and a poor group of low performance channels through a specific linear transformation. This encoding method has been theoretically proven to achieve Shannon capacity.

The main idea of polarization code is to construct code words by arranging sub-channels with different reliability, so as to achieve high reliability communication. Specifically, it utilizes a recursive algorithm to progressively "polarize" the channel by repeatedly applying a matrix transformation operation, converting the original binary input sequence into a series of code words with varying reliability. The code word corresponding to the more reliable subchannel is almost error-free in transmission, while the code word corresponding to the less reliable subchannel may have more errors in transmission. The characteristics of polarization code are as follows:

- polarization code can be flexibly designed according to different communication scenarios and needs. By choosing a suitable construction algorithm, the coding with different length and error-correcting ability can be realized.
- The decoding algorithms of polarization codes are characterized by low complexity, especially the decoding algorithms based on Successive Cancellations (SC). This makes the polarization code have high real-time and feasibility in practical application.
- polarization code in the noise and interference of the channel shows a strong anti-interference performance. By adding a proper number of check bits, error and noise in the channel can be effectively combated.

Polarization codes have been widely used in many communication standards, including the fifth generation mobile communication standard (5G) and satellite communications. It has great potential in high-speed, low latency and high reliability communication systems, and provides an important coding basis for the development of future communication technologies.

3. Application scenario research

3.1. Internet of things scenario

The Internet of Things (IoT) is a technological ecosystem that enables the connection and communication between the physical and digital realms through internet connectivity. It facilitates intelligent interconnection and data exchange among devices. At its core, the IoT aims to link diverse devices and objects to the internet, enabling them to perceive their environment, gather data, and make independent decisions. These devices can establish an internet connection using either wireless or wired methods. They can then store, process, and analyze data utilizing cloud platforms, thereby enhancing

DOI: 10.54254/2755-2721/32/20230183

the intelligence, convenience, and efficiency of various applications. Ultimately, the IoT empowers a seamless integration between the physical and digital worlds, fostering a new level of connectivity and enabling transformative capabilities.

IoT devices are extensively utilized in various domains, including consumer, commercial, industrial, and infrastructure sectors. In the IoT industry, a significant volume of image data is collected, stored, and transmitted by collaborative, low-power devices. For example, surveillance cameras in smart cars capture surveillance images that contribute to this data. Consequently, there is a rise in the number of image files being released and distributed across multiple servers. However, in cases where malicious actors tamper with IoT data, errors may go unresolved, leading to data loss. To address these challenges, Lizhi Xiong proposed a secure and reliable secret image sharing system based on Internet of Things Extended Hamming code (RSIS) [7]. RSIS leverages the distributed architecture of IoT to provide secure information sharing. The hidden images are divided into steganographic images, which can be distributed to multiple servers located outside the IoT network, making it challenging for attackers to detect them. Using RSIS, the data is encoded and embedded within the steganographic images using Hamming code, allowing for a high probability of detecting any potential spoofing attempts.

Now, the demand for IoT applications in current communication networks has been growing. To meet the needs of low-cost and low-power IoT applications, various IoT standardization bodies have proposed different technologies to adapt to this trend. One such technology is Narrowband IoT (NB-IoT), which utilizes Turbo codes with modern error correction coding techniques to achieve high error rate performance close to Shannon's theoretical limits. Turbo codes have demonstrated excellent performance characteristics and are widely applied in many communication standards. However, the computational complexity of Turbo decoders is significantly higher than other modules in NB-IoT receivers, particularly when dealing with higher data rates and low-latency communications. To enhance the flexibility of the uplink system, Mohammed Jajere Adamu [8] and his team proposed an improved approach to Turbo channel decoding. They introduced a frequency domain equalizer (FDE) into their Soft Interference Cancellation (SIC) scheme for uplink NB-IoT systems. This method employs an iterative detection process that exchanges soft decision information between the FDE and Turbo channel decoder to eliminate inter-symbol interference (ISI). Through extensive simulation evaluations of Mean Square Error (MSE) and Block Error Rate (BLER) performance, the researchers have put forward a robust scheme aimed at improving the reliability of user device (UE) data transmission in NB-IoT systems while simultaneously reducing computational complexity. This research offers valuable insights for the development of NB-IoT systems, enabling them to better adapt to the increasing demands of IoT applications.

3.2. 5G scenario

In the new era of information development, people's demand for Mobile Communication has increased significantly, so the 5th Generation Mobile Communication Systems (5G) has been rapidly developed in a short time, and 5G will play an increasingly important role in the future. 5G has higher requirements for information transmission rate, reliability and delay, and channel coding technology is one of the main wireless transmission technologies to meet these needs. In order to receive signals in noisy channels correctly and lossless, new channel coding technologies such as Polar code and LDPC code are gradually proposed, which greatly improves the channel capacity of 5G communication. At the same time, these technologies have also been rigorously demonstrated to achieve the Shannon limit.

Aiming at the serious performance loss of 5G LDPC quantized minimum sum decoding algorithm, Shao proposed a Modified Adapted Min-Sum (MAMS) decoding algorithm [9]. First, based on the situation that the decoding performance is more sensitive to the offset factor in the check node update function when the row degree and column degree of the 5G LDPC check matrix are low, different check node update functions are adopted according to the different values of the row degree and column degree to improve its error correction ability. Then, the third smallest value in the message passed by the variable node to the check node is introduced to adjust the offset factor to reduce the decoding error rate. Finally, by setting the threshold of iteration times, check node update functions of different complexity

DOI: 10.54254/2755-2721/32/20230183

are adopted in the iteration process to cut down the decoding complexity. At the same time, in view of the hardware implementation of 5G LDPC code coders, he also completed the hardware design and implementation of supporting multi-rate compatible coders.

Polarization code was proposed in 2009, due to its excellent performance and huge potential, a large number of researchers invested in it, so in 2016, polarization code was confirmed for the 5th Generation Mobile Communication system. Aiming at the problem of poor error correction performance due to low error detection efficiency of parity-assisted polarization codes (PC-polar codes for short), Zhang proposed a novel coding algorithm for Cyclic Redundancy Check (CRC) code-assisted PC-polar codes [10]. The algorithm uses PC bits and frozen bits with high Hamming weight to replace the information bits with low Hamming weight to optimize the distance spectrum of the polarization code, and combines the five-bit cyclic shift register to optimize the check function of PC code, then adds the CRC code with high error detection efficiency to the PC-polar code, and finally determines the number of the two check codes by the control variable method. At the same time, a Successive Cancellation List (SCL) decoding algorithm based on key sets is proposed to solve the problem of high computational complexity in PC-polar code decoding. In this algorithm, the polarization code is decomposed into 6 sub-polarization codes, and the key set is constructed by selecting the information bits with low Hamming weight from each sub-polarization code, and adding the information bits with low polarization weight value to further improve the key set.

3.3. Satellite communication scenario

As we enter the era of 5G, communication systems are advancing towards high speed, low latency, and superior reliability. Nevertheless, 5G ground base stations are confronted with challenges such as cost and physical conditions, which restrict them from providing communication services to remote mountains, oceans, deserts, and other non-urban areas. Moreover, ground communication networks are susceptible to damage during natural disasters. Low-orbit satellite communication networks, however, cover a vast range, are less impacted by natural conditions, and can still provide services amidst natural disasters. Currently, setting up a low-orbit satellite communication network and integrating it with the terrestrial mobile network offers an effective solution to the issues present in ground communication networks. Satellite communication networks utilize satellites as relays to facilitate information transmission. Given the long communication distance from the terminal, these networks experience significant signal attenuation and large Doppler shifts, leading to low reliability. Channel coding is one strategy to enhance the reliability of the communication system. Polar code, theoretically proven to reach the Shannon capacity limit, boasts excellent performance, thereby establishing it as a channel coding scheme for low-orbit satellite communication.

Taking low-orbit satellite communication as an application scenario, Gao analyzed the encoding and decoding algorithm and construction algorithm of Polar code. He proposed an improved Polarization Weight (PW) construction algorithm for block fading channels [11]. The performance of Polar code was simulated under block fading (BF) channel conditions and compared with two Low-Density Parity-Check codes suggested by the Consultative Committee for Space Data Systems (CCSDS) for current low-orbit satellite communication. Satellite navigation represents another typical application in satellite communications. To further enhance the interference resistance of navigation signals, Turbo code, known for its superior performance, is adopted as the channel coding technology in the new generation satellite navigation system. Among several Turbo decoding algorithms, the logarithmic Maximum A Posteriori (LOG-MAP) decoding algorithm excels in performance and complexity. However, the decoding delay of this algorithm is proportional to the frame length. In the satellite navigation receiver, when the frame length is long, the number of tracking channels is large, and the decoding clock is not fast enough, the LOG-MAP decoding algorithm becomes unsuitable. Xue et al. added a sliding window operation to the LOG-MAP decoding, significantly reducing the decoding delay and the Random Access Memory (RAM) resources required by the decoding module [12].

DOI: 10.54254/2755-2721/32/20230183

4. Conclusion

This treatise engages in a detailed discourse surrounding the prevalent coding techniques and their respective application environments in the realm of channel coding. By elucidating the current status and consequential importance of channel coding, it seeks to provide readers with a comprehensive understanding of the critical role channel coding plays within data communication. The role of channel coding, essentially, is to safeguard data transmission in the face of potential interference and noise disturbances, ensuring high-quality data exchange. This function is pivotal for the smooth operation of any modern digital communication system, from Internet and mobile communications to television broadcasting and more. In scenarios plagued by significant noise, a well-chosen channel coding can even aid in error correction, thus enhancing the reliability of data transmission. To further solidify this understanding, the paper also delves into the typical implementations of various channel codes within specific settings such as the Internet of Things, Fifth-Generation telecommunications technology, and satellite communication. By illustrating these practical examples, it aims to shed light on the feasibility and effectiveness of applying different coding methodologies across diverse scenarios.

In essence, the study of channel coding and its various implementations can prove to be a vital resource for individuals involved in the design and operation of communication systems. By providing this comprehensive exploration, the paper encourages a deeper appreciation of the importance of channel coding, and its vast potential in optimizing modern communication systems.

References

[1] Shannon, C. E. (1948). A mathematical theory of communication. The Bell System Technical Journal, 27(3), 379-423.

[2] Hamming, R. W. (1950). Error detecting and error correcting codes. The Bell System Technical Journal, 29(2), 147-160.

[3] Chien, R. (1964). Cyclic decoding procedures for Bose-Chaudhuri-Hocquenghem codes. IEEE Transactions on Information Theory, 10(4), 357-363.

[4] Berrou, C., Glavieux, A., & Thitimajshima, P. (1993). Near Shannon limit error-correcting coding and decoding: Turbo-codes. 1. In Proceedings of ICC'93-IEEE International Conference on Communications (Vol. 2, pp. 1064-1070). IEEE.

[5] Gallager, R. (1962). Low-density parity-check codes. IRE Transactions on Information Theory, 8(1), 21-28.

[6] Arikan, E. (2009). Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. IEEE Transactions on Information Theory, 55(7), 3051-3073.

[7] Xiong, X. Han, Zhong, C. -N. Yang, & Xiong, N. N. (2023). RSIS: A Secure and Reliable Secret Image Sharing System Based on Extended Hamming Codes in Industrial Internet of Things. IEEE Internet of Things Journal, 10(3), 1933-1945.

[8] Adamu, M. J., Qiang, L., Zakariyya, R. S., et al. (2021). An efficient turbo decoding and frequency domain turbo equalization for lte based narrowband internet of things (nb-iot) systems. Sensors, 21(16), 5351.

[9] Wang, Z., Du, Y., Wei, K., et al. (2022). Vision, application scenarios, and key technology trends for 6G mobile communications. Science China Information Sciences, 65(5), 151301.

[10] Saura, J. R., Palacios-Marqués, D., & Ribeiro-Soriano, D. (2021). Using data mining techniques to explore security issues in smart living environments in Twitter. Computer Communications, 179, 285-295.

[11] Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., et al. (2000). Brain-computer interface technology: a review of the first international meeting. IEEE Transactions on Rehabilitation Engineering, 8(2), 164-173.

[12] Radianti, J., Majchrzak, T. A., Fromm, J., et al. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. Computers & Education, 147, 103778.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230184

Forecasting red wine quality: A comparative examination of machine learning approaches

Bohui Zhan

Art and Science, University of Rochester, Rochester, 14627, United State

bzhan@u.rochester.edu

Abstract. This research explores the forecast of red wine quality utilizing machine learning algorithms, with a particular emphasis on the impact of alcohol content, sulphates, total sulfur dioxide, and citric acid. The original dataset, comprised of Portuguese "Vinho Verde" red wine data from 2009, was bifurcated into binary classes to delineate low-quality (ratings 1-5) and high-quality (ratings 6-10) wines. A heatmap verified the potent correlation between the chosen variables and wine quality, paving the way for their inclusion in our analysis. Four machine learning techniques were employed: Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, and Naive Bayes. Each technique was trained and assessed through resulting metrics and graphical visualizations, with diverse proportions of data assigned for training and testing. Among these techniques, Logistic Regression achieved an accuracy score of 72.08%, while KNN slightly surpassed it with an accuracy rate of 74%. The Decision Tree technique rendered the peak accuracy of 74.7%, while Naive Bayes underperformed with a score of 60.2%. From a comparative viewpoint, the Decision Tree technique exhibited superior performance, positioning it as a viable instrument for future predictions of wine quality. The capacity to predict wine quality carries significant implications for wine production, marketing, customer satisfaction, and quality control. It enables the identification of factors contributing to high-quality wine, optimization of production processes, refinement of marketing strategies, enhancement of customer service, and potential early identification of substandard wines before reaching consumers, thereby safeguarding the brand reputation of wineries.

Keywords: red wine quality, Logistic Regression, decision tree, Naive Bayes, machine learning.

1. Introduction

Among the panoply of wines on offer, red wine stands out not only due to its distinctive flavor profiles but also because of its quality, which can significantly sway consumer preferences and market trends [1]. The quality of wine is a multifaceted attribute, influenced by a myriad of factors including the wine's chemical properties [2]. Gaining an understanding and making predictions about the quality of wine, especially red wine, is an important research pursuit with a wide range of practical implications in the fields of wine production, quality control, and marketing strategy.

In recent years, the deployment of machine learning techniques has broadened into various sectors, encompassing the food and beverage industry, attributed to their ability to forecast outcomes based on input data. Machine learning offers a refined and efficient method to dissect the multitude of factors influencing wine quality, thus enabling the creation of a predictive model for quality evaluation [3]. This

DOI: 10.54254/2755-2721/32/20230184

study ventures into the convergence of machine learning and wine quality prediction. Our goal is to forecast the quality of red wine based on four crucial attributes: alcohol content, sulphates, total sulfur dioxide, and citric acid. These variables were cherry-picked due to their potent correlation with wine quality, as derived from an initial heatmap analysis. In order to achieve our research goal, applied and compared four distinct machine learning models: Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, and Naive Bayes [4]. The performance of each model was assessed, and the most dependable model for predicting wine quality was pinpointed.

2. Data used

The dataset underpinning this study is derived from the Portuguese "Vinho Verde" red wine, produced in 2009. It encompasses 1599 samples, with each sample being defined by 12 distinct attributes.

The original dataset's quality ratings, which range from 1 to 10, were preprocessed for a binary classification approach, recategorizing them into two groups: 0 indicates poor quality (original ratings 1-5), and 1 signifies superior quality (original ratings 6-10). A heatmap analysis of the dataset was executed to discern the attributes most significantly correlated with wine quality [5]. This heatmap provides a color-coded visual representation of the relationships between various wine attributes and wine quality, with the intensity of color reflecting the correlation's strength. This examination unveiled alcohol, sulphates, and citric acid as having a significant correlation with wine quality, as detailed in Figure 1.

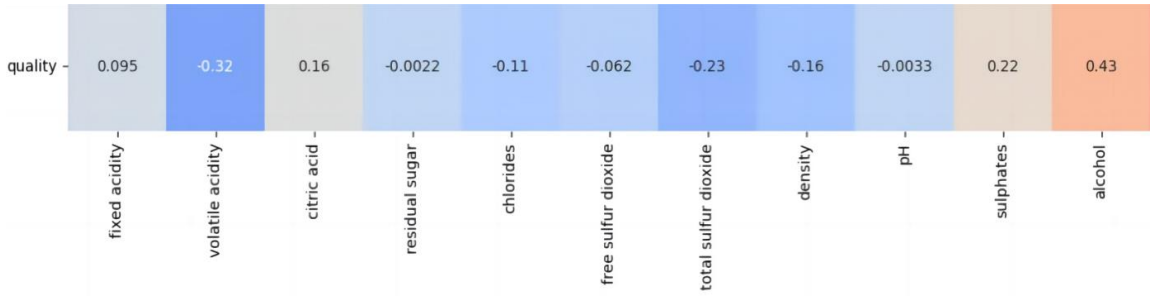


Figure 1. Heatmap of the dataset (Photo/Picture credit: Original).

While total sulfur dioxide did not stand out in the heatmap as one of the strongest correlates, included it in our analysis due to a significant finding from our grouped data. Specifically, the Table 1 displays the mean values of different parameters grouped by wine quality:

Table 1. Average values of wine attributes by quality.

quality	fixed acidity	volatile acidity	citric acid	free sulfur	total sulfur	sulphates	alcohol	...
0	8.14220	0.58950	0.23775	16.56720	54.64516	0.61853	9.92647	...
1	8.47403	0.47414	0.29988	15.27251	39.35204	0.69262	10.8550	...

The observations from this table delineate a clear pattern: wines of superior quality (quality 1) typically exhibit higher mean values for alcohol, sulphates, and citric acid, while presenting a lower mean value for total sulfur dioxide, compared to their inferior quality counterparts (quality 0) [6]. This intriguing contrast with total sulfur dioxide, despite not being listed among the top four correlations on the heatmap, provides compelling grounds for its inclusion as a predictor in the ensuing analysis. As a result, these four parameters—alcohol, sulphates, citric acid, and total sulfur dioxide—were chosen as the main focus of the study, intending to harness the power of machine learning models for predicting wine quality.

DOI: 10.54254/2755-2721/32/20230184

3. Proposed methodology

In the wake of the data preprocessing phase delineated in Section 2, the aim was to predict wine quality based on four attributes: alcohol, sulphates, total sulfur dioxide, and citric acid. To this end, four distinct machine learning methodologies were brought into play.

First off, Logistic Regression was put to use, a statistical method predominantly utilized for predicting binary outcomes. Owing to its simplicity and interpretability, this model set the initial benchmark for the analysis. Subsequently, the KNN method, being distance-based, enabled the classification of wine quality in accordance with the similarity of feature vectors. As the third step, a Decision Tree algorithm was set in motion. This model presented a structured, hierarchical approach to classification hinged on feature thresholds. It provided a visually enriched representation of the decision-making process. Lastly, the Naive Bayes algorithm was put into action. Despite its premise of independence among predictors, it often delivers remarkable performance, thus contributing a unique perspective to the other methods. These methodologies, selected for their distinctive properties, made the results more comprehensive and robust. Collectively, they facilitated a comparative analysis to ascertain the most effective model for predicting red wine quality. The particulars of the application of each methodology will be elucidated in the subsequent sub-sections.

3.1. Logistic regression

3.1.1. Mathematical principle

$$log\left(\frac{p}{1-p}\right) = \beta 0 + \beta 1X1 + \beta 2X2 + \cdots \beta k\, Xk + \epsilon \quad (1)$$

where, p is the probability of the dependent variable (e.g., wine quality) being 1 (e.g., high quality), and Xi are predictors (independent variables such as alcohol, sulphates, total sulfur dioxide, and citric acid). As shown in Figure 2.

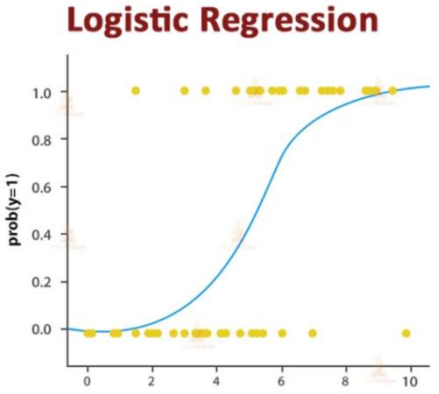


Figure 2. Logistic regression (Photo/Picture credit: Original).

3.1.2. Logistic regression implementation. This study utilizes a dataset founded on the 2009 Portuguese "Vinho Verde" red wine, encompassing 1599 samples. Each sample is defined by 12 unique attributes. The original dataset had quality ratings spanning from 1 to 10. To facilitate binary classification, these ratings underwent preprocessing, where they were grouped into two categories: 0 signifying subpar quality (original ratings 1-5), and 1 denoting superior quality (original ratings 6-10). A heatmap analysis proved instrumental in identifying the attributes most closely linked with wine quality [7]. This visual tool delineated the interrelationships between different wine attributes and the quality of the wine, with color intensity reflecting the strength of each correlation [8]. This analysis led to the identification of

DOI: 10.54254/2755-2721/32/20230184

alcohol, sulphates, and citric acid as attributes displaying significant correlations with wine quality, as depicted in Figure 3.

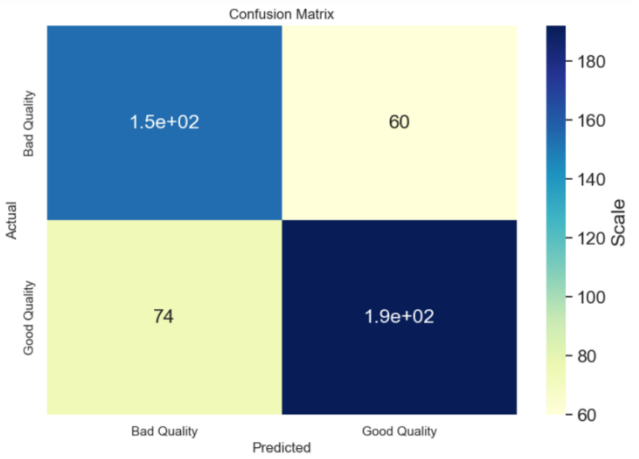


Figure 3. Confusion matrix of logic regression (Photo/Picture credit: Original).

3.2. K-Nearest neighbors

The K-Nearest Neighbors algorithm is a type of instance-based learning method widely used in machine learning. The algorithm predicts the classification of a new observation based on the classifications of its 'K' nearest neighbors in the feature space [9].

3.2.1. Mathematical principle. A KNN model can be visually and conceptually represented as :$Classification\ (Z) = majority\ class\ among\ k\ nearest\ neighbors\ of\ Z\ in\ \{X1, X2, \ldots, Xn\}$

Here, Z is the new instance to be classified, and X1, X2, ..., Xn are instances in the dataset, each having a class label. The 'k' nearest neighbors are selected based on a distance metric, such as Euclidean distance, in the multidimensional feature space. The classification for Z is then determined by the majority class among these 'k' nearest neighbors.

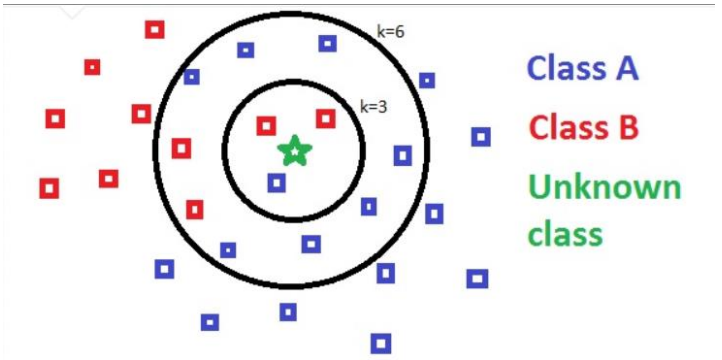


Figure 4. K-Nearest neighbors (k=3 and k=6) (Photo/Picture credit: Original).

K-Nearest Neighbors Implementation: The K-Nearest Neighbors algorithm served as the second machine learning technique utilized for predicting wine quality. The process commenced with the division of data into training and testing sets, employing Scikit-learn's train_test_split function. An 80/20 ratio was adopted for this split, designating the larger portion for training and the remaining for testing. Given KNN's dependence on the distances between feature vectors, it was essential to scale the feature values prior to KNN application. To this end, Scikit-learn's StandardScaler class was employed, performing the required standardization. The KNN algorithm was actualized using Scikit-learn's KNeighborsClassifier class, assigning the number of neighbors as 5 and employing Euclidean distance

DOI: 10.54254/2755-2721/32/20230184

as the metric. Subsequent to training the model on the standardized training data, predictions were made on the standardized test data. Performance evaluation paralleled the method used for the logistic regression model: the calculation of an accuracy score and generation of a classification report. This report incorporated precision, recall, F1-score, and support for both wine quality categories (high and low). A confusion matrix was also formed and visualized to provide a more nuanced understanding of the model's performance. As shown in Figure 4.

The results depicted a commendable performance of the KNN model on the test data, although the accuracy score was slightly below that of the logistic regression model. The classification report suggested that the model exhibited satisfactory performance for both wine quality categories, despite occasional instances of misclassification as discerned from the confusion matrix. As shown in Figure 5.

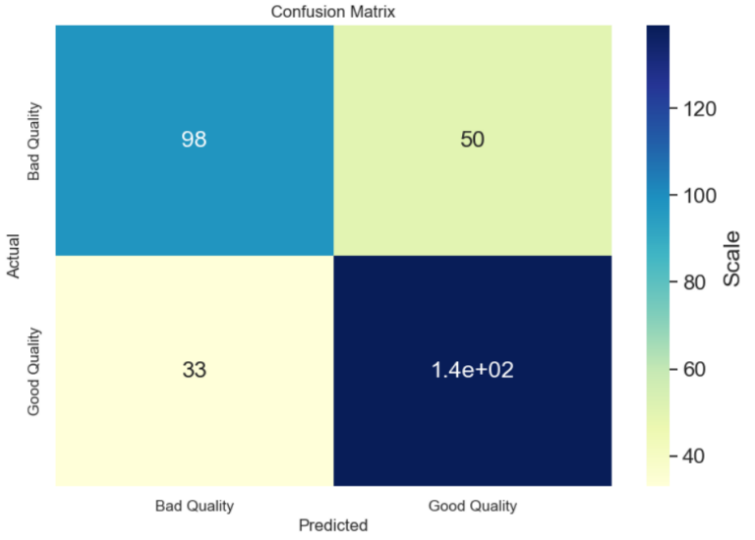


Figure 5. Confusion matrix of KNN (Photo/Picture credit: Original).

To provide a more intuitive understanding of the KNN model's operation, created a 3D scatter plot of the scaled training data, with the colors representing the actual wine quality labels. This visual representation helps to show how the KNN model uses the 'closeness' of data points in the feature space to predict the wine quality. As shown in Figure 6.

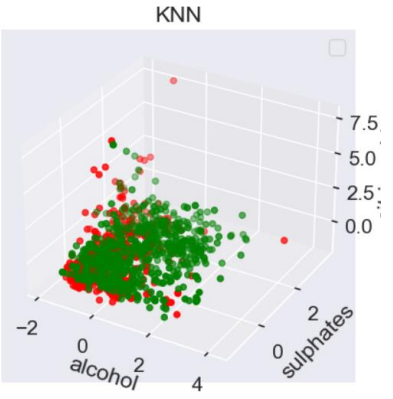


Figure 6. 3D scatter plot of KNN (Photo/Picture credit: Original).

3.3. Decision tree

A decision tree stands as a potent predictive model and is counted among the most comprehensible machine learning algorithms. Functioning in a hierarchical manner, it segmentizes the data into subsets

DOI: 10.54254/2755-2721/32/20230184

based on varying attribute values, essentially driving decisions via specific rules and conditions [10]. Through an iterative procedure, this model continues to establish test conditions for additional attributes, further bifurcating the data. Each decision gives rise to a new branch in the tree. This process perseveres until a predefined stopping criterion is achieved, such as the exhaustion of attributes for future partitioning, or when the maximum tree depth is attained [11].

*3.3.1. Decision tree implementation.*The DecisionTreeClassifier class from Scikit-learn was employed to construct the Decision Tree model. The 'entropy' criterion was selected as the measure of a split's quality, serving as an indicator of the level of uncertainty or disorder within a dataset. Essentially, entropy is computed as (3).

$$Entropy = -\Sigma\,[\,p(x)log2\,p\,(x)] \tag{3}$$

where p(x) is the proportion of the observations that belong to each class.

By electing entropy as the criterion, our algorithm attempts to maximize the information gain at each split. This essentially means that the model prefers the splits that yield the largest information gain. Therefore, high entropy denotes a high degree of disorder and low information gain, while low entropy signifies a well-ordered set and high information gain. As shown in Figure 7.

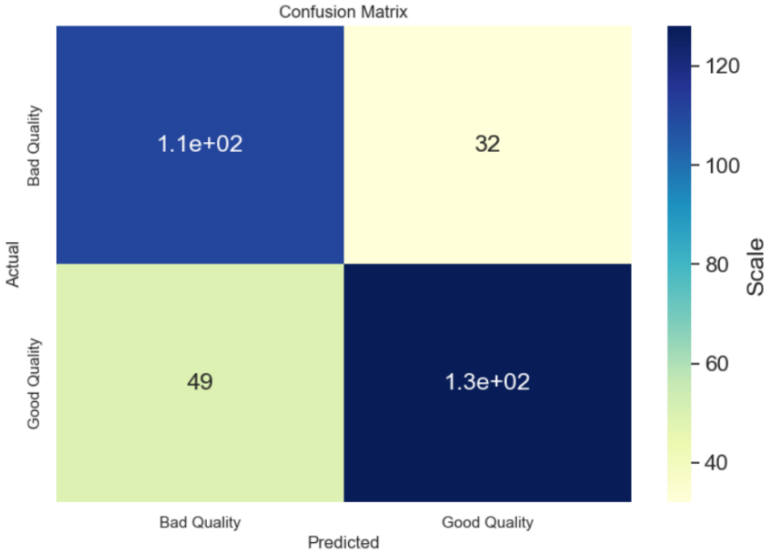


Figure 7. Confusion matrix of decision tree (Photo/Picture credit: Original).

The accuracy score of the Decision Tree model was relatively good. However, the confusion matrix revealed that there were some instances of misclassification. This suggests that the model may be overfitting the training data, which is a common problem with Decision Trees. One of the main advantages of Decision Trees is their interpretability. To leverage this, displayed the trained Decision Tree visually as a plot and as text. The plot provides a clear visualization of the decision-making process of the model, showing the conditions for each split and the distribution of classes in each leaf node. As shown in Figure 8.

DOI: 10.54254/2755-2721/32/20230184

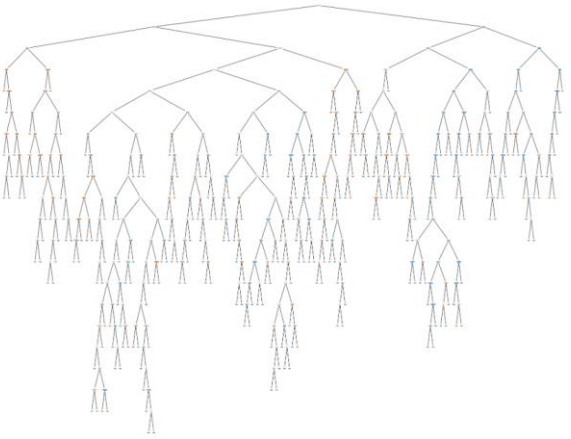


Figure 8. Decision tree (Photo/Picture credit: Original).

3.4. Naive Bayes

This algorithm calculates the probability of each category of the dependent variable given independent variables, and the prediction is made based on which category has the highest probability. Despite its simplicity and strong assumptions, Naive Bayes can be extremely effective, fast, and accurate in many scenarios [12].

3.4.1. Naive Bayes implementation. The Multinomial Naive Bayes variant was utilized, due to its effectiveness with feature vectors that are multinomially distributed.

The Naive Bayes model was initialized and trained on the designated training set utilizing the fit method. Subsequently, the model generated predictions on the test set. The model's accuracy was determined by juxtaposing its predictions on the test set against the actual wine quality classifications. This procedure culminated in the Naive Bayes model achieving an accuracy score of 60.2%, marking the lowest performance among all the evaluated models. Despite its relative simplicity, the Naive Bayes classifier did not match the performance of other models in forecasting wine quality. This may be attributed to the model's assumption of predictor independence, a condition that might not be met by the variables in this dataset. Even though the Naive Bayes model was overshadowed by other models in terms of performance, it retains value in establishing a baseline comparison for more intricate models and could potentially deliver superior results with a different set of features or hyperparameters.

Notwithstanding its lower accuracy, the Naive Bayes technique exhibited the potential of probabilistic classification models in forecasting wine quality predicated on the selected physicochemical properties. Prospective studies might delve into other variants of Naive Bayes or manipulate its parameters to bolster its predictive accuracy.

4. Conclusion

This investigation implemented four distinctive machine learning models, each yielding different insights and levels of success in predicting wine quality. The Logistic Regression model generated an accuracy of 72.08%, while KNN displayed a marginally superior outcome with an accuracy of 74.06%. Remarkably, the Decision Tree model outperformed the others, achieving the peak accuracy of 74.7%, while the Naive Bayes model fell short comparatively with a score of 60.2%.

Despite the varied accuracies, the findings suggest that machine learning can indeed function as a powerful tool within the wine industry. The ability to predict wine quality using these models has the potential to significantly optimize production processes, empowering winemakers to concentrate on the most influential variables for wine quality. This predictive capacity could also enable retailers to hone their marketing strategies and provide more precise information to consumers, culminating in enhanced customer satisfaction. Additionally, the prospect of identifying potentially subpar wines before they hit the market is a notable advantage for quality control, assisting in maintaining the high standard of

DOI: 10.54254/2755-2721/32/20230184

wineries and preserving their brand reputation. It is crucial to recognize, however, that this investigation, as with any study, has its limitations. For example, it adopted a binary rating system for wine quality and examined only a limited set of variables. Future research could consider a more nuanced rating system for wine quality or investigate a broader range of variables impacting wine quality. Furthermore, the efficacy of other machine learning models, or even combinations thereof, could be tested for improved predictive accuracy.

References

[1] Feher J, Lenguello G and Lugasi A. The cultural history of wine—Theoretical background to wine therapy. Cent. Eur. J. Med. 2007,2, 379–391.

[2] Sirén H, Sirén K and Sirén J. Evaluation of organic and inorganic compounds levels of red wines processed from Pinot Noir grapes.Anal. Chem. Res. 2015, 3, 26–36.

[3] Gupta Y. Selection of important features and predicting wine quality using machine learning techniques. Procedia Computer Science. 2018; 125:305-312.

[4] Grewal P, Sharma P, Rathee A and Gupta S. COMPARATIVE ANALYSIS OF MACHINE LEARNING MODELS. EPRA International Journal of Research and Development (IJRD). 2022;7(6):62–75.

[5] Reimann C, Filzmoser P, Hron, K, Kynčlová, P and Garrett, R G. A new method for correlation analysis of compositional (environmental) data – a worked example. Sci. Total Environ. 2017, 607–608, 965-971.

[6] Al-Ghamdi A S, Using logistic regression to estimate the influence of accident factors on accident severity. Accid. Anal. Prev. 2002, 34(6), 729-741.

[7] Bisong E. Introduction to Scikit-learn. In: Building Machine Learning and Deep Learning Models on Google Cloud Platform. Apress, Berkeley, CA, 2019.

[8] Susmaga R. Confusion Matrix Visualization. In: Kłopotek, M.A., Wierzchoń, S.T., Trojanowski, K. (eds) Intelligent Information Processing and Web Mining. Advances in Soft Computing, vol 25. Springer, Berlin, Heidelberg, 2004

[9] Guo G, Wang H, Bell D, Bi Y and Greer K. KNN Model-Based Approach in Classification. In: Meersman, R., Tari, Z., Schmidt, D.C. (eds) On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE. OTM 2003. Lecture Notes in Computer Science, vol 2888. Springer, Berlin, Heidelberg, 2003.

[10] Song Y Y, Lu Y. Decision tree methods: applications for classification and prediction. Shanghai Arch Psychiatry, vol 27, no. 2, 2015, pp.130-5.

[11] Apté C, Weiss S. Data mining with decision trees and decision rules. Future Generation Computer Systems, vol 13, issues 2–3, 1997, pp. 197-210.

[12] Rish I. An empirical study of the naive Bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence, vol. 3, no. 22, 2001, pp. 41-46.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230185

Comprehensive evaluation and enhancement of Reed-Solomon codes in RAID6 data storage systems

Lijun Wei

International College, Wuhan University of Science and Technology, Wuhan, 430081, China.

1910700213@mail.sit.edu.cn

Abstract. This paper provides an in-depth examination and optimization of Reed-Solomon codes within the context of Redundant Array of Independent Disks 6 (RAID6) data storage configurations. With the swift advancement of digital technology, the need for secure and efficient data storage methods has sharply escalated. This study delves into the application of Reed-Solomon codes, which are acclaimed for their unparalleled ability to rectify multiple errors, and their crucial role in maintaining RAID6 system operation even under multiple disk failures. The intricacies of Reed-Solomon codes are scrutinized, and the system's resilience in various disk failure scenarios is evaluated, contrasting the performance of Reed-Solomon codes with other error correction methodologies like Hamming codes, Bose-Chaudhuri-Hocquenghem codes, and Low-Density Parity-Check codes. Rigorous testing underscores the robust error correction capabilities of Reed-Solomon encoding in an array of scenarios, affirming its efficacy. Additionally, potential enhancement strategies for the implementation of these codes are proposed, encompassing refinements to the algorithm, the adoption of efficient data structures, the utilization of parallel computing techniques, and hardware acceleration approaches. The findings underscore the balance that Reed-Solomon codes strike between robust error correction and manageable computational complexity, positioning them as the optimal selection for RAID6 systems.

keywords: Reed-Solomon codes, RAID6 data storage, system optimization.

1. Introduction

The exponential growth of digital technology has escalated the need for efficient and secure methods of data storage [1]. RAID6, a variant of Redundant Array of Independent Disks, stands as a cornerstone in contemporary data storage technology, primarily attributable to its resilience to faults and high availability. The incorporation of Reed-Solomon (RS) codes, renowned for their exceptional capability to rectify multiple errors, has been shown to be effective in RAID6 data storage configurations [2].

This study delves into the utilization of Reed-Solomon codes in Redundant Array of Independent Disks 6 (RAID6) and their indispensable role in ensuring system functionality, even in the event of multiple disk failures. Initially, a succinct overview of RAID6 and Reed-Solomon codes is presented, underscoring the significance of fault tolerance and high availability in data storage. Following this, a detailed exposition of the (N = K + 2, K) configuration and the advantages that Reed-Solomon codes

DOI: 10.54254/2755-2721/32/20230185

confer upon RAID6 is undertaken. Ultimately, the deployment of an encoder and decoder pair within the system is elucidated [3].

2. Theoretical background

Reed-Solomon codes serve as robust error-correcting mechanisms, proficient in detecting and rectifying multiple symbol inaccuracies. These codes are renowned for their exceptional error-correction capacity, making them widely applicable in various fields, including data storage, digital communication, and broadcasting [4]. The effectiveness and functionality of these codes are determined by two key parameters: k, which signifies the number of original data symbols, and m, embodying the number of parity symbols appended for detecting and rectifying errors [5]. In the realm of data storage technologies, RAID6 stands as a superior technology that employs disk striping coupled with double parity. This technology ensures the incorporation of two parity blocks on every disk in the array, thereby elevating the system's resilience against data losses. Its sophisticated design allows it to withstand two simultaneous disk failures, without compromising the integrity of the stored data.

As we delve deeper into the concept, it's vital to comprehend that RAID6's enhanced resilience is heavily reliant on the proficient usage of Reed-Solomon codes. It's this intricate intertwining of the two technologies that fortifies the system against potential data losses, ensuring optimal data protection. Therefore, the interplay between Reed-Solomon codes and RAID6 isn't just complementary; it's a strategic alliance that assures optimal data protection in an ever-evolving digital landscape.

3. Application of Reed-Solomon codes in RAID6

3.1. Explanation of (N = K + 2, K) configuration

In RAID6, the specific implementation of RS codes is denoted as (N = K + 2, K), where the array comprises K data disks and 2 parity disks. N, thus, represents the total number of disks in the array. This configuration ensures that the system remains functional even when two disks fail simultaneously.

3.2. Benefits of Reed-Solomon codes for RAID6

The use of RS codes in RAID6 brings several benefits. The main advantage is the ability of these codes to correct up to two erasures (disk failures) at any location within the data set, thereby making the system highly resilient to data losses. Moreover, this feature renders RS codes an excellent choice for RAID6 storage systems that necessitate high availability and fault tolerance.

3.3. Detailed implementation of encoder and decoder pair

The implementation of an encoder and decoder pair for such a system involves handling any input message and correcting all correctable erasure patterns [6]. The encoder multiplies a k-symbol data vector by a generator matrix to produce an (N = K + 2, K) code word [7]. In case of disk failures, the decoder computes the inverse of this process to recover the original data.

However, it is worth noting that the (N = K + 2, K) RS code structure comes at the cost of storage space (equivalent to 2 disks) and increased computational overhead for parity calculation [8]. The complexity of the system and the number of disks needed for each case of failure will be further examined in subsequent sections of this paper [9].

4. Analysis of system complexity and disk access

4.1. Determining the number of disks needed in various failure scenarios

In the context of RAID6, the number of disks accessed during data recovery largely depends on the extent of disk failure. In the event of a single disk failure, all remaining disks, including the parity disk, need to be accessed to recover the lost data. Essentially, the lost data can be calculated from the data on all the other disks and the data on the first parity disk [10].

DOI: 10.54254/2755-2721/32/20230185

If there is a simultaneous failure of two disks, the recovery process remains the same as for a single disk failure. All remaining disks need to be accessed to recover the lost data. The data from all other disks and the data from the two parity disks are used to solve a system of linear equations and recover the two lost data pieces.

4.2. Examination of system complexity with the use of Reed-Solomon codes

The performance of a system utilizing Reed-Solomon encoding and decoding is significantly influenced by the complexity of these processes. Specifically, the encoding complexity of Reed-Solomon codes is typically O(n^2), where n represents the length of the encoded byte string. This complexity originates from the computations required to generate parity disks from the data disks.

On the other hand, the decoding process is even more complex, exhibiting an O(n^3) complexity. This heightened complexity is attributed to the need to solve a system of linear equations to recover lost data, a task involving the computation process of retrieving lost data from the available disks and parity information.

Therefore, while Reed-Solomon codes offer a robust mechanism for error correction, it's imperative to consider that their implementation might significantly affect the performance of the system due to the associated computational complexity. This factor needs to be carefully evaluated when deciding to use Reed-Solomon codes in any application, weighing the benefits of high error correction against potential performance implications.

5. Comparative analysis of alternatives

5.1. Introduction to alternative error correction codes

Error correction codes are instrumental in maintaining data integrity across various applications, including data storage, digital communication, and broadcasting. In addition to Reed-Solomon (RS) codes, other prevalent error correction codes encompass Hamming codes, Bose-Chaudhuri-Hocquenghem (BCH) codes, and Low-Density Parity-Check (LDPC) codes. Hamming codes, known for their simplicity and efficiency, excel at correcting single-bit errors and detecting double-bit errors. However, when it comes to rectifying multiple errors, their capabilities do not match up to some of the more advanced alternatives. Like Reed-Solomon codes, BCH codes operate within a finite field (Galois Field). They are adept at rectifying multiple random error patterns, making them a prime choice for error correction in a variety of data storage and communication systems.

LDPC codes, contrastingly, are linear error correcting codes that offer robust error correction capabilities, especially for long code lengths. These codes are particularly advantageous in applications such as deep space communication, where upholding data integrity in the face of a noisy channel is absolutely critical.

5.2. Comparison of alternatives with Reed-Solomon codes

Compared to other options, Reed-Solomon codes possess the distinct advantage of being able to correct multiple errors. This positions them as a preferred selection for applications that demand robust error correction.

Hamming codes, though easy to implement, fall short in correcting multiple errors simultaneously, a weakness that Reed-Solomon codes effectively circumvent. While BCH codes, much like Reed-Solomon codes, excel at correcting multiple errors, they are hampered by restrictions in code length, which can limit their use in certain applications. Reed-Solomon codes, however, aren't bound by such limitations, showcasing their versatility. LDPC codes provide powerful error correction capabilities, but their computational intensity and complexity in implementation can be a downside. Here, Reed-Solomon codes, with their relatively lower computational complexity, emerge as a superior choice for systems where computational resources are scarce.

Within the context of RAID6 storage, Reed-Solomon codes strike an exceptional balance between robust error correction and manageable computational complexity, solidifying them as an optimal choice.

DOI: 10.54254/2755-2721/32/20230185

The subsequent section will delve into the specifics of implementing Reed-Solomon codes in a RAID6 system, emphasizing the construction of an encoder-decoder pair and assessing their performance under varying scenarios.

6. Optimizing the implementation of Reed-Solomon codes

A comprehensive exploration of potential optimization techniques to enhance the implementation of Reed-Solomon codes is undertaken. This involves a deep-dive into algorithmic improvements, particularly those targeting the finite field arithmetic operations integral to the encoding and decoding processes. The application of more efficient data structures, capable of streamlining data access and manipulation, significantly reducing execution time, is also investigated. Further consideration is given to parallel computing techniques, employing multiple processors to perform computations concurrently and thus accelerate the encoding and decoding processes. Lastly, hardware acceleration techniques, such as the use of Graphics Processing Units (GPUs) or Field Programmable Gate Arrays (FPGAs), which can further expedite computations, are examined. The ultimate objective of this section is to pinpoint effective strategies for optimizing Reed-Solomon codes, ensuring both robustness and efficiency.

In this exploration of coding theory, focus is placed on two central mechanisms, the Encoder and the Decoder. The task of the encoder is to transmute the input data, such as text or images, into another form, typically a sequence of numbers or binary code. Redundant data is integrated so that if transmission errors occur, this extra information can rectify them. Conversely, the decoder restores the encoded information to its original form. During testing, Reed-Solomon (RS) encoding, an error correction coding capable of rectifying multiple errors, is utilized. The RS encoder elongates the input byte string to incorporate the original message and additional redundant bytes. The redundancy is created through mathematical operations, specifically polynomial operations over a finite field. When alterations occur to the encoded byte string, due to transmission or storage errors, the RS decoder utilizes redundant bytes to correct these errors, recovering the original message. Across diverse scenarios, including short and long messages, multiple errors, and differing RS code parameters, Reed-Solomon encoding and decoding effectively corrects all introduced errors. Whether dealing with a two-letter message like "Hi" or a lengthy sentence, the result remains consistent, demonstrating RS encoding's ability to correct errors in both short and long messages. Even when the number of introduced errors is escalated, or the redundancy bytes adjusted, the original messages are always accurately restored.

Improvements to the code have been made, specifically in the areas of batch optimization. Batch optimization enhances performance by amalgamating multiple messages into a single batch for encoding and decoding operations. Batch processing reduces the number of loops and the overhead of data replication compared to processing each message individually. In batch processing, encoding results and decoding results of batches are generated using list derivation to minimize the number of loops. By optimizing batch processing for different RS encoding parameters, the following results are obtained. The average encoding and decoding time increases slightly with the augmentation of ECC (Error Correction Code) symbols. The average time for encoding and decoding using 5 ECC symbols is 0.006626 seconds, while the average time using 20 ECC symbols is 0.029391 seconds. These findings reveal that batch optimization leads to better performance when processing multiple messages. By amalgamating multiple messages into batches, more encoding and decoding operations can be executed in the same amount of time. Therefore, batch optimization proves to be highly effective for processing numerous messages or improving the efficiency of encoding and decoding.

In conclusion, tests prove that Reed-Solomon encoding is a formidable error-correcting mechanism capable of handling messages of varying lengths and correcting multiple errors. Therefore, it's a highly effective method for maintaining data integrity, finding applications in fields such as data storage and communication.

7. Conclusion

In conclusion, our in-depth analysis underscores the resiliency and efficiency of Reed-Solomon codes for error detection and correction, particularly within RAID6 systems. The capacity to seamlessly handle

DOI: 10.54254/2755-2721/32/20230185

multiple disk failures reinforces the dependability and accessibility of data that these codes facilitate. Our practical application of the encoder-decoder pair further confirms the operational functionality of Reed-Solomon codes across a spectrum of message lengths and error scenarios. However, it's critical to acknowledge the inherent intricacy of Reed-Solomon codes and the potential challenges it brings, most notably the high volume of disk access required during failure recovery. We've put forward potential optimization strategies that could greatly enhance the implementation of these codes. Concentrating on elements such as algorithmic enhancements, streamlined data structures, parallel computing techniques, and hardware acceleration methods, we're optimistic that the performance of Reed-Solomon codes can be significantly amplified. That said, alternative approaches, including mirroring, RAID5, erasure codes, and Local Reconstruction Codes, offer varying trade-offs regarding redundancy, storage efficiency, and recovery speed. It's paramount to consider these alternatives in alignment with the specific demands and workloads of the concerned system. Looking ahead, it would be a rewarding endeavor to delve more deeply into the optimization strategies discussed, quantify their impacts, and potentially weave them into the Reed-Solomon encoding and decoding process. Additionally, a comparative analysis of Reed-Solomon codes and their alternatives under various system conditions could yield invaluable insights for identifying the most effective data protection and fault tolerance mechanisms.

References

[1] Drucker, N., Gueron, S., & Krasnov, V. (2018). The comeback of Reed Solomon codes. 2018 IEEE 25th Symposium on Computer Arithmetic (ARITH), 125–129.

[2] Kadekodi, S., Rashmi, K. V., & Ganger, G. R. (2019). Cluster storage systems gotta have {HeART}: Improving storage efficiency by exploiting disk-reliability heterogeneity. 17th USENIX Conference on File and Storage Technologies (FAST 19), 345–358.

[3] Lee, J.-Y., Kim, M.-H., Raza Shah, S. A., Ahn, S.-U., Yoon, H., & Noh, S.-Y. (2021). Performance evaluations of distributed file systems for scientific big data in FUSE environment. Electronics, 10(12), 1471.

[4] Lin, S.-J., Alloum, A., & Al-Naffouri, T. Y. (2016). Raid-6 Reed-Solomon codes with asymptotically optimal arithmetic complexities. 2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), 1–5.

[5] Mishra, V., & Pateriya, R. K. (2016). Efficient data administration with reed-Solomon code. International Journal of Scientific Research and Management (IJSRM), 4(12).

[6] Tsung, C.-K., Yang, C.-T., Ranjan, R., Chen, Y.-L., & Ou, J.-H. (2021). Performance evaluation of the vSAN application: A case study on the 3D and AI virtual application cloud service. Human-Centric Computing and Information Sciences, 11.

[7] Xie, P., Yuan, Z., & Hu, Y. (2023). Nscale: An efficient RAID-6 online scaling via optimizing data migration. The Journal of Supercomputing, 79(3), 2383–2403.

[8] Yi, C., Sun, X., Zhang, T., Li, C., Li, Y., & Ding, Z. (2022). Random Interleaving Pattern Identification From Interleaved Reed-Solomon Code Symbols. IEEE Transactions on Communications, 70(8), 5059–5070.

[9] Yuan, Z., You, X., Lv, X., Li, M., & Xie, P. (2021). HDS: Optimizing data migration and parity update to realize RAID-6 scaling for HDP. Cluster Computing, 24(4), 3815–3835.

[10] Zou, L., Hou, H., & Zhou, X. (2022). Systematic MDS Array Codes Correcting a Single Criss-Cross Error with Lower Update Complexity. 2022 IEEE International Conference on Big Data (Big Data), 3242–3249.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230186

Exploring the application and performance of extended hamming code in IoT devices

Liuxu Shen

Computer Science & Technology, Nanjing University of Information Science & Technology, Nanjing, 210044, China

202083420072@nuist.edu.cn

Abstract. This study primarily focuses on the implementation of extended Hamming code within Internet of Things (IoT) devices and examines its impact on device performance, particularly in relation to communication protocols. The research begins by introducing and explaining the essential principles surrounding the extended Hamming code and its system. This introduction is followed by a detailed analysis of its practical application in IoT device communication and the subsequent influence on performance. Additionally, the study explores the potential role of extended Hamming code in strengthening the security measures of IoT devices. Experimental findings indicate that incorporating extended Hamming code can effectively enhance the communication efficiency of IoT devices, ensuring accurate data transmission. It also improves the overall operational efficiency of the devices and fortifies their security framework. Yet, despite these promising outcomes, the real-world application of extended Hamming code presents significant challenges. These hurdles highlight the need for continued research and exploration to maximize the potential of the extended Hamming code in the IoT domain. The study concludes with an optimistic outlook, encouraging ongoing investigation and innovation to further optimize the benefits of this code and drive advancements in IoT technology.

Keyword: internet of things, extended hamming code, communication protocol, device performance, device security.

1. Introduction

The evolution of Internet of Things (IoT) technology is dramatically reshaping our daily lives. IoT devices, as essential elements of the IoT landscape, communicate using a variety of protocols to enable smart interaction between devices [1]. However, interruptions in the communication process can significantly compromise the efficiency and accuracy of data in IoT devices. The traditional Hamming code, a well-known tool for error detection and correction, is widely employed in various communication systems. Still, due to its limitations in certain demanding application scenarios, a more advanced solution—namely, the extended Hamming code—is required to meet stricter performance demands [2]. This paper delves into the implementation of the extended Hamming code within IoT devices, focusing on its impact on device communication protocols. The extended Hamming code's role in enhancing the communication efficiency of IoT devices, guaranteeing precise data transmission, and improving overall device performance is scrutinized. Additionally, the potential of the extended Hamming code to bolster the security measures of IoT devices is explored.

DOI: 10.54254/2755-2721/32/20230186

2. Related theories

2.1. Extended hamming code definition

Hamming code, a highly regarded method of error detection and correction, uses three check bits to encode four data bits, exemplified in the (7,4) Hamming code [3]. These check bits are strategically positioned at powers of two, and are generated using XOR operations, ensuring full positional coverage for error detection and correction [4].

Nevertheless, traditional Hamming code can stumble when confronted with multi-bit errors, like two-bit errors. This is because the error checking code can overlap with the one produced during a single-bit error, making it difficult to discern whether it's a single or double-bit error [5]. To combat this, an enhanced Hamming code has been proposed. This version appends an overall check bit, referred to as 'Pa', to the conventional Hamming code, and introduces an overall check code, named 'Ga'. These are both derived from specific XOR operations [6]. The enhanced Hamming code has the capacity to both detect and correct single-bit errors, as well as detect double-bit errors [7]. By analyzing the values of 'Pa' and 'Ga', it can distinguish between error types, thus enabling appropriate corrections. Therefore, this enhanced Hamming code holds significant potential for application within the communication protocols of IoT devices. As shown in Figure 1.

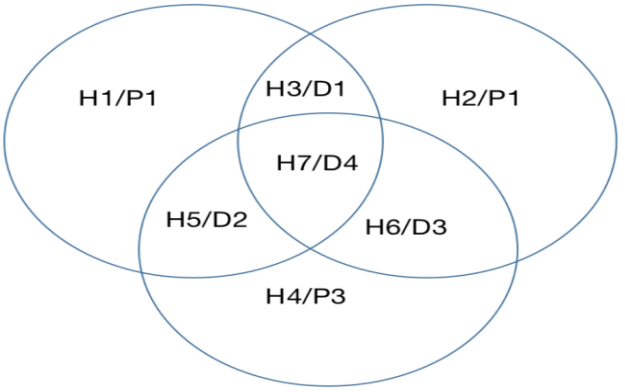


Figure 1. Schematic diagram of the intersection of each position of the hamming code (photo/picture credit: original).

$$G_1 = P_1 \oplus D_1 \oplus D_2 \oplus D_4 \quad (1)$$

$$G_2 = P_2 \oplus D_1 \oplus D_3 \oplus D_4 \quad (2)$$

$$G_3 = P_3 \oplus D_2 \oplus D_3 \oplus D_4 \quad (3)$$

2.2. Expand hamming code code system

In the context of deep space communication, the utilization of the extended Hamming code system is principally showcased through its robust capabilities of error detection and correction. Deep space communication is characterized by extreme distances, substantial signal degradation, and significant environmental noise, leading to a transmission error rate significantly higher than in typical communication scenarios [8]. This necessitates the adoption of a robust coding system to ensure communication reliability, a role fulfilled by the extended Hamming code.

At the transmission end of deep space communication, original data is initially encoded into the extended Hamming code. Each data bit generates corresponding check bits, and an additional parity check bit, 'Pa', is also computed, representing the XOR operation of all data bits and check bits. Furthermore, an overall check code, 'Ga', is calculated as the XOR operation of 'Pa', all data bits, and

DOI: 10.54254/2755-2721/32/20230186

the inverse of all check bits [9]. This encoding procedure ensures that, even in the event of transmission errors, the original data can be reconstructed using the error detection and correction mechanism.

Upon receiving the extended Hamming code, the receiver first checks the overall code 'Ga' to identify if any error has occurred. Subsequently, based on the parity check bit 'Pa' and the error detection code, the type of error is determined, located, and corrected. This procedure guarantees the data's integrity and accuracy, even under conditions of high error rate inherent in deep space communication [10]. Furthermore, in comparison with other coding techniques, the extended Hamming code more accurately locates and distinguishes errors, particularly in handling multi-bit errors resulting from the complex environment of deep space communication. This provides a more reliable error correction mechanism, conferring a substantial advantage in high-demand applications such as deep space communication. Thus, the use of the extended Hamming code in deep space communication is of paramount importance.

3. System analysis and application research

3.1. Implementation of extended hamming code code in IoT

In the communication of IoT devices, Bluetooth, Wi-Fi, and LoRa are three commonly used protocols, and extended Hamming code plays a crucial role in these protocols.

Bluetooth, as a short-range wireless communication technology, is mainly used for small-scale data transmission between devices. Due to its limited working distance, signals may be interfered with or blocked, leading to data loss or errors. In this case, the extended Hamming code becomes especially important. At the sending end, the original data is converted into a binary data stream, and then according to the encoding rules of the extended Hamming code, corresponding check bits are generated for each data bit, and the overall check bit 'Pa' and the overall check code 'Ga' are calculated to form the extended Hamming code. At the receiving end, the device will perform error detection on the received extended Hamming code, and locate the error bit according to the error detection code, and make the corresponding error correction, thereby ensuring the integrity and accuracy of data even in the event of interference or obstruction during data transmission.

Wi-Fi is a common wireless communication protocol, suitable for high-speed data transmission between devices. Due to the long communication distance and fast data transmission speed of Wi-Fi, data transmission errors may occur. This requires the use of extended Hamming code for encoding and decoding, to effectively detect and correct errors in data, ensuring the reliability of high-speed data transmission. Similarly, Hamming code verification processing is performed at both the sending and receiving ends. This way, even if errors occur in high-speed data transmission, the integrity and accuracy of data can be ensured.

LoRa is a long-range low-power wireless communication technology, mainly used for the construction of wide area networks. Due to its long working distance and low data transmission rate, LoRa is more susceptible to environmental noise, leading to data errors. In this environment, the error detection and correction capability of the extended Hamming code becomes particularly important. At the sending end, the device converts the original data into a binary data stream, and then generates the extended Hamming code according to the encoding rules of the extended Hamming code. At the receiving end, the device decodes the received extended Hamming code. If errors exist, it locates the error bit according to the error detection code and makes corresponding error correction. This way, even if errors occur in long-distance, low-speed data transmission, the integrity and accuracy of data can be ensured.

3.2. IoT device communication

IoT device communication faces numerous challenges, including noise interference, electromagnetic reflections, Doppler effects, and multipath propagation. To address these hurdles, time-varying encryption algorithms and extended Hamming accumulation codes have been introduced. The time-varying encryption algorithm is dynamic and can automatically adjust the encryption strategy according to environmental changes, thereby enhancing communication security. Simultaneously, the extended

DOI: 10.54254/2755-2721/32/20230186

Hamming accumulation code, an effective error correction code, can detect, locate, and correct errors, considerably improving communication reliability. This method is particularly beneficial for audio signal transmission, where it can drastically reduce the bit error rate. Moreover, the inclusion of Irregular Repeat Accumulation (IRA) codes, efficient Forward Error Correction (FEC) codes, provides superior performance to turbo codes while maintaining the same coding complexity..

3.3. IoT device performance

Enhancing IoT device performance requires the adoption of multiple strategies. Artificial Neural Networks (ANN) are utilized for Hamming code decoding due to their potent nonlinear fitting abilities and parallel processing capabilities. This approach significantly accelerates the decoding process and further improves decoding accuracy thanks to ANN's adaptive learning ability. Additionally, extended Hamming accumulation codes (EHA) and nonsystematic irregular repeat accumulation codes (IRA) not only enhance communication reliability but also optimize the device's energy consumption. The performance of IRA codes surpasses turbo codes with the same coding complexity, implying higher communication performance at the same energy cost can be achieved.

3.4. IoT security

To strengthen the security of IoT devices, a time-varying encryption algorithm is employed to tackle dynamically changing threats. This algorithm can adjust the encryption strategy based on environmental changes, thereby improving communication security. Simultaneously, the use of extended Hamming accumulation codes and nonsystematic irregular repeat accumulation codes (IRA) not only enhances communication reliability but also fortifies the system's security. These codes can detect and correct errors during the transmission process, thus preventing malicious attackers from exploiting these errors to attack the system. Furthermore, the use of Artificial Neural Networks for decoding, thanks to ANN's adaptive learning ability, can improve the accuracy of decoding, thereby further enhancing the system's security.

4. Experimental methods

In this study, a series of experiments were designed and conducted to test and validate the performance of extended Hamming codes and artificial neural networks in IoT devices. Initially, a simulation of an IoT communication environment was established, which included devices for simulating potential interferences such as noise, electromagnetic reflection, Doppler effects, and multipath propagation. Within this environment, the impact of integrating extended Hamming codes on the communication performance of IoT devices was observed.

The experimental method for extended Hamming codes mainly includes the following steps: Construction of the generator matrix: Firstly, a generator matrix was built, which is used to multiply the input message bits to get a codeword. The generator matrix is composed of a kxk identity matrix and a kxr parity-check matrix, where k is the number of message bits and r is the number of parity-check bits. Design of the decoder: The input to the decoder might be a received codeword with or without errors. In order to detect and correct errors, a parity-check matrix was utilized. The parity-check matrix is made up of the transpose of the parity-check matrix and an identity matrix. Design of the extended Hamming codes: The proposed extended Hamming code design includes an encoder and a decoder for reliable data transmission. With this extended Hamming code technique, single bit and double adjacent bit errors can be identified and corrected. In the encoding part, the parity-check bits are calculated and appended with the message bits for transmission. At the decoder end, if any errors exist, they are corrected, and the parity-check bits are removed. The uncorrected data is then sent as output. Furthermore, artificial neural networks were also employed for Hamming code decoding, and the performance and efficiency of ANN in the decoding process were studied. The specific experimental methods are as follows: Weight initialization: At the beginning of training, small random values were assigned to the weights.

Calculation of forward responses: Neurons in the input layer receive input patterns. These neurons pass this information to neurons in the hidden layer. The hidden layer neurons calculate the output using

DOI: 10.54254/2755-2721/32/20230186

a nonlinear activation function and the weights from the input to the hidden layer, as well as the input. The output of the hidden layer neurons becomes the input for the output layer neurons. Error backpropagation: The error between the target output signal and the actual output signal was calculated, and for all training inputs, this error should be minimized (possibly zero). Weight update: Based on this error, the weights between the hidden-output layer and the input-hidden layer were updated.

A dedicated testing platform was used, allowing for detailed performance testing of devices under controlled conditions. The platform was utilized to assess the performance changes brought about by the introduction of extended Hamming codes and artificial neural networks, such as data transmission speed, transmission accuracy, and device operation efficiency. The innovation of this method lies in its ability to not only detect and correct single bit errors but also detect and correct double adjacent bit errors. This is extremely important for improving the reliability of data transmission. At the same time, artificial neural networks were used to decode the received data, which is a method of real-time operation, self-organization, and adaptive learning. This method has high efficiency and accuracy when dealing with complex decoding problems.

5. Challenges

Despite promising results obtained from the research, several challenges persist in the practical application of the extended Hamming code. Although theoretically, the extended Hamming code enhances communication reliability, its efficacy could be influenced by factors such as the complexity of the communication environment, performance constraints of the devices, and hardware and software limitations to realize extended Hamming codes.

Furthermore, while artificial neural networks (ANN) have been proven effective for Hamming code decoding, handling of large-scale data and high-speed data transmission may present challenges. The requirement of substantial data and computational resources for ANN training could be problematic in resource-constrained environments.

Lastly, combined use of the extended Hamming code and ANN may give rise to new problems, such as how to effectively integrate these two technologies, and how to balance their impact on system performance and security.

6. Conclusion

This study underscores the immense potential of utilizing extended Hamming codes and Artificial Neural Networks to enhance the communication performance and security of IoT devices. Experimental outcomes suggest that these technologies can substantially augment data transmission speed and precision, elevate the operational efficiency of devices, and bolster device security. However, despite these promising findings, certain obstacles still persist when it comes to real-world applications. These include constraints related to hardware and software when implementing complex extended Hamming codes, and issues associated with managing large-scale data and high-speed data transmission. Consequently, there is an urgent need for future research to delve into how these challenges can be surmounted to maximize the advantages of these technologies. Furthermore, exploring the joint utilization of extended Hamming codes and ANN warrants further investigation. Although each technology brings its own set of benefits to the table, identifying an effective synergy to attain superior performance and heightened security remains an open question. In conclusion, this study introduces fresh perspectives and innovative ideas towards the optimization of IoT devices. It is our hope that these insights will inspire future research to propel advancements in this field and offer practical guidelines for the design and implementation of IoT devices.

References

[1] Xiong L, Han X, Zhong X, et al. RSIS: A secure and reliable secret image sharing system based on extended Hamming codes in industrial Internet of Things[J]. IEEE Internet of Things Journal, 2021, 10(3): 1933-1945.

DOI: 10.54254/2755-2721/32/20230186

[2] Isakov D A, Sokolov A V. McELIECE CRYPTOSYSTEM BASED ON QUATERNARY HAMMING CODES[J]. Informatics & Mathematical Methods in Simulation, 2022, 12(4).
[3] Torres-Alvarado A, Morales-Rosales L A, Algredo-Badillo I, et al. An SHA-3 Hardware Architecture against Failures Based on Hamming Codes and Triple Modular Redundancy[J]. Sensors, 2022, 22(8): 2985.
[4] He Y, Xiao C, Wang S, et al. Smart all-time vision: The battery-free video communication for urban administration and law enforcement[J]. Digital Communications and Networks, 2023.
[5] Cintas-Canto A, Kermani M M, Azarderakhsh R. Error Detection Constructions for ITA Finite Field Inversions Over on FPGA Using CRC and Hamming Codes[J]. IEEE Transactions on Reliability, 2022.
[6] Septien-Hernandez J A, Arellano-Vazquez M, Contreras-Cruz M A, et al. A Comparative study of post-quantum cryptosystems for Internet-of-Things applications[J]. Sensors, 2022, 22(2): 489.
[7] Al Homssi B, Dakic K, Maselli S, et al. IoT network design using open-source LoRa coverage emulator[J]. IEEE access, 2021, 9: 53636-53646.
[8] Nguyen C D, Nguyen P D, Nguyen A T, et al. Performance Evaluation Of Neural Network-Based Channel Detection For STT-MRAM[C]//2021 8th NAFOSTED Conference on Information and Computer Science (NICS). IEEE, 2021: 430-434.
[9] Nguyen T A, Lee J. Improving Bit-Error-Rate Performance Using Modulation Coding Techniques for Spin-Torque Transfer Magnetic Random Access Memory[J]. IEEE Access, 2023, 11: 33005-33013.
[10] Larue G, Dufrene L A, Lampin Q, et al. Neural Belief Propagation Auto-Encoder for Linear Block Code Design[J]. IEEE Transactions on Communications, 2022, 70(11): 7250-7264.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230187

Examination of essential technologies and representative applications in RAID 6

Yuxuan Du

Sydney Smart Technology College, Northeastern University, Qinhuangdao, 066004, China

202019152@stu.neuq.edu.cn

Abstract. The evolution of the Internet of Things (IoT) has significantly intensified the interconnectivity between various entities. The robust advancement of information technology has ushered in a societal upswing while simultaneously triggering an exponential increase in data volumes. Consequently, the efficiency of access to storage systems and the reliability of data are severely challenged. Researchers are actively seeking efficient solutions to these challenges. The RAID storage system, with its commendable access performance, excellent scalability, and relative affordability, has become a preferred choice for the storage servers of numerous enterprises. This paper delves into the workings of RAID 6, erasure codes, and capacity expansion, thereby exploring the feasibility of various capacity expansion strategies. Effectively, a well-designed expansion scheme can mitigate issues related to insufficient storage capacity. Simultaneously, the configuration of the code plays a crucial role in determining the expansion time and consequently influences expansion efficiency. Overall, the information and findings presented in this study contribute to enhancing our understanding and management of storage systems in an increasingly data-intensive era.

Keyword: RAID6, erasure, expanded codes.

1. Introduction

Firstly, modern information technology is progressively assuming a central role in our lives. Its influence seeps into every corner of our existence, becoming increasingly ubiquitous. The rapid advancement of technologies such as artificial intelligence, cloud computing, big data, and 5G has driven exponential data growth, creating an increasingly noticeable bottleneck effect. The traditional computer system, initially built around the CPU and memory, is transitioning towards a memory-centric structure, leading to the evolution of storage systems into relatively independent entities [1].

Secondly, the dramatic surge in network users and the swift expansion of application domains have resulted in an unimaginable volume of data. This has imposed a significant strain on data center storage capacity. Long-term exposure to high loads makes the storage media within storage systems more prone to damage. Beyond the losses incurred from system downtime, the financial and temporal cost of data recovery is high. For many businesses that operate on real-time or near-real-time data, such blows can be severe and potentially disastrous [2]. The RAID memory offers excellent

DOI: 10.54254/2755-2721/32/20230187

accessibility and scalability and is relatively cost-effective, making it the primary choice for many corporate memory servers.

The following provides a brief overview of a common RAID 6 code, known as H-code. The H-code comprises an array of (p-1)×(p+1). H-code encoding includes two types of check chains: the reverse diagonal check chain and the horizontal check chain. The H-code uses the layout of the anti-skew check block, evenly placed along the disk array's diagonal, to enhance partial stripe write performance. The horizontal check chain of the H-code ensures optimal continuous data writing performance, and the horizontal check block possesses a special horizontal check disk, as depicted in Figure 1.

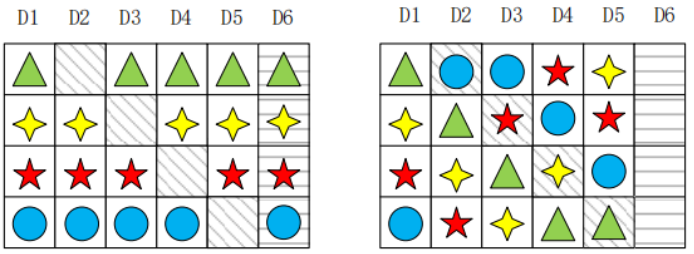


Figure 1. Horizontal check layout and anti-skew check layout of h code (photo/picture credit: original).

2. Relevant theories

2.1. RAID6 technical analysis

The full name of RAID 6 is a separate hard disk array with two independent parity data. To ensure that data is not lost when two hard disks go offline, two different verification algorithms are required [3]. In this way, when two hard disks are disconnected, the data on the disconnected hard disks can be deduced and recovered by combining the equations according to two different calibration algorithms. According to the People's Post and Tele communications (2017), the usual practice is that the first check data is generated by a traditional XOR algorithm, and the other check data is calculated by a reversible function, and the result is generated by XOR again. At present, the most common implementation is to convert the data in Galois Field, and then use the XOR operation to generate a second copy of the verification data. The heart of RAID 6 consists of two copies of the check data to ensure data safety in the event that two hard drives fail simultaneously. RAID6 also achieves better random I/O performance because it inherits the characteristics of striped and distributed parity data storage from RAID 5.

DOI: 10.54254/2755-2721/32/20230187

2.2. Erasure correction code

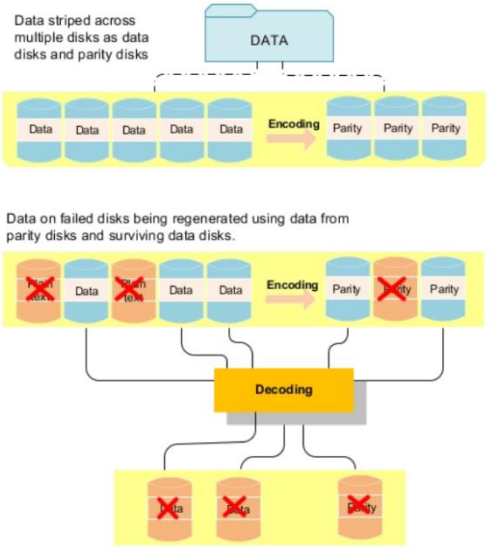


Figure 2. Erasure coding schematic diagram (Photo/Picture credit: Original).

Erasure coding is an advanced error correction technique, as illustrated in Figure 2. It is a fault-tolerant coding technology that can recalculate lost data blocks based on the remaining data blocks and parity blocks. Erasure coding has several advantages that bolster the reliability, availability, and performance of storage as compared to replication. Reliability: Objects are encoded as data and parity blocks and are distributed across multiple nodes and locations [4]. This decentralized method provides a safeguard against site and node failures. Erasure coding improves reliability over replication at a comparable storage cost. Availability: In the context of storage systems, availability refers to the ability to retrieve objects when a storage node fails or becomes inaccessible. Storage Efficiency: For instance, a 10MB object duplicated at two sites would utilize 20MB of disk space (two copies). In contrast, an object encoded between three sites using a 6+3 erasure coding scheme would only use 15MB of disk space. Despite its advantages, erasure coding has some drawbacks. It increases the number of storage areas and locations. In contrast, if only data objects are replicated, a single copy should suffice at one storage location. When erasure coding is implemented across locations distributed in different regions, retrieval delays increase. Extracting fragments of an object encoded with erasure and distributed in remote locations over a WAN connection will take longer than retrieving an object that is replicated and stored locally. Moreover, the utilization of WAN network traffic for retrieval and repair is high when using erasure coding across geographically dispersed sites, especially for objects that are frequently accessed or repaired over a WAN connection. This process also escalates the utilization of computing resources.

2.3. Expansion

To better describe the performance of the RAID expansion solution, we introduce the following basic concepts:

Data: The user's information is stored in a string, which is called data. Block: The most basic unit of storage. It can be classified into data blocks and parity blocks according to the information it carries. Data blocks carry user information, and parity blocks carry redundant information. Stripe: Divides continuous data into data blocks of the same size and stores each piece of data on different disks. Strip: The value can contain only data blocks or parity blocks, or both. Parity Chain: A computing chain consisting of a check block and the data block that generates it. According to the different coding rules of the check chain, it can be divided into: row check chain, oblique check chain and reverse oblique check chain. Encoding: According to certain calculation rules, the generation of redundant data, then

DOI: 10.54254/2755-2721/32/20230187

the process is called encoding. Decoding: The process of making use of both surviving and redundant data to recover when data is lost. Horizontal code: According to the layout rules of erasure codes, the check blocks are stored in separate check disks, then this encoding is called horizontal coding. Vertical code: According to the layout rules of erasure codes, the parity block and data block are stored in the data disk, and this coding is called vertical coding. Scaling up: Adding storage media to RAID to increase storage capacity. Scaling down: Cutting down the disks in the RAID to eliminate damaged disks and reduce energy consumption. Metadata: contains information about data, such as addresses and attributes. It is usually stored in the start location of a disk. Any operation on the data will update the metadata. Read Modify Write: The method of creating a new parity block from the original parity block and the updated data block. Reconstruction Write: An update method that generates new parity blocks from updated data blocks and old data blocks. Degraded Read: A service that continues to provide requests to users despite a disk failure is in degraded read state. Partial Stripe Write: One or more data updates that belong to the same stripe are called partial stripe write [5].

3. Research on 3 erasure code correction technology

3.1. RS coding

The lost data block can be recovered by multiplying (GT) -1 by the codeword vector.

3.2. LDPC coding

The LDPC code is a kind of packet error correcting code with a sparse checking matrix that was proposed by Robert Gallager of MIT in his doctoral thesis in 1963. Its performance is close to the Shannon limit, theoretical analysis and research is easy [6].

The LDPC code is essentially a linear block code that maps the information sequence to the sending sequence via the generation matrix G, i.e. the codeword sequence. For G, there is an exactly equivalent parity matrix H, and all codeword sequences C form the zero space H.

The low-density parity code is an improvement of the parity code.

Because the decoding of parity codes is difficult to apply, there are very suitable decoding schemes using low density. Although LDPC code is not the best code, but because of the existence of a very simple decoding scheme, it makes up for the deficiency of non-optimal code.

3.3. Array coding

Data recovery is completed by XOR operation. The mechanism of single fault tolerance, double fault tolerance and three fault tolerance is used in array coding. Data and redundant data are stored on disks. If some data is faulty, the remaining data is used to restore data. According to whether the array coding meets the optimal storage efficiency, the array coding can be divided into MDS coding and Non-MDS coding. That is, the encoding that meets the optimal storage efficiency is MDS encoding, and vice versa is Non-MDS encoding. In a RAID storage system, the disk that stores data information is called a data disk, and the disk that stores parity information is called a parity disk.

Horizontal parity codes are codes in which data is stored in one area and check information is stored in another area, and the two areas are not on the same disk. Common horizontal codes include RS code, Cauchy RS code, RDP code, Generalized RDP code, etc.

Vertical parity codes are the distribution of data and check information in different areas of the disk. Common vertical coding includes X-Code coding, H-Code coding, HDP Code coding, Short Code coding, D-Code coding, N-Code coding, H-Code coding, P-Code coding, Balance P-Code coding, and balance P-code coding. Hover Code coding, WEAVER Code coding, etc.

DOI: 10.54254/2755-2721/32/20230187

4. Research on capacity expansion technology

4.1. Traditional expansion algorithm

Capacity expansion solutions are divided into two types based on optimization policies: data migration process optimization and minimum data migration.

4.2. Data migration process optimization

This section describes a typical Round-Robin expansion solution. In this expansion solution, including migrating the old disk to the new disk and moving the old disk to the old disk. This solution has a lot of data to migrate, but once scaled out it can respond to a uniform distribution of data and respond very well to user input [7].

Optimized capacity expansion solutions based on data migration such as MDM and ALV are also available.

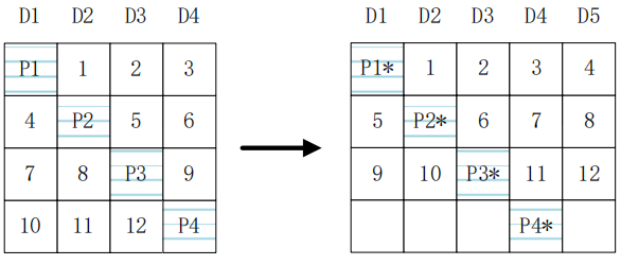


Figure 3. Schematic diagram of comparison before and after expansion (Photo/Picture credit: Original).

The Semi-RR expansion scheme significantly curtails the volume of data requiring migration. As depicted in Figure 3, within the context of the Semi-RR expansion scheme, only data blocks 4, 8, and 12 are transferred from the old disk to the new disk D5. The remaining files are left untouched, thereby minimizing data migration. When considering the least data migration cost for expansion, several methodologies stand out within different encodings, namely, PBM, RS6, Xscale, and FastScale. These techniques aim to streamline the expansion process while ensuring a minimal amount of data is transferred, resulting in significant savings in time and computational resources. The specific application of these methods is represented in Figure 4. Through such strategic management of data migration, we aim to optimize the storage expansion process and improve overall system efficiency.

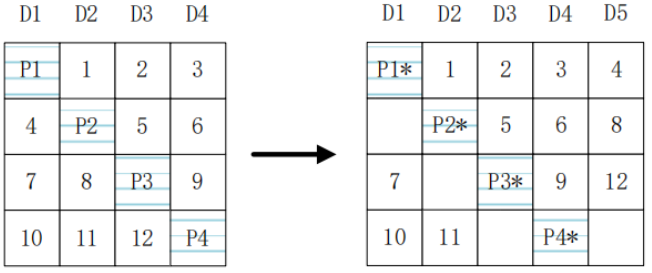


Figure 4. Schematic diagram of comparison before and after expansion (Photo/Picture credit: Original).

4.3. HS6 expansion algorithm

According to the characteristics of H-Code encoding with 2 check chains, HS6 expansion algorithm fully considers the problem of minimum data migration and minimum computing overhead, and carries out its implementation.

Example 1: As shown in Figure 5, before expansion, if p = 5, the RAID-6 storage system has 4 x 6 storage strips [8]. After expansion, the storage strips must be (p-1) x (p +1). Therefore, add disks D5 and D6 forms 6 x 8 storage strips (p = 7).

DOI: 10.54254/2755-2721/32/20230187

D0	D1	D2	D3	D4	D5
0	Q0	1	2	3	P0
4	5	Q1	6	7	P1
8	9	10	Q2	11	P2
12	13	14	15	Q3	P3
16	Q4	17	18	19	P4
20	21	Q5	22	23	P5
24	25	26	Q6	27	P6
28	29	30	31	Q7	P7
32	Q8	33	34	35	P8
36	37	Q9	38	39	P9
40	41	42	Q10	43	P10
44	45	46	47	Q11	P11

Figure 5. RAID-6 storage system with 4* 6 stripes (Photo/Picture credit: Original).

D0	D1	D2	D3	D4	D5	D6	D7
0	Q0	1	2	3			P0
4	5	Q1	6	7			P1
8	9	10	Q2	11			P2
12	13	14	15	Q3			P3
16	Q4	17	18	19			P4
20	21	Q5	22	23			P5
24	25	26	Q6	27			P6
28	29	30	31	Q7			P7
32	Q8	33	34	35			P8
36	37	Q9	38	39			P9
40	41	42	Q10	43			P10
44	45	46	47	Q11			P11

Figure 6. Add two new disks to a RAID-6 storage system (Photo/Picture credit: Original).

After a disk is added, the HS6 expansion algorithm first standardizes the storage system so that the storage strip meets the condition that (p-1) x (p +1) and p is a prime number. Figure 6. During the standardization of the storage system, the original storage strip structure is preserved to the maximum extent, and the data migration and computing costs are reduced. The standardization of HS6 expansion algorithm adopts the practice of forward tuning of the following data.

D0	D1	D2	D3	D4	D5	D6	D7
0	Q0	1	2	3			P0
4	5	Q1	6	7			P1
8	9	10	Q2	11			P2
12	13	14	15	Q3			P3
32	Q8	33	34	35			P8
36	37	Q9	38	39			P9
16	Q4	17	18	19			P4
20	21	Q5	22	23			P5
24	25	26	Q6	27			P6
28	29	30	31	Q7			P7
40	41	42	Q10	43			P10
44	45	46	47	Q11			P11

Figure 7. Standardization of RAID-6 storage systems (Photo/Picture credit: Original).

After standardization, when p = 7, the 6 x 8 storage strips in the storage system are migrated to ensure that all diagonal check blocks are on the reverse diagonal. The HS6 expansion algorithm migrates certain data from the old disk to the new disk, and the migrated data is only moved in the same check chain, which realizes the minimum calculation cost of horizontal check [9]. At the same time, after the data is migrated to the new disk, the original skew check data is retained to the greatest extent, which minimizes the calculation cost of skew check, that is, the calculation cost is minimized.

DOI: 10.54254/2755-2721/32/20230187

D0	D1	D2	D3	D4	D5	D6	D7
0	Q0	1	2	3			P0
4	5	Q1	6	7			P1
8	9	10	Q2	11			P2
12	13	14	15	Q3			P3
32		33	34	35	Q8		P8
36	37		38	39		Q9	P9
16	Q4	17	18	19			P4
20	21	Q5	22	23			P5
24	25	26	Q6	27			P6
28	29	30	31	Q7			P7
40	41	42		43	Q10		P10
44	45	46	47			Q11	P11

Figure 8. Move Q8，Q9，Q10，Q11 skew check blocks to diagonal diagonals (Photo/Picture credit: Original).

Figure 7 shows that before standardization, Q0 = 1⊕6⊕11⊕12, Q1 = 2⊕7⊕8⊕13, Q2 = 3⊕4⊕9⊕14, Q3 = 0⊕5⊕10⊕15. Two new disks are added, they are standardized into 6 x 8 storage strips. The HS6 expansion algorithm migrates certain data blocks to the new disk to maintain the integrity of the original diagonal check chain to the greatest extent. After data is migrated to the new disk, the diagonal check chain changes due to the expansion of the storage strip, as shown in Figure 8 after standardization. Q0' = 1⊕6⊕11⊕12⊕ New data ⊕36, so Q0' =Q0⊕ new data⊕36; Q1' = 2⊕7⊕8⊕13⊕32⊕37, so Q1' =Q1⊕32⊕37; The check chain controlled by the Q2', Q3' check block is all new data after standardization, and the check chain needs to be reconstructed; Q8' = 3⊕4⊕9⊕14⊕34⊕39, so Q8' =Q2⊕34⊕39; Q9' = 0⊕5⊕10⊕15⊕35⊕ New data, so Q9' =Q3⊕34⊕ new data. It can be seen that the HS6 expansion algorithm utilizes data analysis of the first 4.2.4 diagonal check chains, reduces the 5 XOR operations required for chain construction to 2 XOR operations, and reduces overhead costs. As shown in Figure 9 [10].

D0	D1	D2	D3	D4	D5	D6	D7
0	Q0	1	2			3	P0
4	5	Q1	6	7			P1
	9	10	Q2	11	8		P2
		14	15	Q3	12	13	P3
32		33	34	35	Q8		P8
36	37		38	39		Q9	P9
16	Q4	17	18			19	P4
20	21	Q5	22	23			P5
	25	26	Q6	27	24		P6
		30	31	Q7	28	29	P7
40	41	42		43	Q10		P10
44	45	46	47			Q11	P11

Figure 9. Migrate certain data to new disks for minimal computational overhead (Photo/Picture credit: Original).

5. Conclusion

This paper introduces the application of erasure codes to enhance storage reliability and proposes optimizations for H-code encoding within RAID 6 storage systems. We also seek to optimize the online capacity expansion process for RAID 6 storage systems by presenting a new HS6 capacity expansion algorithm. This unique approach ensures that during the HS6 expansion, data migration only occurs between the old and new disks, while maintaining the integrity of the parity data. Through this process, we manage to maximize the use of available storage, significantly reducing the cost of data migration and computational overheads. It's crucial to note that the primary motivation for these advancements is to improve the efficiency and reliability of RAID 6 systems, and consequently, support more robust and efficient data storage solutions in our increasingly data-intensive world. A

DOI: 10.54254/2755-2721/32/20230187

comparative analysis between this new algorithm and traditional expansion techniques clearly demonstrates the advantages of our approach. By minimizing the cost of data migration and maximizing the utilization of storage, our method can streamline the expansion process while maintaining system integrity, ultimately contributing to the overall performance and reliability of RAID 6 storage systems.

References

[1] Kamyod C. CIA analysis for lorawan communication model[C]//2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering. IEEE, 2021: 394-397.

[2] Yuan Z, You X, Lv X, et al. HDS: optimizing data migration and parity update to realize RAID-6 scaling for HDP[J]. Cluster Computing, 2021, 24(4): 3815-3835.

[3] Maneas S, Mahdaviani K, Emami T, et al. Operational Characteristics of {SSDs} in Enterprise Storage Systems: A {Large-Scale} Field Study[C]//20th USENIX Conference on File and Storage Technologies (FAST 22). 2022: 165-180.

[4] Kingsmore K M, Puglisi C E, Grammer A C, et al. An introduction to machine learning and analysis of its use in rheumatic diseases[J]. Nature Reviews Rheumatology, 2021, 17(12): 710-730.

[5] Cheng Y, Wei J, Tan X, et al. Research on key technologies of data-oriented intelligent campus in 5G environment[C]//2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE). IEEE, 2022: 203-208.

[6] Bennett H M, Stephenson W, Rose C M, et al. Single-cell proteomics enabled by next-generation sequencing or mass spectrometry[J]. Nature Methods, 2023, 20(3): 363-374.

[7] Jiang T, Zhang G, Huang Z, et al. {FusionRAID}: Achieving Consistent Low Latency for Commodity {SSD} Arrays[C]//19th USENIX Conference on File and Storage Technologies (FAST 21). 2021: 355-370.

[8] Mohsan S A H, Khan M A, Noor F, et al. Towards the unmanned aerial vehicles (UAVs): A comprehensive review[J]. Drones, 2022, 6(6): 147.

[9] Alzahrani A, Alyas T, Alissa K, et al. Hybrid approach for improving the performance of data reliability in cloud storage management[J]. Sensors, 2022, 22(16): 5966.

[10] Wu Y, Dai H N, Wang H, et al. A survey of intelligent network slicing management for industrial IoT: Integrated approaches for smart transportation, smart energy, and smart factory[J]. IEEE Communications Surveys & Tutorials, 2022, 24(2): 1175-1211.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230188

Research on feature coding theory and typical application analysis in machine learning algorithms

Pengxiang Wang [1], Kailiang Xiao[2,4] and Lihao Zhou [3]

[1]JSNU-SPbPU Institute of Engineering, Jiangsu Normal University, Xuzhou, 221116, China
[2]School of Electronic and Information Engineering, South China University of Technology, Guangzhou, 510641, China
[3]School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, 100876, China

[4]202030242188@mail.scut.edu.cn

Abstract. Nowadays, the world is still in the environment of economic depression. In order to promote economic recovery, improve Relations of production and production efficiency, stimulate consumption expansion and upgrading, and accelerate industrial transformation and upgrading, problems such as industrial upgrading need to be solved urgently. Solving the above problems requires more useful tools, and artificial intelligence is one of them. Machine learning is the key to distinguishing artificial intelligence from ordinary program code. Unlike people learning knowledge, machine learning has its own unique language algorithms and behavioral logic. Machine learning, as a technology active in the field of artificial intelligence in recent years, specializes in studying how computers learn, simulate and realize part of human learning behavior, so as to provide data mining and behavior prediction for humans, to obtain new knowledge or skills, or to strengthen the original basic ability of machines. In this study, a variety of common coding algorithms and learning strategies in machine learning are discussed, supervised learning algorithms are selected as examples in the learning strategies, models are further selected and evaluated for a variety of algorithms, and parameters are adjusted and performance is analyzed. As for the theoretical analysis in the research, the paper makes a tentative application in the three fields of housing price, physical store sales and digital recognition, explores and selects the corresponding application method in the appropriate scenario, and expands the application field of machine learning.

Keywords: artificial intelligence, machine learning, feature encoding.

1. Introduction

Machine learning, is specialized in learning how computers learn, simulate, and use human behavior, in order to provide data mining and human behavior to obtain new knowledge or skills. Machine learning, as the basis of artificial intelligence, is the way to make a computer intelligent.

The strategy of machine learning is the inference strategy adopted by the system during the learning process, which involves selecting appropriate machine learning algorithms throughout the entire process. The algorithm mainly includes Supervised learning and Unsupervised learning; The latter discovers and

DOI: 10.54254/2755-2721/32/20230188

summarizes the relationship between input and output based on unlabeled datasets, and makes predictions for new inputs [1].

Focusing on supervised learning, this study conducted in-depth analysis and exploration of linear models under supervised learning, compared the advantages and disadvantages of different supervised learning algorithms, trained, selected and evaluated the models through cross-validation, and analyzed the applied data and parameters through different scenario application and model practice. In order to extend its application range.

2. Related theory

2.1. Seven common coding algorithms in machine learning

2.1.1. Label coding. Coding labels simply means giving different categories and different numerical labels. It belongs to hard coding, that is, it directly maps a large number of category features, and how many category values represent how many. This hard coding method is simple and rude, convenient and fast. In the encoding, sklearn. Preprocessing. labelencoder is used to encode the target label, and its value is between 0 and n classess-1.

2.1.2. Serial number coding. For ordinal variables, assign values from 1 to n to this n-class ordinal variable in sequence. But in fact, the method of characterizing interval variables is applied to ordinal variables, which is similar to label coding, but the categories are coded according to the number order specified in advance. Sklearn. preprocessing.ordinal encoder can encode in two ways: first, load the data categorical_df into the encoder, and then encode the loaded data; second, directly load the encoder and encode it [2].

2.1.3. Binary coding. Binary coding usually refers to linear block coding, a [n, k] linear Block code, eighteen pieces of data are divided into k characters as a section (called group data), and through the encoder, it becomes a group of n long characters, according to the codeword of [n, k] linear Block code.

2.1.4. Frequency coding. Replace the category features with the counts in the training set (generally counting is based on the training set, which belongs to a kind of statistical coding. Statistical coding is to replace the original category with the statistical features of the category, for example, the code is 100 when the category A appears in the training set for 100 times). This method is very sensitive to outliers, so the results can be normalized or transformed (for example, using logarithmic transformation). Unknown categories can be replaced with 1. There is no special encoder for frequency coding, but it can be realized by CountEncoder in the categorical_encodings package.

2.1.5. Mean coding. In machine learning and data mining, whether it is a classification problem or a regression problem, the collected database often includes categorical feature. Because categorical feature indicate that some data belongs to a specific category, numerically, categorical features are usually discrete integers from 0 to n. Generally speaking, for categorical feature, we only need to use tag coding and one-shot coding mentioned in 2.1.1 and 2.1.2. The former can receive irregular feature columns and convert them into integer values from 0 to n-1 (n is n different categories); The latter can make a sparse matrix of m*n through one-hot coding (assuming that the data has a total of m rows, whether the specific output matrix format can be controlled by sparse parameters) [3].

2.1.6. Helmert coding. Helmert coding is usually used in econometrics. After Helmert coding (each value in the classification feature corresponds to a line in the Helmert matrix), the coded variable coefficient in the linear model will show that the difference between the average value of the dependent variable given a certain value of the category variable and the average value of the dependent variable

DOI: 10.54254/2755-2721/32/20230188

given other values of the category. Helmert coding used in category_encoders package is reverse Helmert coding.

2.2. Supervised learning

2.2.1. Linear model return. Minimizing the sum of squares of a residual with penalty by the coefficient of ridge regression:

$$min_{\omega}||X_{\omega} - y||_2^2+\alpha||\omega||_2^2 \tag{1}$$

Ridge regression has changes in the classifier. RidgeClassified, this classifier is sometimes called a vector machine supporting least squares with a linear core. Similar cross validation scores can be achieved by combining recall, accuracy, or accuracy/recall. RidgeClassifier employs different penalty least squares loss methods to provide different digital solvers with different computational performance summaries, based on their respective numerical performance. The logistic function is used to describe the probability of the output results of a single experiment. Logistic regression is realized in LogisticRegression, where binary, one-to-many classification (One-vs-Rest) and polynomial logistic regression are realized, and the regularization options are ℓ1、ℓ2or elastic net. An IsotonicRegression class that fits a non-decreasing function of real numbers to one-dimensional data.

$$\text{minimise} \sum_{i} \omega_i (y_i - \hat{y}_i)^2 \, subject \, \hat{y}_i \le \hat{y}_j \, whenever X_i \le X_j \tag{2}$$

Among them, the weight is strictly positive, and both x and y are arbitrary real numbers. The parameter increasing can change the constraint to even if. Setting it to' auto' will automatically select the constraint according to Spearman's rank correlation coefficient [4]. In terms of mean square error, IsotonicRegression produces a series of predictions for training data, which is the closest to the target. These predictions are interpolated to predict unknown data. Naive Bayes. Given the importance of categorical variables, the independence of each pair of traits is assumed to be biased. Bayes' theorem gives the relationship between class variables and their associated eigenvectors:

$$\mathrm{P}(y \mid x_1,..., x_n) = \frac{P(y)P(x_1,..., x_n \mid y)}{P(x_1,..., x_n)} \tag{3}$$

Conditional independence assumption using Naive Bayes:

$$\mathrm{P}\left(x_i \mid y, x_1,..., x_{i-1}, x_{i+1},..., x_n\right) = P(x_i \mid y) \tag{4}$$

For all, this relationship can be simplified as:

$$\mathrm{P}\left(y \mid x_1,..., x_n\right) = \frac{P(y)\prod_{i=1}^{n} P(x_i \mid y)}{P(x_1,..., x_n)} \tag{5}$$

Because it is an input constant:

$$\mathrm{P}\left(y \mid x_1,..., x_n\right) \propto P(y)\prod_{i=1}^{n} P(x_i \mid y) \Rightarrow \hat{y} = \arg\max_{y} P(y)\prod_{i=1}^{n} P(x_i \mid y) \tag{6}$$

Quantities can be assessed using the method of maximum a posteriori prediction (MAP), which correlates with the classes observed during the training phase. Although this method may appear simplistic, don't be misled by its seeming simplicity. Indeed, it requires minimal information to predict necessary parameters accurately [5]. Moreover, the decoupling of conditional feature distribution classes

DOI: 10.54254/2755-2721/32/20230188

facilitates the independent estimation of each distribution as a single entity. This process can effectively mitigate the dilemmas presented by the curse of dimensionality. However, while Naive Bayes is generally recognized as a proficient classifier, it falls short as a predictor. Consequently, the results generated by the 'proba' prediction should be taken with a grain of salt.

The Decision Tree (DT) is a non-parametric technique utilized for both classification and regression in supervised learning. Some of the key advantages of the decision tree include its interpretability and simplicity. Decision trees can be visually inspected, offering an intuitive understanding of the model. Furthermore, decision trees usually require little to no data preprocessing, unlike other algorithms which often necessitate data normalization or scaling. Decision trees also have the capability to handle both numerical and categorical data, and are capable of solving multi-output problems. Despite these advantages, decision trees also have their share of drawbacks. To circumvent some of these issues, strategies such as tree pruning or setting a minimum number of samples required at a leaf node can be implemented.

2.3. Model selection and evaluation

2.3.1. Feature selection and extraction. Training and testing the parameters of the predictor function on the same data set is a bad approach. It is important to note that "experimentation" is not only used in academia, as in commercial spheres, machine learning usually begins with experiments. Scikit learn Test can be used the split auxiliary function to randomly split dataset into the training set and test set [6]. If the evaluator evaluates various parameters, if parameter C is the super parameter of the support vector machine, the selection of the parameter determines the optimal performance of the model, and the risk of fitting remains before fitting the model. The above problem can be resolved by splitting several known data sets into a "validation set". Training the model with training data to evaluate the model in the validation set. If the "test" score is good, you can finally evaluate the model on the test set.

Dividing the dataset into three or more files reduces the amount of data available for modeling. Therefore, the results of model evaluation depend on the random distribution of training sets and validation sets. One way to solve the above problems is corss-validation (CV). When the cross-validation method is applied. The most basic cross-validation method is k-fold CV, which refers to dividing the training set into k smallest subsets (other methods will be introduced below, and all have the same principle). The following procedure applies one of the k "folds": K-1 subsets are used for model training; *Cross-validation iterators for independent co-distributed data.* It is assumed that some data are independent and identically distributed, that all samples come from the same generation process, and that the generation process is not based on the memory of past samples.

2.3.2. Model persistence, security and maintainability restrictions. Reuse without reconfiguring the model. Examples will be provided later to explain how to use pickle to start the model to be more persistent. When using pickle serialization, some security and maintainability issues need to be reviewed.

Another option for pickle is to output the model in another form. See Related Projects for details by using the model output tool. Unlike pickle, once output, you can't restore the complete Scikit-learn estimator object, but you can use models to make predictions, usually using tools that support open model exchange formats Pickle (and through extended joblib), there are some examples about maintainability and security [7]. Because, Do not use untrusted data that has not been pickle, because it may lead to the execution of malicious files during operation. When the model is loaded in another version, it is totally unsupported and not recommended. It should always be remembered that performing operations on such data may result in different and unexpected results.

2.3.3. Model evaluation. Bias and variance are characteristics of estimators, and learning algorithms and hyperparameters are often chosen to reduce them. Way to cut the standard deviation is to use more informative data. However, if the Performance of the real difference is too difficult to estimate the low difference index, then only more informative data can be collected. It is difficult to visualize models in

DOI: 10.54254/2755-2721/32/20230188

high-dimensional space. Therefore, using tools for description is very useful. In order to verify the accuracy of models such as classifiers, it is necessary to evaluate the function. It should be noted that when optimizing hyperparameters based on validation scores, the validation scores are biased, which is not a good generalization estimate. To get a good estimate of generalizability, you can use a different set of tests to calculate the score. The estimation fails if the training and validation results are low. Lower training scores and higher validation scores are generally not possible. The learning curve shows the results of the validation set and the estimator training set with different numbers of training samples. For naive Bayes, with the increase of the training set size, the score decreases very low. Because of this, it will not benefit a lot from a larger data set. In contrast, the training score of support vector machine is much higher than the verification score for small data.

3. Application

3.1. House price

3.1.1. Project background. House price prediction is a classic machine learning problem and a common item in data analysis. Kaggle provides a concrete example: according to 79 features given, the corresponding house price is predicted, including the type of house, the width of street, the area of each floor, etc., and the evaluation index given is the common mean square error (RMSE) of C in regression problems:

$$RMSE = \sqrt{\frac{1}{\mathrm{n}}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{7}$$

3.1.2. Data processing. Looking at the data in the first five lines, we will find that the' ID' in line 0 cannot participate in training. In the data set, we have a total of 79 related features, which are all kinds of data types, including numerical features and classification features. The size and scale of numerical features are different, and some features also contain missing values. Next, the classification features are processed and replaced by one-hot coding. One-hot coding is to transform different categories into unique features [8]. For example, "MSZoning" contains the values "RL" and "Rm". Finally, the missing value of the data feature is added to 0. For the category features with obvious sequential relationship, such as garage quality' GarageQual ', the bigger the better, LabelEncode is used, and for those without sequential relationship, single heat encoding is used. At the same time, we can artificially add new features.

3.1.3. Training. After data initialization, we can start training. Here we first construct a simplest linear model with MLP. Obviously, the linear model is difficult for us to win the competition, but the linear model provides a reliability check to see if there is meaningful information in the data. If all goes well, the linear model will be used as a base model so that we can intuitively know how much the best model outperforms the simple model.

Firstly, we define our loss function. For house prices, we're interested in the relative price fluctuation, so we're more interested in the relative error (y-y')/y than the absolute error y-y'. The solution is very simple. Just take the logarithm of y, and the division of relative error will be converted into the subtraction of ordinary loss function. The root mean square error can be obtained as follows:

$$\sqrt{\frac{1}{\mathrm{n}}\sum_{i=1}^{n}(\log y_i - \log \hat{y}_i)^2} \tag{8}$$

3.1.4. K-fold cross training and model parameter adjustment. K-fold cross-validation means that the training data set is divided into training data and validation data. That is, the training data set is divided

DOI: 10.54254/2755-2721/32/20230188

into k folds, one of which is taken as verification data, and the others are used as training data for training [9].

Generally speaking, when training a model, it is not that the smaller the training loss, the better. When the model is large enough, the training is easy to over-fit, and the noise of the training data is all learned, and the generalization ability of the model is very poor at this time. As shown in the figure, we intuitively describe the relationship between the model under-fitting and over-fitting. Sometimes the error of K-fold cross-validation is much higher, which indicates that the model is over-fitted (as shown in the previous figure). In the whole training process, we hope to monitor the training error and verification error. Less over-fitting may indicate that the existing data can support a stronger model, while greater over-fitting may mean that we can benefit from regularization technology.

3.2. Store sales - time series forecasting

In order to verify the accuracy of models such as classifiers, it is necessary to evaluate the function. A good method to choose more Rothman of course operates more than 3000 pharmacies in seven European countries. Root mean square percentage error (RMSPE) given in the question:

$$RMSPE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\frac{y_i - \hat{y}_i}{y_i})^2} \tag{9}$$

Note that there is a phenomenon in the data of the day that it opened on the same day, but the sales volume is 0, so it will not be scored at this time.

3.2.1. Data description. Store.csv (provides additional details about each store). 'Store': Represents the unique ID of each store, totalling to 1115. 'StoreType': Indicates one of the four different store models: A, B, C, and D. The distribution is as follows: A(54%), D(31%) and others (15%). 'Assortment': Classifies the level of variety available: a = basic, b = extra, c = extended. 'CompetitionOpenSinceYear': Approximate year when the closest competitor store opened. 'Promo2': Indicates ongoing promotional activities (0= stores not participating; 1= stores participating). The data is equally split between the two. 'Promo2SinceWeek': The calendar week when the store began participating in the promotional activities. 'Promo2SinceYear': The year when the store started the promotional activities. 'PromoInterval': Refers to continuous promo2 activities, specifying the months when the promotions are restarted. For example, "February, May, August, November" means each round of promotions starts in February, May, August, and November each year [10].

Train.csv (includes historical sales data). 'Store': The unique ID for each store. 'DayOfWeek': Specifies the day of the week. 'Date': The specific date. 'Sales': The store's turnover for the day. 'Customers': The count of customers on that day (note: this information is not available in the test.csv). 'Open': Indicates whether the store was open that day (0 signifies 'No'; 1 signifies 'Yes'). 'Promo': Shows if a promotion was running in the store that day. 'StateHoliday': Identifies if the day was a state holiday (generally, with a few exceptions). 'a' represents a public holiday, 'b' is for Easter holiday, 'c' denotes Christmas, and '0' implies no holiday.

3.2.2. Data analysis. After processing all the data one by one, it is found that there are three missing values in the CompetitionDistance, and the CompetitionDistance of each store is quite different (the maximum is 20, and the minimum is 75860), so the median is chosen to fill in, and because the data of the CompetitionDistance of the store is inclined, it is transformed logarithmically [2]. There are a lot of data missing in the competition opening month & competition opening eye, so it is difficult to determine the filling rules for the approximate year and month of the establishment of competitive stores, and the impact on sales after making its data visualization is not obvious. Promo2SinceWeek, Promo2SinceYear and PromoInterval all have 544 missing values. Looking at the data distribution of promo2, it is found that 544 stores have no long-term promotion activities, so the above three characteristics are not data

DOI: 10.54254/2755-2721/32/20230188

missing. All discrete features are processed in two parts: First numeritization, followed by one-hot encoding.

4. Conclusion

Optimization of Models: Model optimization is paramount. We've already seen the development of various optimization methods such as Adam, RMSprop, Adagrad, and others. As we move forward, we will undoubtedly see the rise of new methods designed to enhance the performance and efficiency of deep learning models. Neural Network Structure: As deep learning applications continue to expand, the need for more versatile neural network structures grows. To adapt to different types of data and tasks, innovative network structures such as residual networks, attention mechanisms, and convolutional neural networks are being proposed. We anticipate the creation of even more novel network structures in the future. Model Interpretability: The 'black box' nature of machine learning and deep learning models has resulted in a bottleneck when it comes to model interpretation and interpretability. The need to explain the decision-making process of these models necessitates the development of more transparent models and interpretation methods. Furthermore, it's vital to strengthen the standardization and normalization of model interpretation. Model Generalization and Transfer Learning: The generalization and transfer learning capabilities of machine learning and deep learning models are of utmost importance. Generalization refers to a model's ability to perform on unseen data, while transfer learning relates to a model's adaptability to different tasks and scenarios. To improve these capabilities, we need to develop more robust and transferable models and algorithms, as well as reinforce the standardization and normalization of model evaluation and comparison. Computing Resources and Energy Consumption: Machine learning often requires significant computing resources and energy, which poses a challenge in certain application scenarios. Therefore, the development of more efficient models and algorithms is required, along with a focus on collaborative optimization of hardware and software. Predicted Future Developments: Advancements in Automation and Intelligence: As machine learning and deep learning technologies advance, we are likely to see increased automation and intelligence. This could lead to more intelligent products and services, including voice assistants, smart furniture, intelligent transportation systems, and smart healthcare solutions. Emergence of Federated Learning and Edge Computing: Federated learning and edge computing will become increasingly prevalent. Federated learning allows model training across multiple devices without data being sent to a central server, thereby ensuring data privacy. Edge computing shifts computational power closer to the device, reducing reliance on cloud computing, and improving computational efficiency and data privacy.

In conclusion, machine learning will continue to be pivotal in driving innovation and progress in the future. However, continuous technological advancements and the exploration of applications are necessary to overcome current challenges. By enhancing our understanding of machine learning coding theory and learning strategy, we can compare the merits and demerits of traditional coding theory, apply different models to various application scenarios, and expand the potential of machine learning through adjustment of model outputs and parameters. This should lead to more comprehensive and profound applications of artificial intelligence, promising a brighter future for humanity.

Authors contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Sanni-Anibire M O, Zin R M, Olatunji S O. Developing a preliminary cost estimation model for tall buildings based on machine learning[M]//Big Data and Information Theory. Routledge, 2022: 94-102.

[2] Canese L, Cardarilli G C, Di Nunzio L, et al. Multi-agent reinforcement learning: A review of challenges and applications[J]. Applied Sciences, 2021, 11(11): 4948.

DOI: 10.54254/2755-2721/32/20230188

[3] Sarker I H. Machine learning: Algorithms, real-world applications and research directions[J]. SN computer science, 2021, 2(3): 160.
[4] Li Y. Research and application of deep learning in image recognition[C]//2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA). IEEE, 2022: 994-999.
[5] Sun Q, Ge Z. A survey on deep learning for data-driven soft sensors[J]. IEEE Transactions on Industrial Informatics, 2021, 17(9): 5853-5866.
[6] Wang P, Fan E, Wang P. Comparative analysis of image classification algorithms based on traditional machine learning and deep learning[J]. Pattern Recognition Letters, 2021, 141: 61-67.
[7] Ginart A A, Naumov M, Mudigere D, et al. Mixed dimension embeddings with application to memory-efficient recommendation systems[C]//2021 IEEE International Symposium on Information Theory (ISIT). IEEE, 2021: 2786-2791.
[8] Zhang W, Li H, Li Y, et al. Application of deep learning algorithms in geotechnical engineering: a short critical review[J]. Artificial Intelligence Review, 2021: 1-41.
[9] Jia W, Sun M, Lian J, et al. Feature dimensionality reduction: a review[J]. Complex & Intelligent Systems, 2022, 8(3): 2663-2693.
[10] Zhang Y, Shi X, Zhang H, et al. Review on deep learning applications in frequency analysis and control of modern power system[J]. International Journal of Electrical Power & Energy Systems, 2022, 136: 107744.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230189

Research and application exploration of WiFi-based identification technology in the context of next-generation communication

Wenxu Han

College of Engineering, Nanyang Technological University, 50 Nanyang Avenue, 639798, Singapore

HANW0018@e.ntu.edu.sg

Abstract. With the rapid advancement of wireless networks and the widespread use of WiFi technology, there is a growing interest in utilizing WiFi information for identification purposes. These emerging identification technologies in the new generation have had a profound impact on various aspects of modern life, such as smart furniture research, intelligent security systems, and human-computer interaction. This paper delves into the research and application exploration of WiFi-based identification technology within the context of the new generation. It introduces the knowledge and working principles of Channel State Information (CSI), and discusses the fundamental technologies of Multiple-Input Multiple-Output (MIMO) and Orthogonal Frequency Division Multiplexing (OFDM) in detail. Additionally, it explores current applications and presents a promising future for identification technology from a developmental perspective. By examining the advantages and challenges associated with WiFi-based identification technology, this essay sheds light on its potential impact in various domains. Understanding and exploring this technology are crucial as it has the potential to enhance user experiences, optimize resource allocation, and facilitate intelligent and adaptive systems in the new generation.

Keywords: channel state information, identify technology, Wi-Fi-sensing.

1. Introduction

As a widely adopted wireless communication technology, Wi-Fi offers several advantages such as affordability, easy accessibility, non-contact nature, and passive perception, making it an ideal choice for wireless sensing and recognition. Innovative technologies like Identify Tech and Wi-Fi Sensing have emerged to leverage Wi-Fi signals in various ways. Identify technology, also known as Wi-Fi sensing or Wi-Fi tracking, utilizes signals emitted by Wi-Fi devices to identify and track the presence and movements of individuals within a specific area. These identification systems can detect individuals and estimate their locations without the need for additional hardware or wearable devices. They achieve this by analyzing Wi-Fi signal characteristics such as signal intensity, time of flight, and other relevant parameters. In the context of the new generation, identification technology based on Wi-Fi information has gained significant attention. It offers promising applications in diverse fields. For example, in smart homes, Wi-Fi-based identification technology can enable personalized services

DOI: 10.54254/2755-2721/32/20230189

and seamless automation by recognizing and adapting to occupants' preferences and behaviors. In the realm of intelligent security systems, Wi-Fi sensing can enhance intrusion detection and access control by detecting and identifying authorized users based on their unique Wi-Fi signal patterns. Moreover, in the domain of human-computer interaction, Wi-Fi-based gesture recognition systems utilize Wi-Fi signals to interpret hand movements and gestures, enabling touchless and intuitive interaction with digital devices.

2. Key technologies

2.1. Channel state information

Channel Status Information is widely recommended as a type of practical data to describe channel in wireless communication. It states the combined effects of scattering, fading, power rate attenuation and other effects in the channel, which is a quite fundamental concept. CSI plays an essential role in optimizing wireless system performance and enabling advanced techniques like beam-forming and adaptive modulation. CSI can be obtained through various techniques, including pilot-based estimation, feedback from the receiver, or even measurements at the base station. With accurate CSI, wireless systems can achieve efficient and reliable communication, leading to improved spectral efficiency and overall reliability. In frequency domain, information channel model could be summarized as:

$$Y = HX + N \tag{1}$$

In this equation, Y stands for the channel model and also the signal vector of receiving terminal. X is the signal vector of transmitting terminal. H is the channel matrix. N is Gaussian white noise. Meanwhile, channel matrix represents the multi-carrier information with dimensions of Nt (number of antennas at the receiving end) * Nr (number of antennas at transmitting end) * S (number of subcarriers). Then channel matrix is described as:

$$H_i = \|H_i\| e^{jsin(H_i)} \tag{2}$$

The channel matrix describes the weakening factor of the signal along each transmission path, and the value along each element in the channel matrix contains information as signal scattering, environmental weakening and distance attenuation. CSI could be illustrated in channel matrix H, which is the set of every subcarrier channel information. Each H_i represents for an amplitude and phase of a subcarrier wave. Meanwhile, H comes out in the form of complex number. Responding amplitude and phase could be obtained so long as getting to know about the modulus and argument of the complex number. The channel matrix perfectly describes the features of channel, and the process manufacturing channel state matrix is called channel estimation. The performance of OFDM MIMO receivers is strongly dependent on the channel estimation accuracy [1].

2.2. Multiple input and multiple output technology

When Wi-Fi devices receiving wireless signals, multiple input and multiple output technology is utilized in larger scales. A majority of transmitting antennas and receiving antennas are used respectively at the transmitting terminal and receiving terminal to satisfy the case that the signal are capable to be transmitted and received through antennas in order to upper the quality of communication to a higher level. The additional antennas in a MIMO system provide spatial diversity, enabling the use of multi-path propagation. MIMO makes the most of the inherent spatial dimensions to enhance signal quality, reduce interference, and boost capacity by utilizing many routes and antennas.

Moreover, independent data streams are delivered on various antennas using spatial multiplexing techniques in MIMO, which are merged at the receiver. Because the available spectrum is used more efficiently, larger data speeds are made possible. The number of antennas at either end determines the number of data streams that can be sent and received at once. In addition, MIMO systems use sophisticated signal processing techniques like beam-forming and precoding to enhance signal

DOI: 10.54254/2755-2721/32/20230189

transmission and reception. Signal strength is increased through beam-forming, which concentrates the delivered signal on the intended receiver. In daily life, there are many necessities can't get rod of the relationship with it: Cellular Networks, Wi-Fi, the IoT (Internet of Things) adopted in our daily intelligent furniture, satellite communication and so on.

2.3. Wireless signal propagation model

The way electromagnetic waves move via a wireless communication medium is described by a wireless signal propagation model. It offers a paradigm for comprehending how signals move through space, weaken, and engage with their surroundings. Wireless signal propagation is influenced by a number of variables, such as terrain, barriers, terrain frequency, and environmental conditions. Concepts like free space loss, path loss, fading, and multi-path propagation are frequently incorporated into the model. For some common models, the Free Space Path Loss (FSPL) model, which makes the assumption that signals flow without any obstacles in free space, is one often used model. According to the FSPL model, the received signal strength is inversely proportional to wavelength and diminishes with the square of the transmitter's distance. However, impediments like terrain, buildings, and trees cause wireless signals to lose more signal in real-world situations. More complex propagation models, such the Two-Ray Ground Reflection model, which takes into account both signals that travel directly from the transmitter and those that are reflected off the ground, are employed to take into account these obstructions. Other models, including the Okumura-Hata model and the Log-Distance Path Loss model, integrate empirical data and statistical analysis. As shown in Figure 1.

These models assist engineers and researchers in optimizing wireless network designs, planning coverage areas, and predicting signal strength and quality under various conditions by utilizing mathematical formulas, simulations, or empirical measurements. They play a crucial role in the advancement of wireless technologies such wireless sensor networks, Wi-Fi networks, and cellular networks.

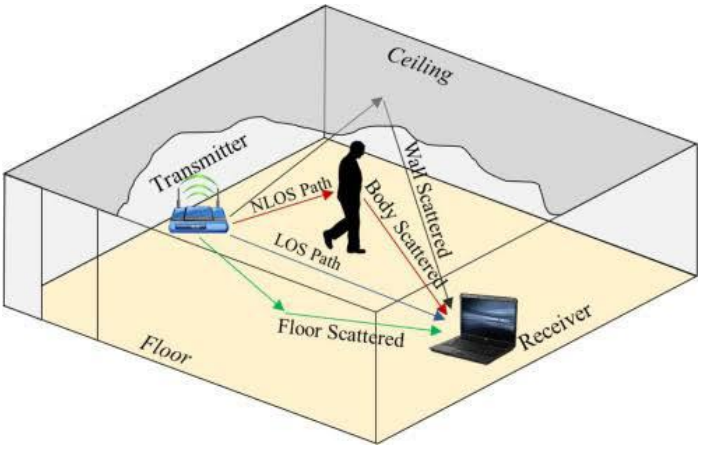


Figure 1. Human body recognition in an enclosure space [2].

2.4. Orthogonal frequency division multiplexing technology

Orthogonal Frequency Division Multiplexing Technology, is treated as a wide ranged modulation and multiplexing technique in modern digital communication systems. It splits up a high-speed data stream into several slower sub streams and sends them all out at once via a number of orthogonal subcarriers. The main goal of OFDM is to mitigate the effects of inter-symbol interference and frequency-selective fading. OFDM achieves resistance against multi-path propagation by employing a significant number of sparsely spaced subcarriers, each carrying a portion of the total data. These subcarriers are specifically created to be orthogonal to one another, preventing interference. Higher spectrum efficiency and effective demodulation are made possible by this orthogonality.

DOI: 10.54254/2755-2721/32/20230189

Data symbols are first transformed into parallel streams in an OFDM system. Then, using modulation techniques like Phase Shift Keying (PSK) or Quadrature Amplitude Modulation (QAM), each stream is assigned to a certain subcarrier. The time-domain signal for transmission is created by combining the resulting modulated subcarriers with an Inverse Fast Fourier Transform (IFFT). Wi-Fi (802.11a/g/n/ac/ax), 4G LTE, and digital audio/video broadcasting (DAB/DVB) all use OFDM in some form or another. It is a crucial technology in contemporary wireless communication systems due to its adaptability and capacity to manage difficult wireless settings.

3. Typical applications and analysis

3.1. Action recognition

Human actions to some extent bring about signal interference. In recent years, data types captured by action recognition in wireless communication are mainly Received Signal Strength (RSS) and CSI. Reference [3] focus more on the application of RSS.

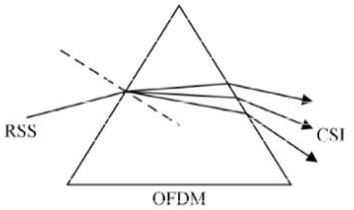


Figure 2. Comparison between RSS and VSI in OFDM system [4].

CSI is considered superior due to its ability to capture steady time-invariant values, providing a large amount of precise data with high granularity. It records the amplitude and phase of subcarrier waves on a subcarrier level, making it highly sensitive and capable of detecting even small movements. Researchers can leverage the varying conditions of wireless signals to identify people's actions in different environments. The process of action recognition based on CSI involves several steps. First, primitive CSI data is collected using the basic operating formula. Next, noise filtering is applied to enhance the reliability of the data, facilitating signal segmentation and feature extraction. Actions' movements are tested, and eigenvalues representing these actions are extracted. The actions are then categorized using classification algorithms, and their accuracy is calculated. CSI-based human action recognition offers advantages such as non-intrusiveness (no need for wearable sensors) and the ability to operate in diverse environments. This technology holds great potential for applications in smart homes, healthcare monitoring, and human-computer interaction, enabling seamless and intuitive interaction between humans and technology. Since people have different body shapes and movement habits, analyzing wireless signals reflecting human activities allows for the analysis of movement speed and modes. As shown in Figure 2.

In the field of action recognition based on CSI, the main classification operation involves placing the processed CSI test data into known action training samples to determine its attribution. Currently, the mainstream methodologies for CSI-based action recognition are the template-based method and the probability-based method. The probability-based approach includes Hidden Markov Models (HMMs), which represent actions as sequences of hidden states and observable features. HMMs offer high flexibility and excellent capabilities for dealing with time-invariant sequences. The essay provides an example where researchers use a decisive Markov model and dynamic programming algorithm to segment and recognize continuous actions simultaneously [5]. Another commonly used model is Conditional Random Fields (CRFs), which model dependencies between neighboring frames and capture the spatiotemporal context of actions. CRFs incorporate both local and global information, enhancing recognition accuracy.

In the template-based method, the Dynamic Time Warping (DTW) algorithm is widely used. It compares temporal sequences by aligning them in time to find the best match. DTW measures the

DOI: 10.54254/2755-2721/32/20230189

similarity between an input sequence and pre-defined templates by warping and stretching the sequences to minimize temporal differences. However, there are still uncertainties and limitations in this technology. Currently, there is no clear definition of human actions or a methodology for hierarchical partitioning of actions. The categorization of small actions versus macro actions is also a topic of discussion, making it challenging to differentiate between them.

3.2. Gesture recognition

The accuracy of gesture recognition heavily relies on capturing the detailed signal reflections of the performed gestures using the Channel State Information (CSI) metric of the Wi-Fi signal [6]. In today's daily life, gesture recognition plays a significant role in various applications, such as controlling intelligent devices like air conditioners through hand gesture interaction. Although most of the mainstream research in hand gesture recognition based on Wi-Fi signals is focused on cataloging gestures, there have been promising advancements in recent years.

In 2018, two periodic hand gesture recognition methods, WiID [7] and SiWi [8], gained attention in the research community. WiID is a user identification system that combines Wi-Fi signals and gesture actions to identify users performing predefined gestures. It utilizes Principle Component Analysis (PCA) for noise reduction on the raw CSI data, extracts features from the velocity time series, and applies radial basis cores to differentiate hand gestures made by different users. SiWi, on the other hand, shares some similarities with WiID and is also a Wi-Fi-based hand gesture sensing identification system. It applies preprocessing techniques such as bandwidth reduction, PCA, and Discrete Wavelet Transform (DWT) on the initial CSI data. Then, it uses Hidden Markov Models (HMM) for cross-scene sensing, training and estimating model parameters to best match the observation sequence.

In 2020, new recognition methods like Finger Pass and WiHF [9] were introduced, bringing fresh perspectives to the field of hand gesture recognition [10]. Finger Pass enables more accurate user identification using lightweight networks, while WiHF achieves significant outcomes in cross-scene gesture recognition and user identification. Both algorithms effectively utilize the amplitude and phase details of the CSI information.

In conclusion, there is still much progress to be made in developing more efficient and accurate methodologies for gesture recognition compared to other mature recognition techniques like action recognition. The ongoing research in this area aims to improve the efficiency, accuracy, and robustness of Wi-Fi-based gesture recognition systems, paving the way for their wider adoption and application in various domains.

3.3. Human body testing

Conventionally, human body testing contain identifications on a wide range of physiological features. Human body possess sorts of unique features. Everybody has different facial features, fingerprint and even pigment contribution on the iris, conveying certain individual’s identity.

Here bring in a daily situation: precise identification on human body’s position. For CSI testing, it typically applies MIMO, besides, by analyzing the channel state information, which provides detailed information about the wireless channel’s characteristics, researchers can also gain insights into the effects of human bodies on signal propagation. During human body capturing, subjects are located in the vicinity of the wireless communication system, and the system measures the channel responses. Then the channel replies contain details on the signal amplitude, phase, and delay at each antenna. Signal attenuation, reflections, and dispersion that the human body causes are all captured in these tests. The impact of body movements, different body orientations, and the presence of barriers like clothing or accessories are just a few of the topics that can be studied using the CSI data that has been acquired. To gauge a system's robustness and dependability in real-world situations, researchers might analyze the signal quality, the signal-to-noise ratio, and other performance metrics.

For action recognition involving multi-person interaction or group movements, different from single-person action recognition, it is supposed to consider the integration of different features and hierarchical modeling of actions. The extraction of various features, the construction of action models

DOI: 10.54254/2755-2721/32/20230189

and the optimization of models are not evolution to perfect versions, and there is still a long way to go on exploiting this technology.

4. Conclusion

In summary, this article provides a concise overview of the basic principles of Wi-Fi sensing and identification technology based on Wi-Fi connection. It highlights the existing identification technologies and their applications across various fields. The essay emphasizes the potential positive effects of Wi-Fi-based identification technology while acknowledging its limitations. The exploration and understanding of this technology are crucial as it holds the potential to enhance user experiences, optimize resource allocation, and facilitate the development of intelligent and adaptive systems for the next generation of communication. As we continue to progress in the era of new generation communication, further research and continuous development in identification technology will undoubtedly lead to exciting advancements that have a beneficial impact on our daily lives, shaping a brighter future driven by high technology.

References

[1] Hoefel, R. P. F. (2012). IEEE 802.11n: On the Performance of Channel Estimation Schemes over OFDM MIMO Spatially-Correlated Frequency Selective Fading TGn Channels. In Brazilian Symposium on Telecommunications, Brasilia, Brazil.

[2] Faheem, M., Shah, S. B. H., Butt, R. A., Raza, B., Anwar, M., Ashraf, M. W., ... & Gungor, V. C. (2018). Smart grid communication and information technologies in the perspective of Industry 4.0: Opportunities and challenges. Computer Science Review, 30, 1-30.

[3] Bianchi, V., Ciampolini, P., & De Munari, I. (2019). RSSI-Based Indoor Localization and Identification for ZigBee Wireless Sensor Networks in Smart Homes. IEEE Transactions on Instrumentation and Measurement, 68(2), 566-575..

[4] Lv, J., Man, D., Yang, W., Du, X., & Yu, M. (2017). Robust WLAN-based indoor intrusion detection using PHY layer information. IEEE Access, 6, 30117-30127.

[5] Wulfmeier, M., Ondruska, P., & Posner, I. (2017). Maximum entropy deep inverse reinforcement learning [EB/OL]. Retrieved November 16.

[6] Ahmed, T., Ahmad, H. F., & C.V., A. (2020). Device-free human gesture recognition using Wi-Fi CSI: A survey. Engineering Applications of Artificial Intelligence, 87, 103281.

[7] Shahzad, M., & Zhang, S. (2018). Augmenting user identification with WiFi-based gesture recognition. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2(3), 1-27.

[8] Zheng, R. Y. (2018). Research on identification technology based on CSI wireless sensing [Doctoral dissertation, Nanjing University of Aeronautics and Astronautics].

[9] Kong, H., Lu, L., Yu, J., et al. (2020). Continuous Authentication through Finger Gesture Interaction for Smart Homes Using WiFi. IEEE Transactions on Mobile Computing.

[10] Li, C., Liu, M., & Cao, Z. (2020). WiHF: Enable User Identified Gesture Recognition with WiFi. In IEEE INFOCOM 2020-IEEE Conference on Computer Communications (pp. 586-595). IEEE.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230190

Comparison of different algorithms in Reversi AI

Xi Chen

Liberal Arts and Sciences, Columbus Community State College, Columbus, 550 East Spring St. Columbus, OH 43215, the United States of America

xchen29@student.cscc.edu

Abstract. Minimax and alpha-beta pruning have been widely applied in AI for various strategic board games, the utilization of greedy algorithms in this context has received less attention. The Greedy algorithms aim to make locally optimal choices at each step, exploiting immediate gains. This research aims to reveal the potential benefits and limitations of applying greedy algorithms in Reversi gaming AI, specifically through a comparison with the Minimax algorithm. A series of AI versus AI matches were conducted to evaluate and compare the performance of the three different AI algorithms. The objective was to assess their gameplay strategies and decision-making abilities by measuring their average execution time and win rates. Relevant codes and experiments will be carried out in a C++ environment, and the shown codes in this article will only have pseudocode and comments. In conclusion, the findings of this study indicate that the Greedy Algorithm is not a superior alternative to the Minimax Algorithm in competitive scenarios, particularly with increased searching depth. However, greedy algorithms still have weak competitiveness with reduced computational time. For future research, concentrating on improving the performance of the Greedy Algorithm by incorporating more advanced heuristics or adaptive strategies maybe a good choice. Additionally, combining the strengths of both the Greedy Algorithm and the Minimax Algorithm could be a promising direction for further investigation.

Keywords: Reversi, greedy algorithms, Minimax, alpha-beta pruning, artificial intelligence.

1. Introduction

In recent years, significant advancements have been made in the field of artificial intelligence (AI), enabling intelligent systems to excel in various fields. Strategic board games have long served as popular test beds for evaluating the capabilities of AI algorithms, including groundbreaking achievements like Alpha-Go. Among these games, Reversi, also known as Othello, presents a complex and challenging environment for AI systems to demonstrate their decision-making prowess. This paper is going to focus on the application of AI techniques, including greedy algorithms, Minimax, alpha-beta pruning, and game theory principles, to enhance the performance of AI systems in playing Reversi.

Reversi is a two-player board game, involves placing and flipping pieces with the objective of gaining the majority of the board. Its simplicity, well-defined rules, and strategic depth make it an ideal choice for studying AI algorithms. While techniques like Minimax and alpha-beta pruning have been widely applied in AI for various strategic board games, the utilization of greedy algorithms in this context has received less attention. Greedy algorithms aim to make locally optimal choices at each

DOI: 10.54254/2755-2721/32/20230190

step, exploiting immediate gains. On the other hand, Minimax allows the AI to evaluate future positions by considering a certain depth of the game tree and alternating between maximizing its own score and minimizing the opponent's score. And by using alpha-beta pruning, we can enhance the efficiency of the Minimax algorithm, reducing the number of explored nodes and improving decision-making speed.

This research aims to reveal the potential benefits and limitations of applying greedy algorithms in gaming scenarios, specifically through a comparison with the Minimax algorithm. This paper aims to reveal whether the utilization of greedy algorithms can lead to increased victories for AI players. By exploring the effectiveness of various techniques in Reversi, we can better understand their applicability in the strategic domain and contribute to deeper and more detailed research on the role of AI in game theory.

2. Different AI strategies

Early game theory can be traced back to [1]. In this section, three distinct Reversi AI algorithms (random move algorithm, Greedy Algorithm, Minimax algorithms with alpha-beta pruning) will be mentioned, and each is designed with different move stragy. The objective of this section is to offer an understanding of the logic and implementation of these algorithms. Pseudocode will be presented in the figures to facilitate comprehension.

2.1. Random move algorithm

In the below pseudocode:

```
// Pseudocode: randomMoveAlgorithm
function randomMoveAlgorithm(board, currentPlayer):
    validMoves = get all valid move positions
    If validMoves is empty:
        return

        randomIndex = generate random integer between 0 and length of validMoves - 1
        randomMove = validMoves[randomIndex]
        row = randomMove.row
        col = randomMove.col

        place currentPlayer’s piece at board[row][col]
```

The “random Move Algorithm” function iterates through all the available positions to make a move and selects one position randomly. It first checks if there are any valid move positions. If there are, it generates a random index within the range of valid moves and selects the corresponding move. Then, it retrieves the row and column of the randomly selected move and places the current player's piece at that position on the board.

This random move algorithm doesn't evaluate the current state of the game and lacks competitiveness, but it provides a simple and quick way for the AI to make a move without spending time on board evaluations. It is commonly used in the simplest difficulty level of AI for the game of Reversi.

2.2. Greedy algorithm

At each turn, the algorithm aims to maximize the number of opponent's pieces that can be flipped. However, if a move is possible in one of the four corners of the board, the AI will add a higher score on it.

In the code below,

DOI: 10.54254/2755-2721/32/20230190

```
// Pseudocode: evaluateMove
function evaluateMove(board, row, col, currentPlayer):
    flippedCount = 0
    for each direction in all directions:
        count = 0
        find all pieces can be flipped
        if piece can be flipped:
            flippedCount += count

    if (row, col) is a corner position:
        flippedCount += 10 // Add a higher score for corner positions
    return flippedCount
```

The the “evaluateMove” function is responsible for evaluating a specific move position on the board. It calculates the scores by considering the number of opponent's pieces that can be flipped by making that move. It iterates through all possible directions from the given position. For each direction, it checks how many of the opponent's pieces can be flipped by extending in that direction. It counts the number of opponent's pieces that can be flipped and adds it to the “flippedCount”. Additionally, if the move position is in the corner of the board, it adds a higher weight to the “flippedCount” to prioritize corner moves. The “evaluateMove” function then returns the “flippedCount” as the evaluation score for that move position.

Then in the “ai2MakeMove” function,

```
// Pseudocode: ai2MakeMove
function ai2MakeMove(board, currentPlayer):
    validMoves = get all valid move positions
    if validMoves is empty:
        return

    bestMove = validMoves [0]
    maxScore = -infinity

    for each move in validMoves:
        row = move.row
        col = move.col
        score = evaluateMove(board, row, col, currentPlayer)

        if score > maxScore:
            maxScore = score
            bestMove = move

    Place currentPlayer’s piece at bestMove.row, beatMove.col
```

This function is responsible for the AI's decision-making process to select the best move position to play. It starts by obtaining all valid move positions available for the current player on the board. If there are no valid moves, the function exits. Otherwise, it initializes variables “bestMove” and “maxScore” to track the move with the highest score. It then iterates through each move position. For each move, it calls the “evaluateMove” function to calculate the score of that move position. If the calculated score is higher than the current “maxScore”, it updates the “maxScore” and sets the current move position as the “bestMove”. After evaluating all the available moves, the function selects the “bestMove” and places the current player's piece at that position on the board.

DOI: 10.54254/2755-2721/32/20230190

2.3. Minimax algorithms with alpha-beta pruning

2.3.1. Minimax. The minimax algorithm is originally from the Park-McClellan algorithm, published by James McClellan and Thomas Parks in 1972 [2]. Minimax is a widely used decision-making algorithm in game theory and AI. It is particularly suitable for turn-based games like Reversi. “Minimax is used to identify the best moves in a game tree generated by each player’s legal actions. Terminal nodes represent finished games; these are scored according to the game rules.”[3]. The Minimax algorithm considers the game as a zero-sum competition between two players: one player maximizes their own score, while the other minimizes it. By recursively evaluating possible moves and their consequences on the game state, the AI can anticipate the opponent's responses and select the most advantageous move. “For any game, we can define a rooted tree (the "game tree") in which the nodes correspond to game positions, and the children of a node are the positions that can be reached from it in one move.” [4]. The depth of the game tree search determines the level of strategic analysis performed by the algorithm, as shown in figure 1.

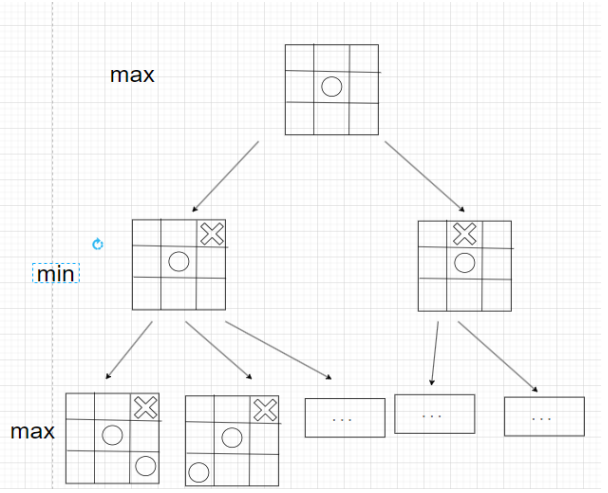


Figure 1. An example of minimax search tree in Tic Tac Toe.

2.3.2. Alpha-beta pruning. Alpha-beta pruning is an optimization technique applied to the Minimax algorithm to reduce the number of nodes evaluated during the search process. It eliminates the evaluation of certain branches that are guaranteed to be less optimal, thus significantly reducing the time cost. By maintaining upper and lower bounds, known as alpha and beta, the algorithm prunes branches that cannot possibly affect the final decision. “Alpha-Beta Pruning is a good optimization of Minimax because achieves the same results using less time and memory, as less moves and less states are evaluated.” [5]. This pruning technique allows the AI to explore a deeper depth of the game tree in less time, as shown in figure 2 [6].

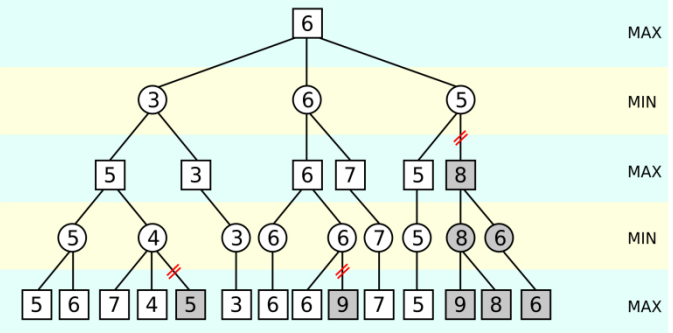


Figure 2. An illustration of alpha-beta pruning.

DOI: 10.54254/2755-2721/32/20230190

3. Evaluation

This section will have two subsections: Test and test result. The test section will cover research methods, research objects, and research tools. The test result section will show the data of the test in figures. Simple analysis will also be included.

3.1. Test

To evaluate and compare the performance of the three different AI algorithms, a series of AI versus AI matches were conducted. The objective was to assess their gameplay strategies, decision-making abilities by measuring their average execution time and win rates.

In these conducted multiple matches, each algorithm played against the other two algorithms in black and white. In these multiple games, each algorithm is played against the other two algorithms 1000 times in black and white, and the winning percentage and average running time between different algorithms will be recorded.

All matches were played on a 6x6 Reversi board. Relevant codes and experiments will be carried out in a C++ environment

3.2. Test result

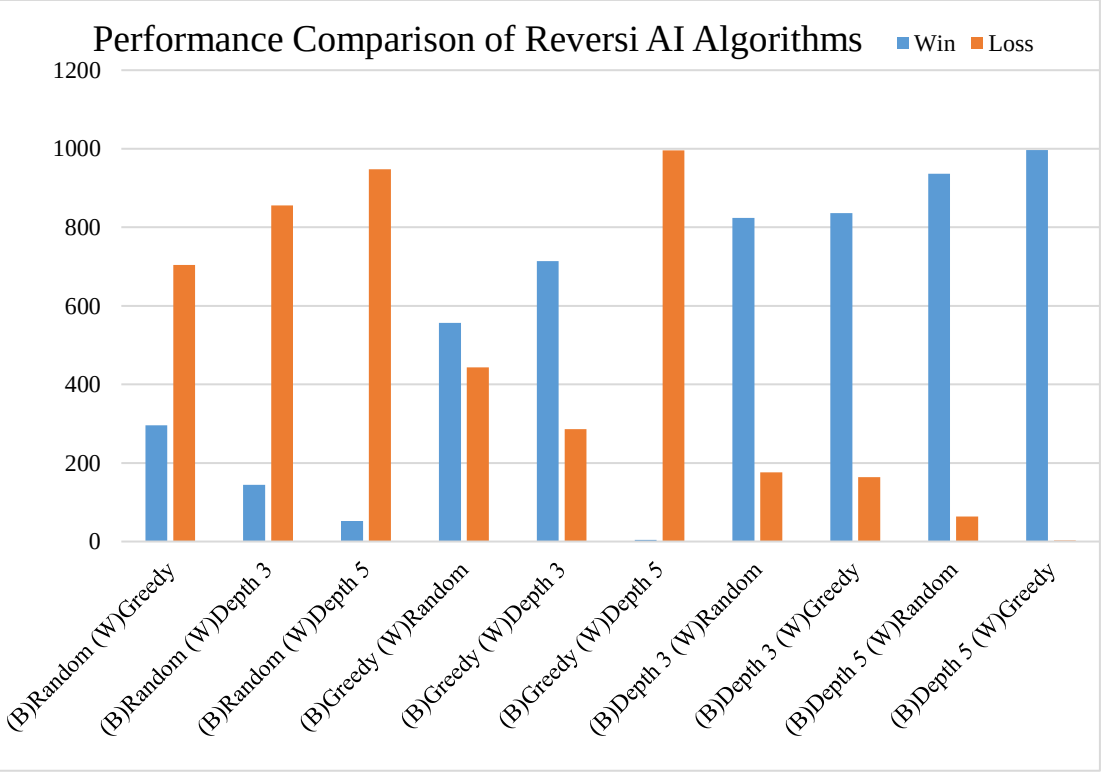


Figure 3. Performance comparison of each algorithms.

According to figure 3, the notation “B” and “W” means the AI are holding black and white pieces, Depth X indicates the depth of this minimax tree is X. For example, (B)Depth 3 means the AI setting as the depth 3 minimax algorithms is holding black piece.

According to the data in figure 6, the minimax algorithms shown a majority of winning, especially as depth increases. The greedy algorithm has shown a major advantage in playing against random algorithms. When holding black pieces, weak advantage is shown when against depth 3 minimax algorithms, but disadvantage when holding white pieces. In general, the greedy algorithm can maintain

DOI: 10.54254/2755-2721/32/20230190

a win rate of 44% against minimax opponents in depth 3. However, as the depth increases, the Greedy Algorithm loses its competitiveness, and its win rate dramatically drops to less than 1%.

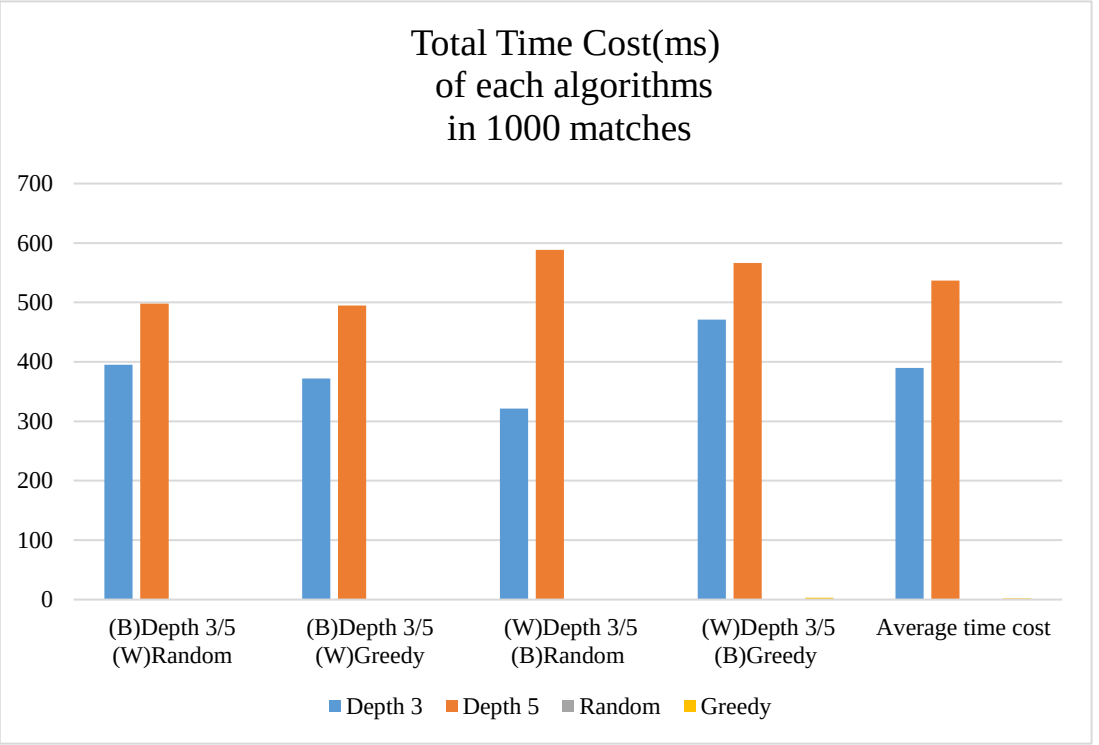


Figure 4. Running time of each algorithms.

According to figure 4, the total runtime of each algorithms in each 1000 matches is shown in millisecond(ms). The runtime of minimax algorithm is increasing with depth, the average runtime time increase by 37.7% as the depth change from 3 to 5, which the minimax algorithm in depth 3 has average runtime in 389.88ms, depth 5 has average runtime in 536.95ms. Random and Greedy algorithms perform great in the runtime test with both of them have a average runtime less than 2ms.

4. Conclusion

The research conducted in this study addressed the question of whether the application of Greedy Algorithms can lead to increased victories for AI players in gaming scenarios, compared to the Minimax Algorithm. Based on the research findings presented in this paper, it can be concluded that the Greedy Algorithm, although efficient in terms of execution time, does not perform better than the Minimax Algorithm, especially when the searching depth of Minimax is increased. The results indicate that the Greedy Algorithm is not a highly intelligent strategy in competitive scenarios, where the Minimax Algorithm with a deeper searching depth showcases superior performance. “The AI preformed much better the more it looked ahead versus looking for moves valuable in the short term.” [7].

However, it is worth noting that the Greedy Algorithm does possess certain advantages. One of its strengths lies in its low computational time, allowing for quick decision-making. In comparison to the Random Algorithm, the Greedy Algorithm gains more victories overall while maintaining a similar execution time.

In conclusion, the findings of this study indicate that the Greedy Algorithm is not a superior alternative to the Minimax Algorithm in competitive scenarios, particularly with increased searching depth. However, its advantage lies in its reduced computational time, making it a viable option for scenarios where time efficiency is prioritized. It is important to further refine and enhance the

DOI: 10.54254/2755-2721/32/20230190

algorithmic strategies in order to develop more effective AI systems for strategic decision-making in various gaming scenarios. For future research, concentrating on improving the performance of the Greedy Algorithm by incorporating more advanced heuristics or adaptive strategies maybe a good choice. Additionally, combining the strengths of both the Greedy Algorithm and the Minimax Algorithm could be a promising direction for further investigation.

References

[1] v. Neumann, J. (1928). Zur theorie der gesellschaftsspiele. Mathematische annalen, 100(1), 295-320.

[2] McClellan, JH, & Parks, TW (2005). A personal history of the Parks-McClellan algorithm. IEEE signal processing magazine, 22 (2), 82-86.

[3] Engel, K. T. (2023). Learning a Reversi Board Evaluator with Minimax. https://www.cs.umd.edu/sites/default/files/scholarly_papers/Engel.pdf

[4] David Eppstein, (1997), Lecture notes for Minimax and negamax search. https://www.ics.uci.edu/~eppstein/180a/970417.html

[5] Festa, J., & Davino, S. (2013). " IAgo vs Othello": An Artificial Intelligence Agent Playing Reversi. In PAI@ AI* IA (pp. 43-50).

[6] By Jez9999, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=3708424

[7] Ross, G. D. (2019). Reversi Artificial Intelligence: A Project Management Analysis. https://digitalcommons.olivet.edu/csis_stsc/13

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230191

Stochastic simulation methods in the study of cell rhythm

Haiqing Xu

The Academy for Software Engineering College; Jilin University, China, Changchun

ciederx@gmail.com

Abstract. Stochastic simulation methods play a crucial role in the study of cellular rhythms. Based on the characteristics of stochastic algorithms, we can more accurately capture the noise effects existing in biological systems and explore their impact on cell rhythms. The findings from stochastic simulation methods shed light on how cell rhythms operate at the molecular level, and this paper presents them inductively for different algorithm types, enabling a deeper understanding of their characteristics. Furthermore, based on the analysis of existing studies, this paper finds that a stochastic simulation approach that considers spatial heterogeneity and intercellular coupling helps reveal the design principles and functional characteristics of the cellular rhythmic system. However, existing stochastic methods also have limitations, including the arbitrariness of parameters and ignoring spatial features. This paper argues that future improvements should focus on integrating quantitative data, accounting for spatial effects, and increasing computational efficiency. These enhancements will contribute to a comprehensive understanding of the generation of cellular rhythms and their importance in biological processes.

Keywords: circadian rhythm, mathematical modeling, Gillespie algorithm, SDEs, stochastic simulation.

1. Introduction

The biological clock, also known as the circadian clock, is an adaptive mechanism that has evolved in organisms on Earth in response to the alternating cycles of day and night. Under constant external conditions, biological rhythms typically follow a 24-hour period and operate autonomously or with the ability to resist external disturbances [1, 2, 3]. The source of these rhythms lies in the gene regulatory feedback reactions, which occur in nearly every cell that comprises life.

To approach the real intracellular reactions, researchers have chosen to abstractly model specific pathways of the cell rhythm.

In the 1950s, computers began to be applied in scientific research [4]. However, due to hardware limitations at that time, verifying complex models still presented significant challenges. This is mainly because the hardware conditions at that time could not meet the needs of the experimenters for continuous simulation of the coupled ordinary differential equation model. To solve this contradiction, the researchers split the concurrent reactions that occur simultaneously in a short period into several sub-reactions that occur continuously and rapidly for simulation. The product of this attempt is the stochastic simulation method we will discuss.

Different stochastic simulation methods commonly used in the study of cell rhythms will be the focus of detailed discussion in this paper. Specifically, we will review the applications of the Gillespie

DOI: 10.54254/2755-2721/32/20230191

algorithm (Stochastic Simulation Algorithm, SSA) and stochastic differential equation simulations (SDEs) in cell rhythm research over the past five years.

In the following sections, we will delve into an in-depth investigation of the Gillespie algorithm, focusing on its stochastic nature and its ability to accurately capture the dynamics of cell rhythms. Additionally, we will analyze the advantages and limitations of the Gillespie algorithm, including factors such as computational efficiency, accuracy, and applicability in different research contexts. Subsequently, we will shift our focus to the application of stochastic differential equation simulations in the study of cell rhythms. We will explore the mathematical foundations of stochastic differential equations, which allow the incorporation of stochastic factors into the models. We will also assess the effectiveness and advantages of SDE simulations in capturing the inherent randomness and dynamic characteristics of cell rhythms.

2. Gillespie algorithm in cellular rhythm research

As one of the most important stochastic simulation algorithms in history, the Gillespie algorithm has a simple core concept and provides reliable effectiveness for experimentalists. In cases where reaction networks are relatively simple, one can solve the governing equations analytically using computers. However, when the situation becomes more complex (e.g., with dozens of different reactions occurring simultaneously), brute force methods are impractical. In contrast, the Gillespie algorithm allows us to simulate the exact dynamics described by the continuous master equation. By discretizing the system through time-division-based stochastic simulation, a single simulation trajectory represents an exact sample of the probability mass function of the master equation solution.

We can represent the algorithm using the following figure 1:

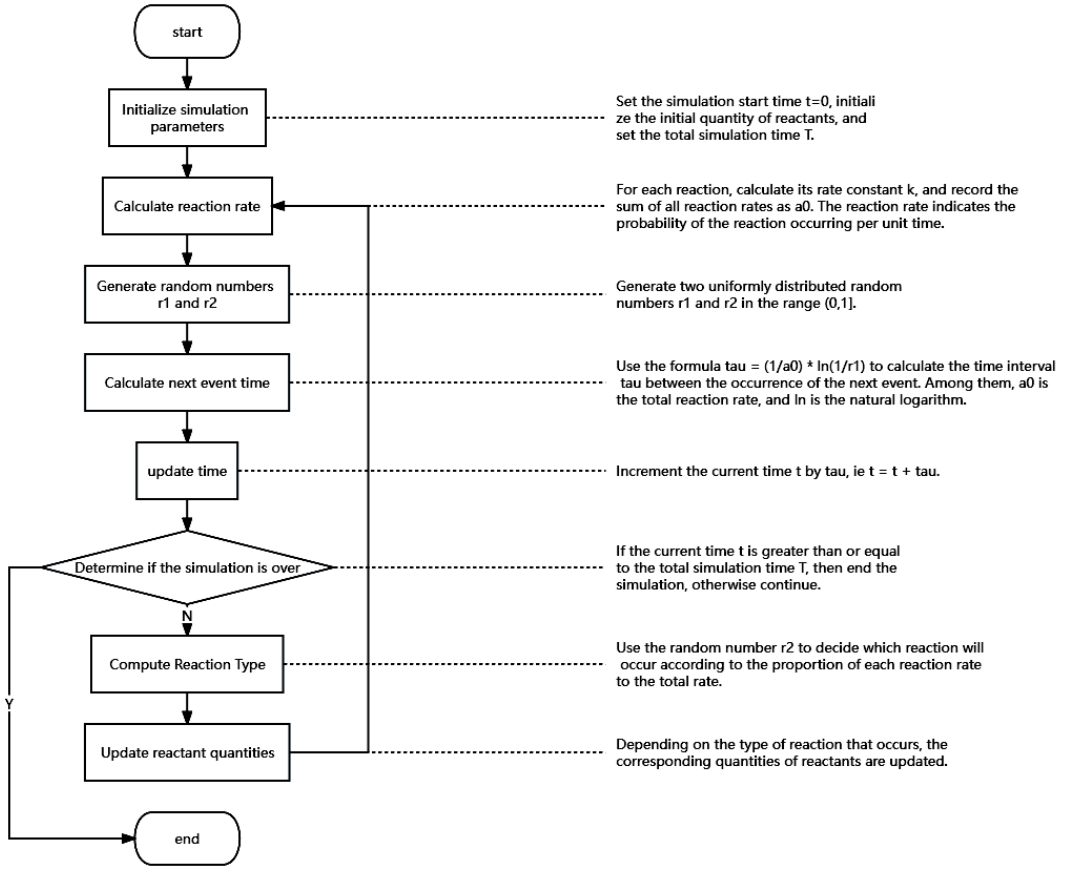


Figure 1. Gillespie algorithm flowchart.

DOI: 10.54254/2755-2721/32/20230191

The discretization of the algorithm is based on individual reaction events rather than a given time step (modeling based on events rather than time). This means that any simulation trajectory faithfully follows the dynamics of the master equation. However, due to the stochastic nature of the algorithm's simulation trajectories, multiple iterations of the above steps are required to obtain the most representative reaction path.

Certainly, due to the different fundamental concepts of the algorithm, deterministic models dominated by the governing equations and stochastic models dominated by the Gillespie algorithm differ significantly in parameter selection as well. B. Xu, H.-W. Kang, and others compared and analyzed the resistance to noise of oscillatory patterns in deterministic and stochastic models based on the dynamics of the Cdc42 GTPase oscillation in yeast [5]. Through numerical simulations and analysis techniques, they observed significant differences between deterministic and stochastic models within the parameter range of interest.

It is worth noting that deterministic models converge to stable limit cycles, while stochastic simulations indicate the existence of noise-induced limit cycles and quasi-cycles. Near the bifurcation point of an infinite period, the deterministic model exhibits sustained oscillations, while the stochastic trajectories initially exhibit oscillatory patterns but tend to approach deterministic steady states. Furthermore, within the low copy number regime, the stochastic model exhibits a transition from oscillatory to stable behavior. In summary, the research by Xu, Kang, and Jilkine reveals the role of stochasticity in Cdc42 oscillations in yeast. Their study demonstrates that through the Gillespie algorithm and its stochastic modeling capabilities, valuable insights into the complex dynamics of cell oscillations can be obtained. Understanding the interplay between determinism and stochasticity is crucial for unraveling the biological oscillatory mechanisms and their functional significance.

R. Zhang, D. Gonze, and others explored the behavior of plant cell circadian rhythms using the Gillespie algorithm embedded in xpp-auto to simulate various aspects [6, 7]. They investigated the effects of molecular noise, light-dark cycles, multiple light inputs, intercellular coupling, and mutations on the variability of rhythm period, phase, and amplitude. Additionally, they utilized discrete simulation methods to capture the stochastic nature of biochemical reactions and studied the robustness of plant circadian rhythms under different conditions. The study observed that light-dark synchrony enhances the robustness of circadian rhythms compared to constant conditions (DD or LL). Multiple light inputs and intercellular coupling improve the robustness of rhythms to noise. Noise leads to phase diffusion, resulting in cells gradually becoming desynchronized under constant conditions.

The Gillespie algorithm provides essential tools for Zhang et al.'s investigation of plant circadian rhythms. By explicitly considering the stochastic nature of biochemical reactions, the algorithm captures the influence of molecular noise on circadian rhythm dynamics. The simulation results demonstrate that discretization provides a more comprehensive understanding of oscillatory behavior, revealing the behavior of mutants and the role of intercellular coupling. It enables researchers to explore different parameters and conditions to gain insights into the robustness and precision of plant circadian rhythms.

3. SDEs in cellular rhythm research

Stochastic Differential Equations (SDEs) are mathematical tools used to describe dynamic systems with inherent randomness. By introducing stochastic terms, SDEs offer a flexible approach to modeling uncertainties and noise in the system. These stochastic terms appear as stochastic differentials in the equations, working alongside deterministic differentials to describe the system's variations [8].

SDEs possess the following characteristics: Firstly, they provide a more accurate representation of complex phenomena in the real world and offer a detailed description of system dynamics. Secondly, the evolution of SDEs has a probabilistic nature, where the same initial conditions can lead to different trajectories. Additionally, SDEs are typically simulated and solved using methods such as Itô calculus, which offers mature analytical techniques. Overall, SDEs provide researchers with a powerful tool for understanding and analyzing dynamic systems with inherent randomness.

S. Miura and T. Shimokawa investigated a cell rhythm model in fruit flies and compared the dynamic outcomes of the Stochastic Simulation Algorithm (SSA) with those of the Chemical Langevin Equation

DOI: 10.54254/2755-2721/32/20230191

(a type of SDE) [9]. They evaluated the system behavior using the oscillation period of the circadian rhythm and the decay time constant of the ensemble-averaged waveform as quantitative metrics. The results showed that the oscillation period of the circadian rhythm was similar in both the deterministic model and the two stochastic methods, regardless of the system size. However, some differences were observed in the decay time constant of the ensemble-averaged waveform, which reflects the oscillation coherence in the presence of noise. Particularly for small system sizes, the Chemical Langevin Equation demonstrated a more significant impact of molecular noise compared to the direct method. The study also indicated that the Chemical Langevin Equation is more suitable for simulating systems with characteristics similar to real biological systems, such as small volume and low molecule numbers in single-cell systems.

As a commonly used type of SDE, the Langevin equation provides a systematic quantification approach for incorporating stochastic effects into mathematical models. By introducing noise intensity as a function of state and time, the Chemical Langevin Equation accurately describes biochemical reactions with low molecular counts. It contributes to a better understanding of the influence of randomness on circadian rhythm generation and provides insights into system behavior in the presence of molecular noise.

In summary, SDEs, including the Chemical Langevin Equation, are powerful tools for studying dynamic systems with inherent randomness. They can capture the probabilistic nature of system dynamics and provide a more accurate representation of real-world phenomena offers valuable insights. These insights have enhanced the understanding of the impact of stochasticity on circadian rhythm generation and provide a comprehensive view of system behavior in the presence of molecular noise.

4. Conclusion

Stochastic simulation methods have played a crucial role in advancing our understanding of cell rhythm dynamics. In this paper, we have explored the application of stochastic simulation methods in cell rhythm research and discussed their limitations and potential future improvements.

Stochastic simulation methods, such as the Gillespie algorithm and the Chemical Langevin Equation, enable researchers to capture the influence of molecular noise on cell rhythm generation. These methods provide valuable insights into the stochastic effects on the behavior of biological systems with low molecule numbers. By incorporating noise intensity as a function of state and time, the Chemical Langevin Equation offers a more accurate representation of intracellular biochemical reactions.

One key advantage of stochastic simulation methods is their ability to model complex biological systems with spatial and temporal heterogeneity. These methods have been applied to investigate the dynamics of circadian rhythms, where the interplay between cellular circuits, synchronization, and robustness to noise is crucial. By incorporating spatial features and intercellular coupling mechanisms, stochastic simulation methods help uncover the design principles and functional characteristics of circadian rhythms.

However, we must acknowledge the limitations of current stochastic simulation methods. One common criticism is the arbitrariness of the choice of equations and parameter values. While these models provide qualitative insights into cell rhythm dynamics, they lack quantitative data such as precise parameter values and absolute concentrations. Furthermore, most models neglect spatial aspects and assume molecular-free diffusion within cells, which may not accurately represent the crowding and high-order nature of cellular processes.

To address these limitations, future improvements in stochastic simulation methods can focus on several aspects. Firstly, efforts should be made to incorporate more quantitative data, such as accurate parameter values, into the models. This will enable more precise quantitative predictions and facilitate the comparison of simulation results with experimental measurements. Secondly, incorporating spatial features and diffusion processes into stochastic simulations will provide a more realistic representation of cellular systems. This requires considering factors such as spatial organization, local interactions, and diffusion rates within cells. Lastly, further advancements in the analysis of large-scale stochastic state

DOI: 10.54254/2755-2721/32/20230191

spaces and the development of efficient computational algorithms will allow us to explore larger and more complex biological systems.

In conclusion, stochastic simulation methods play a significant role in advancing our understanding of cell rhythm dynamics. While current methods have limitations, ongoing improvements in integrating quantitative data, spatial features, and efficient computational algorithms offer promising directions for future stochastic simulation methods. These advancements will contribute to a more comprehensive understanding of cell rhythm generation and its significance in biological processes.

References

[1] Dunlap J.C., Loros J.J., Decoursey P.J., (2004). Chronobiology: Biological Timekeeping, Sinauer Associates Inc., Sunderland.

[2] Edmunds L.N. (1987). Cellular and Molecular Bases of Biological Clocks: Models and Mechanisms for Circadian Timekeeping, Springer, New York.

[3] Mairan J. J., (1729). Botanical observation, History of the Royal Academy of Science, 35-36 (in French)

[4] Ceruzzi, P. E. (2003). A History of Modern Computing (pp. 30-32). MIT Press.

[5] Xu, B., Kang, H.-W., & Jilkine, A. (2019). Comparison of Deterministic and Stochastic Regime in a Model for Cdc42 Oscillations in Fission Yeast. Bulletin of Mathematical Biology, 81(5), 1268-1302.

[6] Ermentrout, B., & Mahajan, A. (2003). Simulating, analyzing, and animating dynamical systems: a guide to XPPAUT for researchers and students. Appl. Mech. Rev., 56(4), B53-B53.

[7] Zhang, R., & Gonze, D. (2021). Stochastic simulation of a model for circadian rhythms in plants. Journal of theoretical biology, 527, 110790.

[8] Oksendal, B. (2013). Stochastic differential equations: an introduction with applications. Springer Science & Business Media.

[9] Miura, S., Shimokawa, T., & Nomura, T. (2008). Stochastic simulations on a model of circadian rhythm generation. BioSystems, 93(1-2), 133-140.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230194

Biodegradable materials in tissue engineering and regenerative medicine

Jiarong He

School of Materials Science and Engineering, Nanjing University of Posts and Telecommunications, Nanjing, 210023, China.

196131124@mail.sit.edu.cn

Abstract. Regenerative medicine signifies that medicine will step into a new era of reconstruction, manufacturing, and replacement of tissues and organs. At the same time,the mankind face many challenges which the development of medicine brings. Along with the progress and development of medical science and technology and the concepts of tissue engineering and regenerative medicine, it has a significant role in advancing the development of human medical technology and future tissue and organ regeneration and repair. This paper will introduce the concept of biodegradable materials and categorize biodegradable bio-materials into synthetic biodegradable bio-materials such as polylactic acid derivatives, copolymers of polyhydroxyacetic acid (PHA) and polylactic acid (PLA), and naturally occurring biodegradable bio-materials such as collagen, chitosan, as well as including their applications and research in tissue engineering. Finally, we make a beautiful outlook on the role of regenerative medicine in the future of human life for human health management and repair.

Keywords: tissue engineering, biodegradable materials, synthetic degradable materials, naturally occurring biodegradable material.

1. Introduction

In the survival needs are gradually satisfied on the basis of people's demand for medical level is also gradually rising, due to some major diseases will cause irreversible damage to human organs and their functions, so the regeneration of organs and tissues for the research has become a new hot spot in today's society. At the same time for some repair damage need to be ingested drugs and their carrier materials have also become a new challenge, as the drug into the body for human tissue repair or other to achieve other functions, such as the realization of the function of the fixed-point time-release drugs, the requirements of the drug in the inhalation of the human body at the same time can not be absorbed, and need to be a certain length of time to produce the effect. In order to realize this function, it is necessary for biodegradable materials to play its characteristics, because it is a biological material, so the biocompatibility should be as high as possible and the human body's rejection reaction should be reduced to none, in addition to ensure that there is a certain degree of mechanical strength and solubility, so that the carrier material can be realized to encapsulate the drug, so that the drug can’t be absorbed by the human body for its period of time to get protection from the encapsulated material is too short, and in a certain time, that is, the encapsulated material is degraded, so that the human body is unable to absorb the drug. A certain time that is, after the degradation of the carrier material, the drug can be absorbed by

DOI: 10.54254/2755-2721/32/20230194

the human body, and as a biodegradable material, so the carrier material does not need to take great pains to take it out, but only needs to be allowed to be slowly decomposed and absorbed in the human body. Besides, Regenerative medicine and tissue engineering still has a lot of problems to overcome, and the research on the application of biodegradable materials in regenerative medicine and tissue engineering has only just begun, and there is still a lot of space to explore and discover. As a summary article, this paper explains the types, characteristics and applications of biodegradable materials. The biodegradable materials are divided into two categories: artificial degradable materials and natural degradable materials, and then branch out from these two categories into subdivided categories, and elaborate the characteristics and applications of each of them, for example, artificial degradable materials are divided into polylactic acid derivatives and copolymers of poly (hydroxyacetic acid) and polylactic acid, and natural degradable materials are divided into collagen and chitosan.

Biodegradable polymers can undergo chain-breaking reactions through three degradation mechanisms: photochemical degradation, thermochemical degradation, and biochemical degradation, resulting in a decrease in mechanical properties and an increase in solubility, and the degraded oligomers are further broken down into small monomeric molecules that enter the body fluid cycle. Biodegradable polymers can help simplify medical procedures as non-permanent implant materials in living organisms.

Tissue engineering as a new concept of regenerative medicine was introduced by American scientists Langer and Vacanti and is defined as the research and development of life science technologies that can repair, improve, or replace human tissues or functions [1].

2. Synthetic biodegradable polymer materials

2.1. Poly (lactic acid)-based derivatives

Polylactic acid derivatives in recent years in the biological scaffold material research room biomedical tissue engineering a hot spot, biological scaffold material as a key topic of tissue engineering research, in addition to providing a good survival microenvironment for the cells and to the cell attachment, growth, adhesion, and other positive interactions, and can be appropriate to induce the ability of differentiation. Polylactic acid (PLA), as a biodegradable material that can be used as a bioscaffold material, has its own unusual characteristics [2-3].

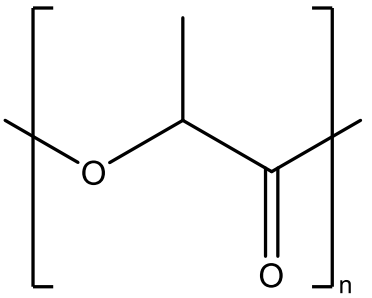


Figure 1. Molecular structure formula of polylactic acid derivatives(origin).

The molecular formula of the polylactic acid derivative is shown in the figure 1 above. Polylactic acid derivatives have been widely applied as scaffolds in tissue engineering for many years and cellular building blocks, and these biodegradable polymers are distributed and absorbed in vivo as the tissue is reconstructed. Polylactic acid has good biocompatibility and biodegradability [4] with human tissues and does not cause rejection in the body, and the hydrolysis products of Polylactic acid can participate in the metabolic process of sugars in the body without causing any residual or biological side effects, so Polylactic acid is now widely used in the production of medical materials such as sustained release drug carriers and implants for biological tissue repair.

DOI: 10.54254/2755-2721/32/20230194

Lactic acid has two functional groups, hydroxyl and carboxyl, and can therefore undergo a self-condensation reaction under certain conditions to produce polylactic acid, which has a life span of 4-6 months in physiological salt water. By adjusting the ratio of different polylactic acids polymerized from different chiral configurations of lactic acid, there will have an adjustment for the strength and degradation time of the material. Since hydroxyapatite is the basic component of human bones, which promotes bone growth by close bonding with collagen or cells, but presents brittle characteristics when it receives tension, which means its mechanical properties are poor, so compounding hydroxyapatite with poly (levulinic acid) to form a new biodegradable composite can improve its mechanical strength. It has been shown that introducing polymeric phase molecules into poly (levulinic acid), the mechanical strength of the material rose by five times. The combination of Polylactic acid and other materials into a composite material can make the material mechanically supportive as well as a drug carrier. Also there is a close relationship between the mechanical strength and hydrolysis rate of Polylactic acid and molecular weight distribution of the polymer and the polymer's average molecular weight; when the molecular weight increases, the mechanical strength also increases and the hydrolysis time becomes longer. When the molecular weight of Polylactic acid is lower than 2500, the mechanical strength is poor and the hydrolysis rate is too fast, while when the molecular weight is more than 1 million, the degradation reaction will not occur for a long time, so the molecular weight of Polylactic acid is a important method to regulate the degradation rate and performance. Polylactic acid is not toxic, has a certain life span in the human body, and the degraded product can be absorbed and metabolized by the human body, so it can be used as a tissue defect reinforcement material. Non-woven and mesh fabrics made from PLA fibers can be used to repair surgically removed chest walls, for example, and can be used for tendon reconstruction after de-stretching and strengthening. Polylactic acid can also be used as a bonding and fixation material for surgical procedures. Since the bending strength and bending modulus of stretched Polylactic acid are comparable to those of the original bone, it can be implanted in the body to protect the broken bone and can be absorbed by the body after healing to prevent secondary damage.

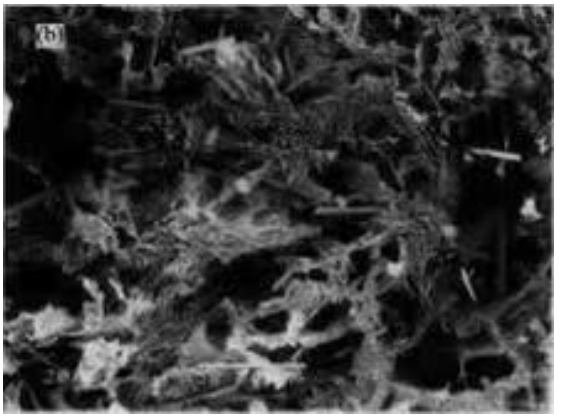


Figure 2. Scanning electron microscope micrographs of PLLA and $CaCO_3$/calcium polyphosphate/ poly-(L-Lactic acid) composite tissue engineered scaffolds [5].

The above figure 2 shows the SEM diagram of Zhu et al. showed for the construction of ball-milled calcium carbonate/polycalcium phosphate fiber/polylactic acid tissue engineering composite scaffold material, and it is observed from the figure that the scaffold composite material has a three-dimensional mesh structure within the cross-section, microporous structure, random distribution of fibers, and relatively uniform distribution of microporosity, and these structures can provide suitable nutrient transport conditions, space conditions, and excretion exchange conditions for the growth and reproduction of tissue cells, which are conducive to the specific physiological functions of bone and cartilage tissues [5].

DOI: 10.54254/2755-2721/32/20230194

2.2. Polylactic acid and polyhydroxyacetic acid copolymer

Biodegradable synthetic polymers have highly controllable degradation properties as well as excellent mechanical and physical properties are suitable for making tissue engineering scaffolds. Pan et al. prepared electrospun dextran/polylactic acid-hydroxyacetic acid copolymer composite scaffolds and applied them to mice for in vivo degradation experiments. After three days, the scaffolds degraded to half of their original size, and after three weeks, three-quarters of the scaffolds were absorbed and the surrounding tissues were completely repaired. [6] Polylactic acid and polyhydroxyacetic acid can bind to cells and biomolecules to enhance biological properties, and biodegradable polymers cannot structurally replicate the mineral composition of bone. However, the biodegradability of copolymers and the ability to incorporate release signaling molecules allow copolymers to be a support for bone formation, and the biological properties of low bioactive poly(lactic acid) poly(hydroxyacetic acid) copolymers can be improved by incorporating bioactive substances to improve bone regeneration and repair.Liu et al. constructed poly(lactic acid) poly(hydroxyacetic acid) scaffolds containing simvastatin (SIM) and stromal cell-derived factors and found that SIM at 0.2 μmg/L increased the alkaline phosphatase activity of mouse bone marrow MSDs.The chemotactic ability of SDF-1α is also enhanced, and the test results showed that poly(lactic acid) polyhydroxyacetic acid with these modifications significantly promoted bone regeneration [7].

3. Natural biodegradable polymers

3.1. Collagen

Collagen is found in many tissues of humans and animals and is one of the main components of bone tissue. Collagen exists in the body in the form of collagen fibers and induces mineral deposition, which provides protection and support to cells and is closely related to cell growth. The fibrin monomers in collagen fibrils can be polymerized by thrombin into a fibrin coagulase with a homogeneous reticular cross-linked structure. Collagen and fibrin thrombin have poor mechanical properties, but fibrin thrombin can promote cell adhesion and multiplication by releasing tumor necrosis factor and platelet-derived growth factor, and fibrin thrombin has good biocompatibility and biodegradability. Zhao et al. achieved cell growth propagation and adhesion with improved mechanical properties by composing fibrin gel and chitosan into a scaffold with a three-dimensional reticular cross-linked structure [8].

3.2. Chitosan

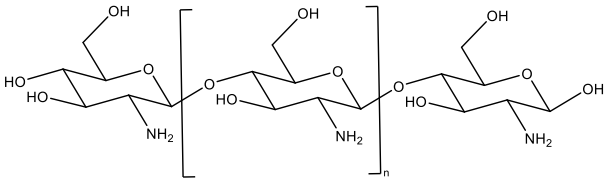


Figure 3. Chitosan molecular structure formula(original).

The molecular formula of chitosan is shown in the figure 3 above. Amaral et al. cultured murine bone marrow stromal cells in vitro knowing chitosan membranes and showed that 4% of chitosan membranes showed substantial cell adhesion, cell proliferation as well as osteogenic differentiation. Chitosan is produced by deacetylation of chitin, a natural polysaccharide found mostly in the shells of insects, etc., as well as in the cell walls of fungi. Chitosan has good biocompatibility and degradability [9]. If chitosan is made into a three-dimensional porous scaffold and implanted into the body, it can be hydrolyzed into oligosaccharides under the action of lysozyme, and then further broken down into small monomeric molecules to be metabolized and absorbed by the body. To realize the application of chitosan in tissue engineering, its mechanical properties pally an important role in this aspect. Xu et al. make a composite

DOI: 10.54254/2755-2721/32/20230194

material of calcium carbonate and chitosan to realize the application of chitosan in bone tissue engineering to help to get the mechanical properties improved [10].

4. Conclusion

This paper firstly introduces the concepts of biodegradable materials and regenerative medicine in tissue engineering, and clarifies the prospects of biodegradable materials in biomedical applications. Then it explains and introduces different biodegradable materials and the different functional effects of different materials in tissue engineering. Firstly, we introduce the synthetic biodegradable materials, focusing on the polylactic acid derivatives and copolymers of polyhydroxyacetic acid and polylactic acid, polylactic acid derivatives because of their special physical and chemical properties, such as greater mechanical strength, higher biocompatibility, degradation of substances can be absorbed and metabolized by the human body and so on, become a slow-release drug carriers, tendon reconstruction, bone regeneration and so on, a good medical material. Then the copolymers of polyhydroxyacetic acid and polylactic acid are introduced. The biodegradable biomaterials are synthesized and polymerized to take advantage of the advantages of the monomer materials and realize the improvement of the composites biocompatibility and mechanical strength by the combination of them, and to better achieve the functions of biological tissues' repair and regeneration. Subsequently, natural biodegradable biomaterials were introduced, taking collagen and chitosan as an example, collagen can be used to achieve cell growth and reproduction by forming fibrin clotting enzymes under the action of enzymes, and polymerization with chitosan can improve the mechanical properties to achieve cell growth and reproduction, whereas chitosan improves the mechanical properties to achieve the utilization of chitosan in the bone tissues to achieve the regeneration of the bone tissues in the case of polymerization with calcium carbonate.

In the era of rapid development of medical technology, people can still see many people with physical organ defects or disabilities have many inconveniences in their lives and long-time disabilities have irreversible damage to their psychology, and the breakdown of body organs and functions will lead to the health of human beings is not optimistic. In order to restore the health of those who are physically handicapped by accidental injuries and need to take prolonged-release medication to maintain their lives, not only respect and care, but also solve the problem from the root. Researchers has make many efforts to tackle the problem at its roots to help the development and research of biodegradable materials in tissue engineering and regenerative medicine, and people are looking forward to it! We all look forward to a future where everyone is physically and mentally complete.

References

[1] Langer R, and J P. Vacanti 2016 *J.* Advances in tissue engineering. *J Pediatr Surg*. **51**(1):8-12.
[2] Dawson E, Mapili G, Erickson K, Taqvi S and Roy K. 2008 *J.* Biomaterials for stem cell differentiation. *Adv. Drug Delivery Rev.* **60**(2):215-28.
[3] Nakagawa M, Teraoka, F, Fujimoto S, Hamada Y, Kibayashi H and Takahashi J. 2006 *J.* Improvement of cell adhesion on poly(L-lactide) by atmospheric plasma treatment. *J Biomed Mater Res A.* **77**(1): 112-8.
[4] Hong Z, Zhang P, He C, Qiu X, Liu A, Chen L, Chen X and Jing X 2005 *J.* Nano-composite of poly(L-lactide) and surface grafted hydroxyapatite: mechanical properties and biocompatibility. *Biomaterials* **26**(32):296-304.
[5] Zhu L, Wang Y, Shi Z and Zhang H 2010 *J.* Construction of CaCO3/calcium polyphosphate / poly-(L-Lactic acid) composite tissue engineered scaffolds. *Chin J Tissue Eng Res Clin Rehabil.* **14**(21):3823-6. doi:10.3969/j.issn.1673-8225.2010.21.006.
[6] Pan H, Jiang H, and Chen W 2008 *J.* The biodegradability of electrospun Dextran/PLGA scaffold in a fibroblast/macrophage co-culture. *Biomaterials*. **29**(11):1583-92.
[7] Liu Y, et al. 2014 *J.* The effect of simvastatin on chemotactic capability of SDF-1α and the promotion of bone regeneration. *Biomaterials* **35**(15):4489-98.

DOI: 10.54254/2755-2721/32/20230194

[8] Zhao F, Yin Y, Lu WW, Leong JC, Zhang W, Zhang J, Zhang M and Yao K. 2002 *J*. Preparation and histological evaluation of biomimetic three-dimensional hydroxyapatite/chitosan-gelatin network composite scaffolds. *Biomaterials* **23**(15):3227-34.

[9] Amaral IF, Lamghari M, Sousa SR, Sampaio P and Barbosa MA 2005 *J*. Rat bone marrow stromal cell osteogenic differentiation and fibronectin adsorption on chitosan membranes: The effect of the degree of acetylation. *J Biomed Mater Res A.* **75**(2):387-97.

[10] Xu HH and Simon CG Jr 2005 *J*. Fast setting calcium phosphate–chitosan scaffold: mechanical properties and biocompatibility. *Biomaterials* **26**(12):1337-48.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230195

Flexible wearable biosensor for physiological parameters monitoring during exercising

Haochen Sun[1,4], Ziyao Xu[2] and Runquan Zhou[3]

[1]School of Communication Engineering, Shanghai University, Shanghai, China
[2]The Affiliated International School of Shenzhen University, Shenzhen, China
[3]Damien High School, La Verne, United States

[4]sunhc1231@shu.edu.cn

Abstract. With the continuous improvement of the quality of life, people's demand for sports is gradually increasing, therefore, the monitoring of physiological parameters during exercise not only has the effect of improving the performance of athletes, but also provides a guarantee of the quality of training for the majority of sports enthusiasts.Monitoring blood oxygen saturation is essential for assessing a person's oxygen levels, and it can be achieved through electrochemical or optical methods. In this paper, we focus on the optical method, which utilizes a pulse oximeter equipped with an infrared light source and a light sensor. This device measures the absorption of light by hemoglobin, allowing the calculation of oxygen saturation. Flexible temperature biosensors are designed to measure environmental or object temperatures using thermosensitive materials or thermistors. The flexibility of these sensors allows them to adapt to irregular surfaces and curved shapes, making them suitable for monitoring human body temperature during exercise. Wearable heart rate monitors use various detection techniques, such as photoplethysmography (PPG) and electrocardiogram (ECG). PPG measures changes in blood volume using LEDs and photodiodes, while ECG detects the heart's electrical impulses. These monitors are widely used in fitness and sports settings, enabling users to track their cardiovascular health, adjust exercise intensity, and set performance goals. The incorporation of additional accelerometers or gyroscopes enhances heart rate monitoring accuracy during exercise, filtering out motion-induced noise for reliable data.

Keywords: monitor, flexible, blood oxygen, temperature, heart rate.

1. Introduction

With the continuous progress of science and technology and people's increasing concern for sports and health, people have higher requirements for the accuracy and professionalism of monitoring equipment, conventional sports monitoring methods can no longer meet people's needs, and wearable technology has gradually become a concern. As an emerging technology, flexible wearable biosensors can real-time, accurate monitoring of the body's physiological conditions during exercise, providing an effective tool for sports training and health management. Wearable flexible biosensors have wide applicability in exercise monitoring applications. Whether it is step counting and sleep monitoring in daily life or training data analysis for professional athletes, wearable flexible biosensors are able to meet the demand. Through real-time monitoring of sports status and body parameters, wearable

DOI: 10.54254/2755-2721/32/20230195

flexible biosensors can help users understand their sports conditions and provide personalised health advice and training guidance to improve their sports performance and health. Compared with traditional hard sensors, flexible biosensors have better deformability and flexibility, and can more accurately obtain human movement information. At the same time, the characteristics of flexible materials also make the sensor can be closely integrated with the human body, which can provide a comfortable wearing experience and will not cause interference to the movement, people will be more willing to wear and use, which provides a more portable option for sports monitoring and more reliable data support. However, wearable flexogen sensors currently face some challenges in motion monitoring. The stability and precision of the sensors need to be further improved to ensure the accuracy and reliability of the data.

This paper introduces three types of wearable flexible biosensors for blood oxygen saturation monitoring, skin temperature monitoring and heart rate monitoring, respectively. It also describes in detail the components of these wearable biosensors, device structure, and working principle.

Wearable flexible biosensors have great potential and value in motion monitoring applications. Through further research and development, it is believed that it can play a greater role in the field of exercise monitoring in the future and bring more convenience to people's healthy life.

2. Biosensors used to monitor blood oxygen saturation

Blood oxygen saturation measurement is usually divided into electrochemical and optical methods. The electrochemical method is based on the sampling of human blood for electrochemical analysis using a hematology analyzer, and the optical method is based on the non-invasive measurement of the difference in light absorption between oxyhemoglobin and deoxyhemoglobin in blood at different wavelengths. Since our research is based on a wearable flexible monitoring device during motion, we choose optical method sensors here.

Oximetry is generally achieved by the device called Pulse Oximeter. Its working principle is based on the light-absorbing properties of hemoglobin. Hemoglobin is an important molecule in the blood that carries oxygen, and it can combine with oxygen to form oxyhemoglobin, when oxygen binds to hemoglobin, the absorption spectrum of hemoglobin changes. Pulse oximeters use this property to measure blood oxygen saturation. A pulse oximeter typically consists of an infrared light source and a light sensor. The infrared light source tends to emit two different wavelengths of light, one wavelength closer to the absorption peak of hemoglobin and the other unaffected by hemoglobin absorption. These two types of light pass through the skin surface into the blood and are then received by the photosensor. When the oxygenated hemoglobin in the blood meets a wavelength close to the light emitted by the infrared light source, it absorbs more light. In comparison, the deoxygenated hemoglobin in the blood absorbs less light. By measuring the difference in the amount of light absorbed by the two wavelengths emitted by the infrared light source, the pulse oximeter can calculate the percentage of oxygenated hemoglobin in the blood and thus obtain a value for oxygen saturation[1].The detection mechanism of the blood oxygen detector and the PPG signal are as shown in Figure 1.

DOI: 10.54254/2755-2721/32/20230195

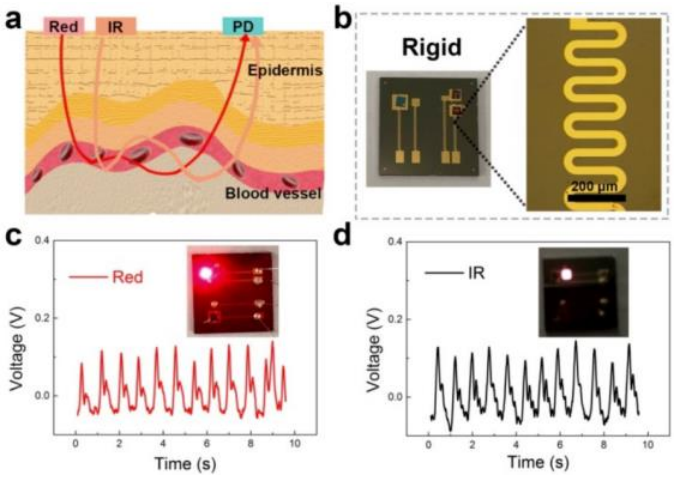


Figure 1. Detection mechanism of the blood oxygen detector [2].

Yuanyuan He, Di Yu, Jiaheng He et al. [3]designed an IoT-based wearable blood oxygen saturation and heart rate real-time monitoring system. Their system collects the photoelectric pulse wave signals detected by optical sensors through a terminal bracelet, and then obtains more accurate heart rate and blood oxygen saturation values after low-pass filtering and dynamic motion compensation processing, and sends them to the cloud platform through the network combined with the positioning information, and manages them using the back-end database, and finally extracts them to the user's cell phone client applet or computer terminal for display. Their system test results show that the network communication error rate of the system is 0; compared with the same type of bracelets, the error of heart rate and blood oxygen saturation detected by the system bracelet is less than 2%, so it has important research significance for real-time monitoring of sports that need high accuracy. Haiming Ai,Fulai Peng, Minlu Hu et al.[4] used a dual-wavelength light emitting diode (LED) as the light source to irradiate the finger in turn, and detected the volumetric pulse wave, pulse rate and oxygen saturation by photodetector of the emitted light. Since the wavelength of the light source is easily shifted by temperature changes and the LED light source is usually not purely monochromatic, they decided to choose the wavelength of the light source in the region of gentle changes in the absorption coefficient of hemoglobin, i.e., 660 nm and 940 nm for red light and near-infrared light, respectively. The AFE4400 series chip with integrated LED driver and signal conditioning circuit was used to improve the system performance, reduce power consumption and design complexity while monitoring the blood oxygen signal in real time. The experimental results show that the correlation coefficient of blood oxygen measurement is 0.995, so it can be considered as a high accuracy. It can be applied to sports real-time monitoring wearable devices.

The future research of blood oxygen saturation monitoring will focus on miniaturization and portability, multi-functional, accuracy improvement and personalized health management. As technology advances, oximeters will tend to be more compact and portable, making it convenient for users to perform oximetry anytime and anywhere. This will promote the popularity of blood oxygen monitoring and the expansion of applications. Future oximeters may integrate more functions, such as heart rate monitoring, sleep quality assessment, and exercise monitoring. This will give users a more comprehensive understanding of their health status and take appropriate measures to improve it. Future blood oxygen detectors may further improve the accuracy and stability of measurement to meet the needs of clinical medicine and health management. Continuous technological innovation and algorithm optimization can improve the accuracy and sensitivity of blood oxygen detection. Future blood oxygen detectors may combine personal health data and lifestyle habits to provide users with personalized health management solutions. More refined and customized health advice can be provided based on individual characteristics and needs through deep learning and other artificial intelligence technologies.

DOI: 10.54254/2755-2721/32/20230195

3. Biosensors used to monitor skin temperature

The flexible temperature biosensor is a flexible and bendable sensor for measuring the temperature of the environment or an object. It is made of a flexible material that is highly deformable and adaptable to contact with curved or curved objects. Flexible temperature biosensors usually use thermosensitive materials or thermistors to measure temperature. The resistance value or other electrical properties of these materials change accordingly to the temperature. The temperature of the environment or object can be calculated by measuring the change in resistance or other electrical properties. The advantages of flexible temperature biosensors are that they can adapt to irregular surfaces and curved shapes, can operate over a wide temperature range and have a high sensitivity and response rate. They are often thin, lightweight, soft and comfortable, and can be applied in contact with human skin or attached to the surface of an object.

There are currently two approaches, the first is to assemble devices using conductors that can be stretched, usually made by mixing a conductive substance into a flexible base [5]. The second is the direct bonding of thin conductive materials with low Young's modulus to a flexible base. From a study by Tan and others, they applied the second method to make biosensors. This biosensor comprises four layers: insulating, sensitive, conductive silver wire, and flexible PET (polyethylene terephthalate substrate). The PET is plasma oxygenated to make its surface rough. Conductive silver gel electrodes are then coated onto the PET substrate, coated with a carbon material suspension, encapsulated with RGO (Reduced Graphene Oxide), and finally an insulating layer . This biosensor has many advantages, such as the good TCR (temperature coefficient of resistance) of the RGO material as a sensitive layer, the good linearity of the temperature-resistance curve in experiments, and the good response time and response time in temperature response tests. Due to the tape press encapsulation technique, the RGO layer spacing is significantly reduced, the sensor's resistance is not affected by the applied stresses and, due to the insulation layer, the external humidity does not affect the resistance or the sensor performance .The detecting process of the PET method is as shown in Figure 2.

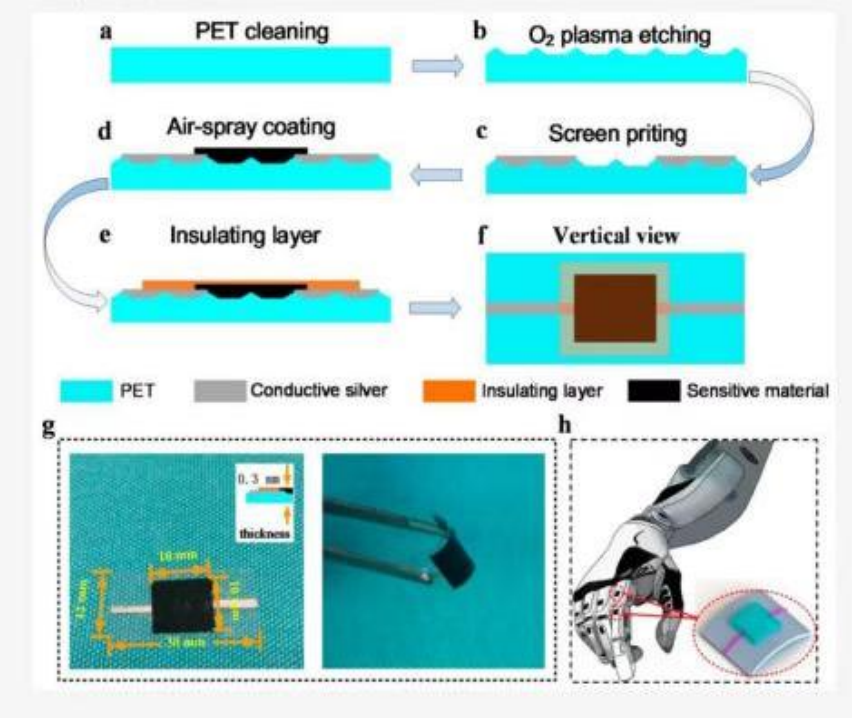


Figure 2. The detecting process of the PET method [6].

Skin temperature biosensor can play an important role in many fields, firstly body temperature is an important health indicator and flexible biosensors are so sensitive to changes in temperature that they

DOI: 10.54254/2755-2721/32/20230195

can capture subtle temperature changes that allow doctors to assess changes in the physiological characteristics of the human body, such as the thermal conductivity of human tissues, skin water content, blood flow and wound repair processes [7]. For example, during the new coronavirus outbreak in 2020-2022, doctors could use flexible sensors to monitor a patient's temperature and other health indicators in real time without disturbing the patient's rest.

In sports, wearable flexible biosensors also have an important role to play. Flexible biosensors can collect physiological signals from all directions, at multiple levels and from multiple angles, during the athletes' movement to build a database of high-level athletes, and through a large amount of data analysis, monitor the athletes' physiological indicators in real-time, reasonably predict the potential risk of injury during the athletes' training, reduce the athletes' risk of injury This will help to reduce the risk of injury and to develop a more scientific training plan for athletes [8]. Under normal physiological conditions, body temperature can be affected by day and night, age, gender, environment, temperature mental and physical exercise conditions, but body temperature but fluctuates little, so if the body temperature if increased will change the body's ability to exercise. When exercising in a thermal ring, a sharp rise in body temperature increases the probability of shock, and the brain is very susceptible to damage when the body temperature is too high[9]. Damage to the brain caused by hyperthermia may be one of the causes of heat stroke[10] . High body temperatures reduce the body's ability to perform long-term work and can cause significant damage to the cardiovascular system. As a result of dynamic exercise in a hot environment, the cardiovascular system is subjected to heavier load loads, such as accelerated blood flow through the skin to disperse heat and accelerated oxygenation of the muscles to fill their new age. As peripheral blood circulation increases, blood volume increases, leading to a decrease in intracardiac blood volume, a decrease in stroke output and an increase in compensatory heart rate [11]. Therefore, using flexible temperature sensors for real-time body temperature monitoring is essential. Many consumer electronic products now use this technology of flexible temperature sensors, such as sports bracelets and watches, which allow more sports enthusiasts and ordinary consumers to enjoy the benefits of technology for life and health.

4. Biosensors used to monitor heart rate

Wearable heart rate monitors have changed the way we monitor and track cardiovascular health. These devices detect and measure the electrical activity of the heart and are called heart rate monitors. The wearable heart rate detector uses sensors in direct contact with the skin and various methods to record and analyze heart rate data.

The most widely used detection technique today is photodensitometry (PPG), which uses light-emitting diodes (leds) and photodiodes to monitor changes in blood volume. The amount of light absorbed by the blood and the blood flow in the capillaries under the skin change with the heartbeat. These changes are detected by the wearable's photodiode, which then converts them into electrical pulses. These signals are then processed and used to calculate heart rate. ppg based heart rate monitors are widely used in fitness trackers and smartwatches due to their non-invasive nature and ease of use.

The electrocardiogram (ECG) is another instrument that is becoming increasingly popular. An electrocardiogram measures the electrical impulses produced by the heart. In an ECG based heart rate monitor, electrodes in the wrist or chest capture electrical pulses. Electrodes can detect and amplify the smallest electrical changes on the skin's surface. However,

The enhanced signals were then analyzed to accurately calculate the heart rate. Due to their high accuracy, ECG based heart rate monitors are often used in medical Settings and professional sports.

Advanced wearables may include additional accelerometers or gyroscopes to improve the accuracy of heart rate monitoring and account for motion distortion. These sensors detect and quantify movement, allowing the device to filter out noise caused by movement and provide more accurate heart rate measurements. These wearable heart rate trackers provide reliable data even during intense exercise by calculating exercise artifacts, making them beneficial for fitness enthusiasts and sports. Aiming at the low accuracy of heart rate measurement in current wearable heart rate detection devices under exercise conditions, a deep learning algorithm combining convolutional neural network with

DOI: 10.54254/2755-2721/32/20230195

sequence-to-sequence network (CNN-seq2seq) was proposed to extract heart rate in photoplethysmograph (PPG) under exercise conditions Method of value. Combined with the features of convolutional neural network in feature extraction and the advantage of long short-term memory network in time series data processing, a network model of convolutional neural network combined with sequence-to-sequence + attention mechanism is established. Methods The PPG signals of 30 healthy subjects were collected at rest, walking, jogging and fast running, and their electrocardiogram (ECG) signals were collected synchronously by an electrocardiogram device with anti-interference capability. The PPG signals were used as neural network input signals, and the ECG signals were simplified and retained The CNN-seq2seq network is then trained, and the network outputs PPG-like signals with accurate heart rate characteristics, so as to achieve heart rate measurement under exercise conditions. CNN-seq2seq network output and the corresponding ECG signal were used to calculate the heart rate per minute. The mean error and mean square error of heart rate estimation were 0.25±1.31. The experimental results show that CNN-seq2seq network model can obtain ideal results for the prediction of exercise heart rate. This provides a feasible scheme for portable measurement of exercise heart rate[12].

Wearable heart rate monitors have a wide range of uses. These gadgets allow people to track their heart rate while exercising, provide accurate data and recommendations on their level of physical activity, and enhance their exercise habits in the areas of fitness and sports. To achieve the best results, athletes can monitor their performance and adjust their intensity by focusing on their target heart rate zone. Fitness enthusiasts can also set goals, monitor their progress, and make clearer decisions about what workouts to do.

Medical organizations are also finding important uses for wearable heart rate monitors. A good example of its use in practice is remote patient monitoring. Patients can continuously monitor their heart rate with the convenience of a heart rate monitor at home. This allows doctors to remotely assess a patient's cardiovascular health, identify any abnormalities or irregularities, and seek immediate medical assistance if necessary. This remote monitoring enhances patient convenience while also enabling proactive medical care and reducing the frequency of hospital visits. [13] Heart rate refers to the number of beats of the heart in 1min under peaceful conditions, and the changes of heart rate and other parameters can reflect the operation of various functions of the human body. This paper designs a single chip microcomputer heart rate detection system whose core component is STC89C52. By means of different electrical signals generated by photoelectric sensors, the electrical signals are sent to the single chip microcomputer for analysis and processing, so as to measure the user's heart rate indirectly. This system can improve the portability and accuracy of the heart rate detector, and reduce the cost.

In short, a wearable heart rate device can detect and measure the electrical activity of the heart. With technologies such as PPG and ECG, these devices can accurately and consistently measure heart rate. The practical application of wearable heart rate monitors has evolved into an important tool for enhancing cardiovascular health and well-being, from fitness tracking and exercise performance improvement to remote patient monitoring and proactive healthcare.

5. Conclusion

This study found that optical sensing technology is the most commonly used detection technology when monitoring blood oxygen saturation. Its sensors usually use infrared and red light through the skin to measure oxygen saturation in the blood. In recent years, researchers have improved the measurements' accuracy and stability by improving the sensors' design and algorithms. To improve the accuracy of the oximetry measurements, the researchers also optimized the algorithms of the sensors. Using machine learning and deep learning techniques, sensor data can be analyzed and processed more accurately, reducing errors and providing more accurate results.Common biosensing technologies used in monitoring skin temperature include thermistors, infrared, and thermocouples. Thermistors are among the most commonly used technologies, which utilize the change in resistance of a material with

DOI: 10.54254/2755-2721/32/20230195

temperature to measure skin temperature. Infrared technology, on the other hand, estimates skin temperature by measuring infrared radiation from the skin surface.

On the other hand, thermocouple technology utilizes the voltage difference created by the difference in the conductive ability of two different materials to measure temperature. Skin temperature monitoring wearable sensors have a wide range of applications in medical, sports, and sleep monitoring areas. Especially in sports, real-time temperature monitoring can be a good way to prevent athletes from becoming overdrawn, thus avoiding injuries and illnesses, and can provide athletes with personalized exercise and rest recommendations based on real-time body temperature changes.For monitoring heart rate, the biosensors they use usually employ electrodes in contact with the skin to measure ECG signals to calculate heart rate. This technology has a high degree of accuracy, but requires full contact with the skin and therefore may be disturbed during exercise. In recent years, heart rate monitoring sensors during exercise are usually connected to a smartphone or other device that can monitor heart rate data in real time and analyze and display the data through an app. This allows users to keep track of their heart rate changes and conduct exercise training and health management based on the data.In general, researchers at home and abroad have made a lot of progress in the study of sensors for monitoring physiological parameters during exercise. These researches are significant for improving people's sports health and monitoring their physical conditions. However, further research and development are still needed to improve the sensors' accuracy, comfort, and reliability to meet the needs in different exercise scenarios.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Zhang L. 2021. Chongqing University of Technology.
[2] Huamin C,Yun X, Jiushuang Z... & Guofeng S. 2019. 5.
[3] Yuanyuan H, Yu D, Jiaheng H. et al. 2023.*Internet of things technology*,**13(02)**:59-62+65.
[4] Ai H-M, Peng F-L, Hu M-L, et al. 2020 *Sensors and Microsystems*, **39(07)**:92-94+97.
[5] Lee P; Lee J; Lee H; Yeo J; Hong S; Nam K. H; Lee D; Lee S. S; Ko S. H. Adv. Mater. 2012, **24**, 3326.
[6] Liu G, Tan Q, Kou H, Zhang L, Wang J, Lv W, ... & Xiong J. 2018. *Sensors*, **18(5)**, 1400.
[7] Hattori, Y.; Falgout, L.; Lee, W.; Jung, S. Y.; Poon, E.; Lee, J. W.; Na, I.; Geisler, A.; Sadhwani, D.; Zhang, Y. H.; Su, Y. W.; Wang, X. Q.; Liu, Z. J.; Xia, J.; Cheng, H. Y.; Webb, R. C.; Bonifas, A. P.; Won, P.; Jeong, J. W.; Jang, K. I.; Song, Y. M.; Nardone, B.; Nodzenski, M.; Fan, J. A.; Huang, Y. G.; West, D. P.; Paller, A. S.; Alam, M.; Yeo, W. H.; Rogers, J. A. 2014 *Adv. Healthcare Mater.* , **3**,1597.
[8] Bingtian S, Jianliang L, Huihua X, Ze X, Jianxin M, Xiaoping C,and Fengyu L.2022. *Science China: Information Science*, **52**, 54-74.
[9] Brinnel H, CabanacM, Hales YRS. 1987,*Amsterdam: Excerpta Madica*, 209-240.
[10] Hales JRS, Hubbard RW, Gaffin SL.1996,New York: Oxford Universitv Press. 285-355.
[11] Rowell LB. 1986,NewYork:Oxford University Press,363-406.
[12] Kai Q, Xu Z, Jiaqi G & Junfeng G.2021. *Journal of South-Central University for Nationalities*, **05**, 489-495.
[13] Guangjing Z, Ming Z, Youhao Z, & Mengyao X. 2021. *Intelligent computer and application*, **5**, 4.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230196

Biosensors for ocean acidification detection

Ziyi Guo

College of Chemical Engineering, HuaQiao University, Xiamen, China

1910732115@mail.sit.edu.cn

Abstract. Ocean acidification is a global environmental problem that significantly impacts Marine ecosystems and biodiversity. The traditional chemical analysis method has the problems of complex equipment and high cost in ocean acidification monitoring. In recent years, fluorescent protein biosensor technology, as an innovative monitoring method, has provided a new solution for the real-time detection of ocean acidification. Compared with traditional chemical analysis methods, fluorescent protein biosensors have the advantages of simple operation, high sensitivity and low cost. Current studies have demonstrated the potential of fluorescent protein biosensors for ocean acidification monitoring. The researchers designed a variety of fluorescent protein biosensors and conducted indoor and outdoor experimental validation. These results show that fluorescent protein biosensors can detect ocean acidification quickly and accurately and maintain stable performance under different environmental conditions. Further studies are needed to verify the consistency and reliability of fluorescent protein biosensors and traditional chemical analysis methods for ocean acidification monitoring. Future research directions include further improving the performance of the fluorescent protein biosensor, increasing its sensitivity and stability, and verifying its application in real Marine environments. This will help establish a better monitoring network for ocean acidification and provide a reliable scientific basis for Marine environmental protection and management decisions. The development and application of fluorescent protein biosensor technology will provide important support and guidance for us to better understand the impact of ocean acidification.

Keywords: ocean acidification, biosensors, fluorescent proteins.

1. Introduction

Ocean acidification is one of the global environmental problems that has attracted much attention in recent years. With the continuous increase of carbon dioxide emissions from human activities, a large amount of carbon dioxide is dissolved in seawater to form carbonic acid, resulting in a continuous decline in the pH of the ocean [1]. This acidification phenomenon has a broad and profound impact on Marine ecosystems and biodiversity.

To monitor and assess the extent of ocean acidification, traditional chemical analysis methods are widely used, but these methods often require complex experimental equipment and expensive costs, and there are uncertainties in the sampling and analysis process [2]. Therefore, finding new, fast, accurate and cost-effective monitoring methods is very important [3].

In recent years, fluorescent protein biosensor technology has become a popular method [4]. This technique uses fluorescent proteins as markers to detect the pH of the environment in real time by

DOI: 10.54254/2755-2721/32/20230196

monitoring changes in fluorescence properties. Using the unique properties of fluorescent proteins, biosensors can provide immediate and quantitative measurements, opening up new possibilities for ocean acidification monitoring and research [5].

This paper reviews the application potential and related research progress of fluorescent protein biosensors in ocean acidification monitoring. First, the impact of ocean acidification on ecosystems and biodiversity will be briefly described. Then, the principle and design method of fluorescent protein biosensor are discussed. Finally, the existing research results of ocean acidification biosensors will be reviewed, and the future development direction and challenges will be discussed.

Through in-depth understanding and research of fluorescent protein biosensor technology, we are expected to develop a new, simple and efficient monitoring method for ocean acidification, and provide more accurate and reliable data support for protecting the Marine ecological environment and formulating corresponding protection strategies.

2. Algae sensors

Certain algae are very sensitive to ocean acidification, and their physiological responses are closely related to environmental pH. By monitoring changes in fluorescence or other metabolic substances inside the cells or on the surface of the leaves of these algae, changes in pH in the ocean can be measured indirectly [6]. This algae-based sensor has high sensitivity and real-time monitoring capability.

2.1. Working principle

Algae sensors reflect the pH of the surrounding environment based on changes in fluorescence or other metabolic substances inside the algae cells or on the surface of the leaves. Ocean acidification can lead to changes in the acid-base balance within cells, affecting the algae's physiological state and metabolic processes. Sensors can infer pH changes in the ocean by monitoring changes in these physiological parameters.

2.2. Technical possibilities

2.2.1. Using bioluminescence. a common algal sensor uses algae's fluorescent properties to measure environmental pH. these algal sensors work by introducing specific fluorescent markers or fluorescent proteins into the algal cells, and the intensity of the fluorescent signal changes when the cells are subjected to a specific pH level. By measuring the intensity of the fluorescent signal, the pH change of the environment can be measured indirectly. Algae biosensors have the ability to monitor environmental parameters in real time, and can quickly respond to target molecules or environmental changes in a short time to achieve real-time data acquisition. This capability makes it important in environmental monitoring and pollution warning applications [3].

2.2.2. Versatility and customizability. Algae biosensors can be optimally designed and engineered for specific target molecules to meet specific detection and monitoring needs. By analyzing the influence mechanism of specific target molecules on algae organisms, the selection of biological elements and parameter regulation can be optimized to improve the sensitivity and specificity of the sensor. Algae can sense and respond physiologically to environmental changes such as pH and carbon dioxide concentration, so that environmental parameters can be measured by monitoring their physiological parameters or metabolites

2.2.3. Selection of sensitive algal species. Algal biosensors enable accurate detection of specific molecular and environmental parameters based on the selectivity and sensitivity of algal organisms to target substances. Algae can sense and respond physiologically to environmental changes such as pH and carbon dioxide concentration, so environmental parameters can be measured by monitoring their physiological parameters or metabolites [4].

DOI: 10.54254/2755-2721/32/20230196

2.3. Application prospects

Algal sensors have the advantages of high sensitivity, real-time, non-invasive and low cost. These sensors can be used in ocean acidification monitoring, marine protected area management, environmental research, and other fields. In addition, algal sensors can be combined with other monitoring technologies and sensors to provide more comprehensive monitoring and assessment of ocean acidification.

3. Shellfish sensors

Shellfish can reflect changes in environmental pH through changes in their shells' oxygen isotope ratios and carbon isotope composition. These sensors can measure the degree of ocean acidification by analyzing isotopic ratios in the shells or molluscan tissues of shellfish. Shellfish sensors have high accuracy and stability and can be used for long-term monitoring [6].

3.1. Principle of operation

Shellfish sensors use changes in oxygen isotope ratios and carbon isotope compositions in shells or molluscan tissues to indirectly measure changes in environmental pH. The pH of the environment influences the chemical composition of shells and molluscan tissues of shellfish growth. Ocean acidification leads to an increase in the solubility of carbon dioxide in water, causing changes in the carbon isotopic composition of *shells and tissues of shellfish.*

3.2. Technical possibilities

3.2.1. Isotopic analysis. Shellfish sensors infer changes in environmental pH by analyzing the carbon isotopic composition in shells or tissues of shellfish. The oxygen isotopic composition influences the oxygen isotopic composition in shellfish shells in the surrounding water. In contrast, the carbon isotopic composition is related to the concentration of carbon dioxide in the surrounding water column. Changes in the degree of acidification in the environment can be inferred from measurements of isotopic ratios in shells or tissues of shellfish [6].

3.2.2. Sample collection and analysis. Performing shellfish sensor analysis requires collecting shell or mollusk tissue samples of shellfish and performing appropriate chemical processing and isotopic analysis. The isotopic content of shellfish shells is usually measured by mass spectrometry techniques (e.g., stable isotope mass spectrometry).

3.2.3. Application areas. Shellfish sensors can be applied to the monitoring and assessing ocean acidification. The impact of acidification on shellfish growth and ecosystems can be understood. Shellfish sensors are widely used in protecting and managing marine ecosystems, assessing the risk of regional ocean acidification, and developing related policies.

3.3. Application prospects

Shellfish sensors still face some challenges in practical applications, such as the difficulty of sample collection, complexity and accuracy of analysis. In addition, standardized methods and further development of data analysis techniques are the future directions of shellfish sensor research [7]. However, shellfish sensors have great potential in ocean acidification research as a non-invasive and sustainable monitoring method.

4. Coral sensor

Using corals as biosensors to detect ocean acidification is a very promising area of research. Corals are very sensitive creatures in Marine ecosystems, and they are very sensitive to changes in the Marine environment, especially changes in pH [8]. Ocean acidification is caused by increased carbon dioxide

DOI: 10.54254/2755-2721/32/20230196

in the atmosphere, which dissolves in seawater to form carbonic acid, causing the pH of seawater to drop.

Using corals as biosensors can assess ocean acidification by observing how corals respond to ocean acidification. A common approach is to look at physiological indicators such as coral growth rate, bone morphology and chemical composition [9]. Ocean acidification directly impacts the physiological processes of corals, so it is possible to assess the degree of acidification of seawater by monitoring these indicators.

In addition, using coral gene expression can also be used as a biosensor to detect ocean acidification. Ocean acidification causes changes in many genes in the coral genome that can act as acidification indicators. Changes in carbon dioxide concentrations in seawater can be assessed by analyzing changes in coral gene expression.

It is important to note that using coral as a biosensor is still in the research stage, with many challenges and limitations. For example, accurately interpreting corals' physiological indicators and gene expression changes needs more in-depth research. In addition, interference from other environmental factors also needs to be considered. However, harnessing the potential of corals as biosensors could provide important information to better understand the effects of ocean acidification.

4.1. How it workss

Coral sensors use the physiological and chemical response of corals to reflect changes in environmental conditions. The physiological processes of corals are affected by environmental pH, and their growth rate, composition of skeletal compounds, and oxygen isotope composition may change [10]. By analyzing these characteristics, it is possible to infer the degree of acidification and health of the environment to which the coral is exposed.

4.2. Technical possibilities

4.2.1. Skeletal compound analysis. The skeletons of corals are important indicators for recording environmental changes. Coral skeletal compounds (e.g., calcareous structures) are closely related to the environment's pH and carbon isotopic composition. Analysis of composition, microstructure and isotopic composition in coral skeletons provides insight into the effects of ocean acidification and other environmental changes on corals [11].

4.2.2. Growth rate monitoring. environmental acidification may affect the growth rate of corals. By monitoring changes in coral growth rates, changes in pH of the surrounding water column can be inferred. The calcium-carbon balance involved in corals has been used in some studies to estimate changes in environmental pH [12].

4.2.3. Temperature and light analysis. Coral sensors can also monitor temperature and light changes. Ambient temperature and light conditions impact corals' physiological processes and their relationship with symbiotic algae. Analysis of coral temperature and light response can provide insight into other environmental changes that coexist with environmental acidification.Coral biosensor can realize real-time monitoring and rapid response of environmental parameters [13]. Corals have high sensitivity and response speed to environmental changes. They can respond to external environmental changes in a short time, so as to achieve real-time data collection and monitoring.

4.3. Application prospects

Coral sensors can be used in ocean acidification monitoring, coral health assessment, climate change research and environmental monitoring. These sensors provide real-time monitoring and assessment of coral ecosystem health and environmental stress.

DOI: 10.54254/2755-2721/32/20230196

5. Conclusion

Ocean acidification biosensor technology is a promising method to monitor ocean acidification. Using fluorescent protein as a marker, the technology can accurately monitor ocean pH changes and provide fast, real-time measurement results. Fluorescent protein biosensors have the following advantages: easy to use, high sensitivity, and low cost. By designing and optimizing fluorescent proteins, more accurate and reliable monitoring of ocean acidification can be achieved.

Current studies have demonstrated the potential of fluorescent protein biosensors in ocean acidification monitoring. Researchers have successfully designed a variety of fluorescent protein biosensors and conducted indoor and outdoor experimental validation. These results show that fluorescent protein biosensors can detect ocean acidification quickly and accurately and work stably under different environmental conditions.

However, fluorescent protein biosensor technology still faces challenges, including selecting suitable fluorescent proteins, optimizing sensor design, and handling complex environmental samples. In addition, for ocean acidification monitoring, further research is needed to verify the consistency and reliability of fluorescent protein biosensors with traditional chemical analysis methods.

In the future, continuing to advance the research and development of fluorescent protein biosensor technology will help better understand ocean acidification's impact and provide accurate data support for Marine environmental protection and management decisions. Further optimizing the performance of fluorescent protein biosensors, improving their sensitivity and stability, and carrying out application validation in real Marine environments will be the focus of future research. This will help build a more complete monitoring network for ocean acidification and provide a stronger scientific basis for ocean protection and sustainable development.

References

[1] Matoo O B.Dissertations & Theses - Gradworks, 2013.
[2] Doney S C, Fabry V J, Feely R A, & Kleypas, J. A. *Annual Review of Marine Science*, 2009.
[3] Kilinc S, Alpat S, Kutlu B, Oezbayrak O, Bueyuekisik H B. *Sensors & Actuators B Chemical*, p273-278.
[4] Alexander and J. Wiest, 2015, pp.
[5] ZHANG Chenglong, HUANG Hui, HUANG Liangmin, *Acta Ecologica Sinica*, 2012 p 1606-1615.
[6] Przeslawski R, Ahyong S, Byrne M ,et al.*Global Change Biology*, 2008 p 14(12).
[7] Angewandte Chemi.*Journal of Chinese Pharmaceutical Sciences*,2022 p 558-560.
[8] Guinotte J M, Fabry V J.*Annals of the New York Academy of Sciences*, 2008.
[9] Crook E D.*Izvestiya Vysshikh Uchebnykh Zavedenij Chernaya Metallurgiya*, 2013 p 57(6).
[10] Habicht K A. 2018.
[11] Brachert, Thomas C. Correge, ThierryReuter, MarkusWrozyna, ClaudiaLondeix, LaurentSpreter, PhilippPerrin, Christine.*Earth-Science Reviews: The International Geological Journal Bridging the Gap between Research Articles and Textbooks*, 2020 p 204(1).
[12] PASCALPANDARD, PAULEMASSEUR, Rawson D M. *Water Research*, 1993 p 427-431.
[13] Renneberg Y R. *Biosensors and Bioelectronics*, 2005.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230199

Enhanced diffusion model based on similarity for handwritten digit generation

Wenjing Kang[1,†] and Wenbo Li [2,3,†]

[1]School of electronic engineering, Xi'an University of Posts and Telecommunications, Xi'an, China
[2]College of computer science and cyber security, Chengdu University of Technology, Chengdu, China

[3]li.wenbo1@student.zy.cdut.edu.cn
[†]All the authors contributed equally and their names were listed in alphabetical order.

Abstract. In recent years with the rise of deep learning, there has been a major revolution in image generation technology. Deep learning models, especially the diffusion model. have brought about breakthrough progress in image generation. Various deep generation models have recently demonstrated a wide variety of high-quality sample data patterns. Although image generation technology has achieved remarkable achievement. There are still challenges and issues, such as quality control in generated images. In order to improve the robustness and performance of diffusion model in image generation, an enhanced diffusion model based on similarity is proposed in this paper. Based on the original diffusion model, the similarity loss function is added to narrow the semantic distance between the original image and the generated image, so that the generated image is more robust. Extensive experiments were carried out on the MINIST dataset, and the experimental results showed that compared with the other generation models, the enhanced diffusion model based on similarity obtained the best scores of IS=31.61 and FID=175.21, which verified the validity of the similarity loss.

Keywords: diffusion model, handwritten digit, similarity loss, generation.

1. Introduction

Image generation refers to the process of using artificial intelligence technology to generate images in single mode or cross-mode according to given data. According to different task objectives and input modes, image generation mainly includes image composition, image-to-image generation based on existing images, and text-to-image generation based on text description [1]. It is widely used in graphic design, game production, animation production and other fields. In addition, image generation also has great application potential in medical image synthesis and analysis, compound synthesis and drug discovery. Therefore, image generation has attracted more and more researchers' attention.

In order to realize image generation, many efficient generation models have been developed in recent years, such as generative adversarial model and autoregressive generation model [2]. Generative adversarial network (GAN) is the mainstream image generation model of the last generation. GAN continuously improves its generative ability and discrimination ability through game training of generator and discriminator, so that the data of generative network is more and more close to the real

DOI: 10.54254/2755-2721/32/20230199

data, so as to achieve the purpose of generating realistic images. However, in the process of development, GAN also has some problems, such as poor stability, lack of diversity of generated images, and mode collapse. The inspiration of autoregressive model for image generation comes from the successful experience of NLP pre-training mode. The self-attention mechanism in Transformer structure can optimize the training mode of GAN, improve the stability of the model and the rationality of image generation. However, the image generation based on autoregressive model has problems in reasoning speed and training cost. Make its practical application limited.

Due to the limitations of the previous Model in terms of performance, the Diffusion Model was proposed to solve these problems, and its effect on training stability and result accuracy was significantly improved, so it quickly replaced the application of GAN. The diffusion model is the process of gradually applying noise to the image in the forward stage until the image is destroyed into complete Gaussian noise, and then learning to restore the original image from Gaussian noise in the reverse stage. The diffusion model can restore the real data more accurately, and the processing ability of image details is stronger. However, the classical generation model only considers the noise loss in the process of diffusion, and does not consider the similarity and semantic consistency between image contents.

In order to ensure the quality and stability of the generated images, an enhanced diffusion model based on similarity is proposed in this paper. On the basis of the classical diffusion model, the semantic similarity between the original image and the generated image is described by comparing them, and it is taken as a part of the loss function and noise loss to guide the model optimization [3].

2. Related works

In fields including text-to-image translation and image creation, diffusion models (DMs), a revolutionary generative model based on deep learning and computational vision, have been put to use. Diffusion models have a number of benefits over conventional autoregressive models, energy-based models, and generative adversarial network (GAN) models, including the ability to create pictures with substantial variety and large details [4].

In Chen Li's Comparison of Image Generation methods based on Diffusion Models, IDDPM model is put forward (Improved Denoising Diffusion Probabilistic Models), By defining the goal function and enhancing the calculation's logarithmic likelihood function, which is used to compute variance variance learning, and lowering the degree of difficulty of the sampling step, accelerated sampling. The performance boost is modest, the Markov process is still used, it needs more processing power, and the sample steps are longer [5].

Therefore, Chen Li proposed a de-noising diffusion implicit generation model (DDIM) for effective sampling. This model is an implicit generation model that relies on edge distribution, so it only needs to let the sampling process meet the edge distribution conditions, rather than relying on Markov random process, so it does not need many sampling steps to get high-quality image samples, and the speed is faster [5].

At the same time, the loss function of the diffusion model most often adopts the simplified optimization objective based on the predicted noise. But there are other options, and prediction targets can be constructed based on raw data x0. In addition to the prediction target of the model, the loss function can also adopt different weight coefficients, which has a certain impact on the training of the diffusion model.

In the Progressive Distillation for Fast Sampling of Diffusion Models proposed by Tim Salimans [6]. the loss function based on the original data and the fitting data is adopted. SNR+1 and truncated SNR are used as the weight coefficients. The truncated SNR weight coefficient is designed to prevent the weight coefficient from being 0 when the SNR is close to 0, which is not conducive to distillation.

In addition, another weight coefficient, min-SNr-γ, is proposed by Efficient Diffusion Training via Min-SNR Weighting Strategy [7]. The main purpose of this paper is to avoid paying too much attention to the small noise level (that is, the number of diffusion steps t is small) during model training. One of its advantages is to accelerate the training process.

DOI: 10.54254/2755-2721/32/20230199

3. Method

3.1. Classical diffusion model

The Diffusion Model is a type of Generative model, which also includes the Variational Autoencoder (VAE) and the Generative Adversarial Network (GAN). Unlike other generative networks like GAN, the diffusion model progressively introduces noise to an image during the forward stage until the image is completely scrambled into Gaussian noise. It then learns to reconstruct the original image from this Gaussian noise during the reverse stage [8].

In particular, the forward stage progressively adds noise to the initial image x_0. The image x_t produced at each step is solely influenced by the result x_{t-1} from the preceding step until the image x_t at step T transforms into pure Gaussian noise [9-10].

The reverse stage involves the continuous process of noise reduction. Initially, Gaussian noise x_T is provided and gradually de-noised until the original image x_0 is completely restored. This process is guided by a loss function, as shown in equation (1).

$$Loss = E_{x_0,\varepsilon}\left(\left\|\varepsilon - \varepsilon_\theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\varepsilon, t)\right\|^2\right) \tag{1}$$

x_0 is the original image,ε is the real nois,ε_θ is the predicted noise,t represents the time step, $\bar{\alpha}_t = \sum_{s=0}^{t}\alpha_t$, $\alpha_t = 1-\beta_t$, $\beta_t$represents the variance of the Gaussian noise at time *t*.

3.2. Diffusion model based on similarity loss

In the traditional diffusion model, the reduction of noise loss equates to eliminating noise and enhancing the image, concentrating more on the clarity and reproducibility of the produced image. If the created image is noisy, discerning its content becomes challenging, thereby limiting its generalization capability. To boost the generalization and robustness of the produced model, it is crucial that the desired content remains clearly identifiable even when the image generation results are subpar. Hence, this paper emphasizes the semantic representation of the produced and original images, aiming to achieve maximum similarity in the semantic space. This approach ensures that a discernible target image can still be obtained even if the image is blurry.

With this in mind, this paper introduces an improved diffusion model that leverages similarity to generate more stable and realistic images. Unlike the traditional diffusion model that only concerns with noise or data, this paper develops loss functions that consider both noise and data as predictive targets. The model's structure is depicted in Figure 1.

The model is bifurcated into two phases: the diffusion generation phase and the similarity comparison phase. In the diffusion phase, noise is introduced and eliminated through the forward and reverse processes, culminating in the generation of the target image. Following that, in the similarity comparison phase, the semantic similarity between the created image and the original image is computed. The model's training is directed by these two phases.

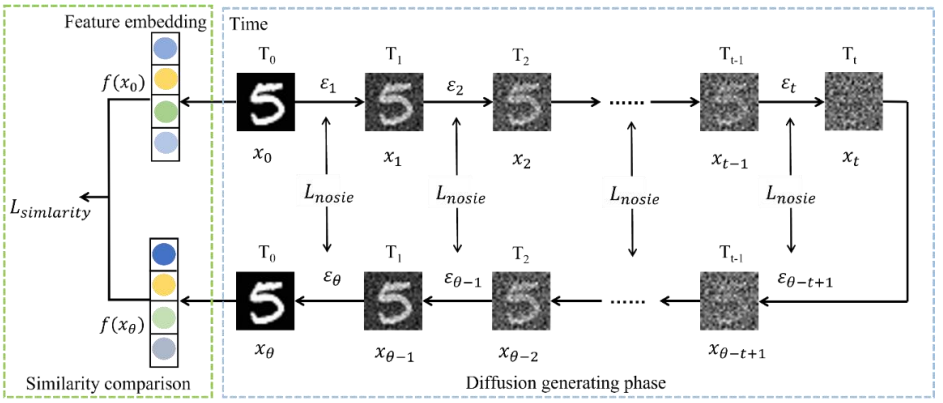


Figure 1. Enhanced diffusion model based on similarity.

DOI: 10.54254/2755-2721/32/20230199

Specifically, the loss function of the similarity-based enhanced diffusion model consists of two parts, namely, noise loss and similarity loss.

As shown in formula (2), noise loss follows the loss function of the classical diffusion model and takes noise as the prediction target.

$$L_{noise} = E_{x_0,\varepsilon}[\|\varepsilon - \varepsilon_\theta(x_t, t)\|^2] \tag{2}$$

The similarity loss, as depicted in formula (3), characterizes the content resemblance between the original and generated images. This paper ensures the semantic uniformity between the original and generated image by evaluating the quality of the produced image through the computation of the cosine similarity between the features of the two images.

$$L_{similarity} = \frac{|<f(x_0), f(x_\theta)>|}{|f(x_0)| \cdot |f(x_\theta)|} \tag{3}$$

<·>Represents inner product operation,f (·) represents feature extraction,|·|represents module.

The final optimization objective is shown in formula (4).

$$L = L_{noise} + L_{simliarity} \tag{4}$$

4. Experiment

In this section, in order to verify the effectiveness of the proposed model, we will demonstrate the improvement of the model's performance by comparing experimental data analysis and generating visual perception of images.

4.1. Experimental settings

The experiment was conducted in Ubuntu20.04, the programming language was Python3.9, the deep learning framework of the experiment was Pytorch2.0, CUDA12.1, the CPU processor was i9-13900KF, and the graphics card was 4090.

All experiments will be trained and tested on the MINIST dataset. We set the number of training rounds for all experiments to 50, the time step T to 500, the Batchsize to 256, the optimizer to adopt the Adam algorithm, and the learning rate to 0.0005. We set the forward process variance consistent with DDPM, increasing linearly from 0.0001 to 0.02. In the direction process, the Unet network structure is used as the common denoising structure. Unet takes the image as the entry point, finds the low-dimensional representation of the image by reducing the sampling, and then restores the image by increasing the sampling.

In order to verify the generalization ability of the model, all performance indicators will be calculated on the test set.

4.2. Analysis of experimental results

Figure 2 shows the results generated by the enhanced diffusion model based on similarity after training on the MNIST dataset. The figure shows the process of the image gradually becoming an image from random noise. It can be seen from the figure that the enhanced diffusion model based on similarity can generate high-quality clear pictures, only some numbers are unrecognizable. In the process of batch generation, it can be found that the number "7" generates the least amount, which may be caused by the unbalanced number of samples in the MINIST dataset. Some numbers have more samples than others, which can cause the model to be biased toward producing numbers that occur frequently.

DOI: 10.54254/2755-2721/32/20230199

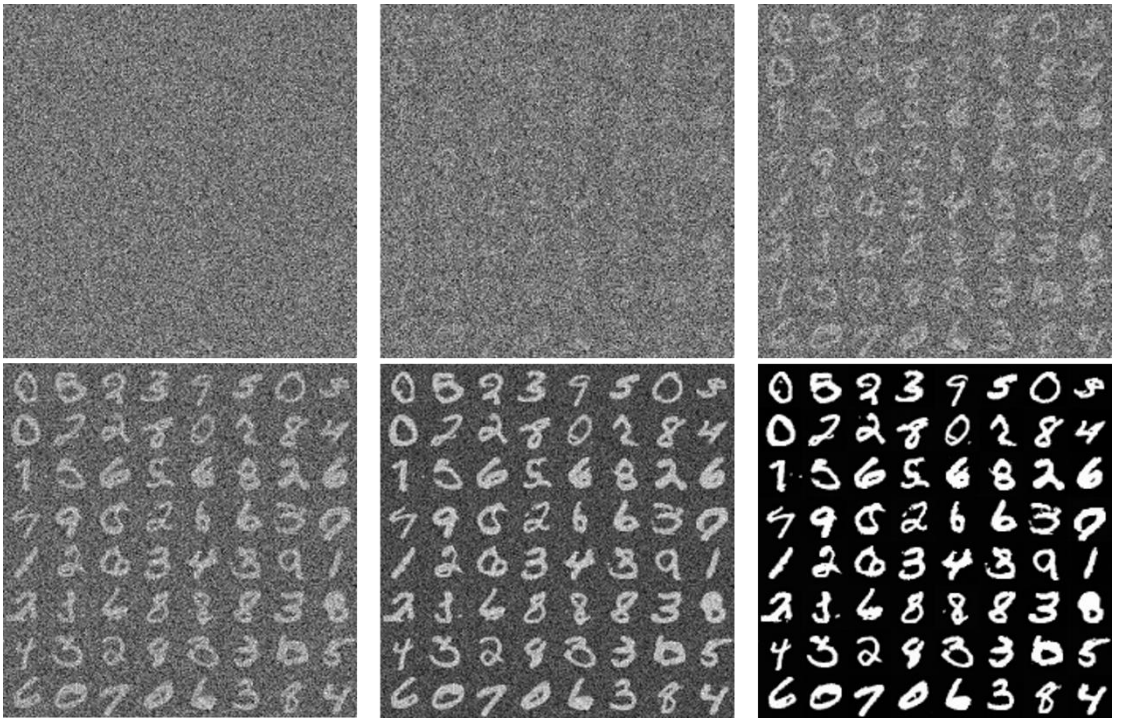


Figure 2. Image generation process of enhanced diffusion model based on similarity.

4.3. Contrast experiment

First, compared with the most basic diffusion model (the diffusion model with only noise loss function), the model performance before and after adding the similarity loss is compared. Secondly, the classical generation model is selected for comparison. The IS (Inception Score) and FID (Frechet Inception Distance score) were used as performance evaluation indexes. It IS worth noting that the IS and FID are calculated from the test set. The comparison results are shown in Table 1.

The following conclusions can be drawn from Table 1.

1) The enhanced diffusion model FID and IS based on similarity have achieved the best results in comparison.

2) IS measures the sharpness and variety of the generated image, while FID measures the distance of the generated image from the original image. The experimental results show that the FID value of the model proposed in this paper IS 31.61, and the IS value is 175.21. After adding the similarity loss to the enhanced diffusion model based on similarity, both FID and IS performance indicators have improved to some extent. Therefore, the addition of similarity loss is effective to improve the performance of the model, and can generate more realistic and clear images to a certain extent.

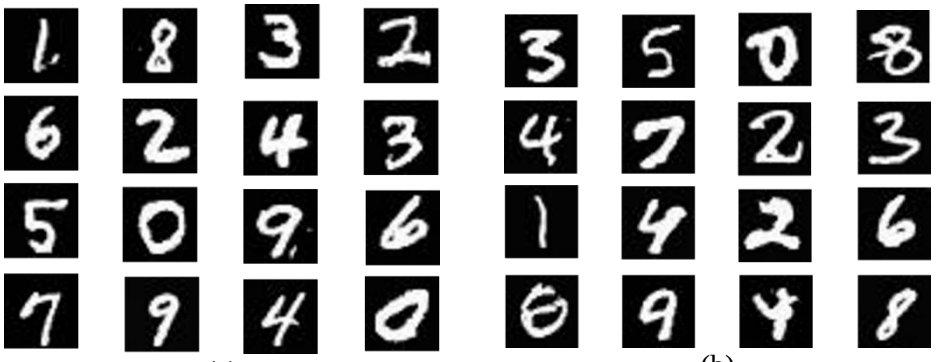
3) In comparison with other types of generative models, the enhanced diffusion model based on similarity also shows better performance and ranks first in the overall ranking.

Table 1. Comparative experimental results.

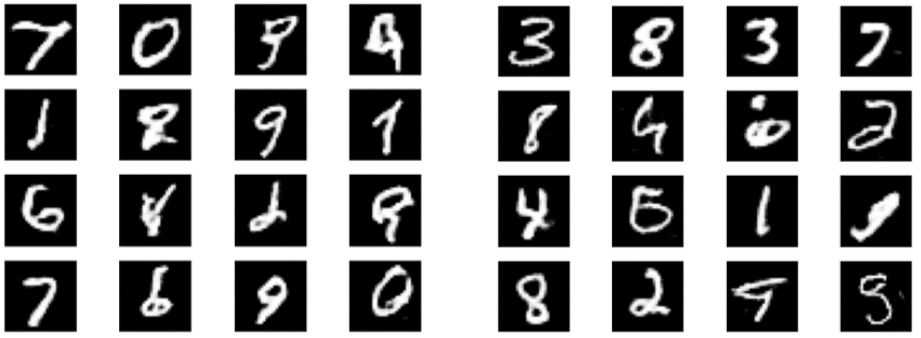
Model	Assessment Criteria	
	FID	IS
DDMP(L_{noise})	32.17	166.60
CGAN	35.20	148.68
DCGAN	36.19	132.98
VAE	39.98	121.67
Ours(L_{noise}+L_{sim})	**31.61**	**175.21**

DOI: 10.54254/2755-2721/32/20230199

In order to more intuitively feel the effect of image generation by the model, Figure 3 shows the generation effect of each comparison model. In the figure, each model shows 4*4 generated images, which are randomly sampled from the generated images. (a) is a similarity-based enhanced diffusion model, (b) is a diffusion model, (c) is a variational autoencoder (VAE), and (d) is a Deep Convolutional Generative Adversarial Networks (DCGAN). As can be seen from the figure, both the similarity-based enhanced diffusion model and the diffusion model can generate high-quality handwritten digital pictures, but the similarity-based enhanced diffusion model has better stability in image generation, and the image details are more clearly distinguishable.



(a) (b)



(c) (d)

Figure 3. Images generated by different comparison models.

5. Conclusion

In order to realize image generation, an enhanced diffusion model based on similarity is proposed in this paper. After reversing the diffusion process of the natural image, a new natural image can be gradually generated from a completely random noise image. Based on this, this paper improves the model by comparing the original image and the generated image to describe the semantic similarity between the two as part of the loss function, improves and optimizes the model, and uses the cosine similarity to measure the mass diffusion model of the generated image. Diffusion model has a strong development prospect. Diffusion model can be applied in various fields, such as image denoising, image restoration, super resolution imaging, image generation and so on. The simultaneous diffusion model is important for producing images with strong diversity and important details, and is a topic worthy of continue. In order to generate images, this paper proposes an enhanced Diffusion model based on similarity. After reversing the diffusion process of natural images, new natural images can be gradually generated from completely random noise images. Based on this, this paper improves the model by comparing the original image and the generated image, describes the semantic similarity between the two as part of the Loss function, improves and optimizes the model, and uses Cosine similarity to measure the quality of the generated image Diffusion model. The denoising diffusion Statistical model has achieved remarkable success in various image generation tasks, and can be applied to image denoising, image

DOI: 10.54254/2755-2721/32/20230199

restoration, super-resolution imaging, image generation and other fields. At the same time, Diffusion model is very important for generating images with strong diversity and important details, which is a subject worthy of further study.

Reference

[1] Aderhold J, Davydov V Yu, Fedler F, Klausing H, Mistele D, Rotter T, Semchinova O, Stemmer J and Graul J 2001 *J. Cryst. Growth* **222** 701

[2] Mingwen Shao, Wentao Zhang, Multi-scale generative adversarial inpainting network based on cross-layer attention transfer mechanism, *Knowledge-Based Systems*, 2020

[3] Jonathan Ho, Ajay Jain, Pieter Abbeel Denoising Diffusion Probabilistic Models arXiv:2006.11239v2 [cs. LG]

[4] Yanxi Wei, Yuru Kang, Fenggang Yao. Image Feature Understanding and Semantic Representation Based on Deep Learning, 2022 *International Conference on Artificial Intelligence of Things and Crowdsensing*, 2022

[5] Tiankai Hang, Shuyang Gu, Chen Li, Jianmin Bao, Efficient Diffusion Training via Min-SNR Weighting Strategy, *arXiv preprint arXiv:2303.09556*, 2023.

[6] Wijmans J G, Baker R W. The solution-diffusion model: a review. *Journal of membrane science*, 1995, 107(1-2): 1-21.

[7] Cao H, Tan C, Gao Z, et al. A survey on generative diffusion model. *arXiv preprint arXiv:2209.02646*, 2022.

[8] C. Li, Y. Qi, Q. Zeng and L. Lu, Comparison of Image Generation methods based on Diffusion Models, 2023 *4th International Conference on Computer Vision, Image and Deep Learning* 2023, 1-4.

[9] Tianrui Huang, Yang Gao, Zhenglin Li, Yue Hu, Fuzhen Xuan. A Hybrid Deep Learning Framework Based on Diffffusion Model and Deep Residual Neural Network for Defect Detection in Composite Plates, *Applied Sciences*, 2023

[10] Pai Zhang, Hanqing Chen, Qinrui Li. Research on Vehicle Recognition Algorithm based on Convolution Neural Network, *Journal of Physics: Conference Series*, 2021

[11] Weilun Wang, Jianmin Bao, Wengang Zhou, Semantic Image Synthesis via Diffusion Models 2022

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230200

A study of human pose estimation in low-light environments using YOLOv8 model

Kaiming Gu[1,3,†], Boyu Su[2,†]

[1]International Engineering College, Xi'an University of Technology, Xi'an, 710054, China
[2]School of Intelligent Engineering, Zhengzhou University of Aeronautics, Zhengzhou, 450046, China

[3]3222241013@stu.xaut.edu.cn
[†]All the authors contributed equally and their names were listed in alphabetical order.

Abstract. Human pose estimation is a formidable task in the field of computer vision., often constrained by limited training samples and various complexities encountered during target detection, including complex backgrounds, object occlusion, crowded scenes, and varying perspectives. The primary objective of this research paper is to explore the performance disparities of the recently introduced YOLOv8 model in the context of human pose estimation. We conduct a comprehensive evaluation of six different models with varying complexities on the same low-light photograph to assess their precision and speed. The objective is to determine the suitability of each model for specific environmental contexts. The experimental results reveal that our findings demonstrate a partial regression in accuracy for the yolov8s-pose and yolov8m-pose models when tested on our sampled images. The increase in model layers indicates enhanced complexity and expressive power, while additional parameters signify improved learning capabilities at the expense of increased computational resource requirements.

Keywords: human detection, pose estimation, YOLOv8, low-light environments.

1. Introduction

The advancements in human pose estimation and object detection have resulted in substantial progress within the domain of computer vision. Human pose estimation plays a pivotal role in the detection and localization of human keypoints in images, and it holds immense significance for the advancement of technologies like behavior recognition and pedestrian re-identification. In the field of human pose estimation, deep learning methods founded on convolutional neural networks have exhibited remarkable progress, achieving high accuracy in detecting and localizing human keypoints in both images and videos. As an illustrative example, the DeepPose model approaches the 2D human pose estimation task as a regression problem for keypoint coordinates [1]. Leveraging convolutional neural networks, it extracts pose features from images, thereby attaining elevated and more precise features to predict human keypoint coordinates [2]. This methodology has demonstrated improved performance in the accuracy of human pose estimation. Additionally, models like CPM and Hourglass have also achieved success in 2D human pose estimation, with wide applications not only in behavior recognition and pedestrian re-identification but also in pose analysis and motion tracking [3-4]. These models, when

DOI: 10.54254/2755-2721/32/20230200

applied to the field of object detection, help to make human life more convenient by detecting human pose targets to achieve research objectives.

In the domain of object detection, conventional methods often depend on sequential steps, such as region extraction and feature matching, which can hinder the achievement of efficient real-time detection. Indeed, the emergence of YOLO (You Only Look Once) has effectively addressed this issue. By adopting a unified approach that processes the entire image at once, YOLO achieves real-time object detection without the need for complex intermediate steps like region extraction and feature matching [5]. This streamlined methodology has significantly improved the efficiency and accuracy of object detection tasks. The YOLO model revolutionizes object detection by transforming the problem into an end-to-end regression task. It accomplishes this by predicting both the object's class and bounding box coordinates through a single network. This unique approach enables high-speed and real-time object detection, making it a significant breakthrough in the field of computer vision. The introduction of the latest YOLO model, YOLOv8, brings about significant advancements with three distinct components. Firstly, it incorporates a new and improved backbone network, enhancing the model's overall performance. Secondly, the Anchor-Free detection head offers an innovative approach to object detection, contributing to improved accuracy [6]. Lastly, the utilization of new loss functions further refines the model's capabilities in detecting target objects, making it adaptable to various application scenarios. YOLOv8 effectively achieves enhanced detection performance and accuracy, catering to the diverse needs of different real-world applications.

Both human pose estimation and object detection have seen significant advancements with the utilization of deep learning techniques and YOLO series models, which have not only improved the accuracy of detection and estimation but also accelerated the development of related applications. Among these models, the YOLOv8-based human pose detection methods have attracted considerable attention due to their impressive efficiency and accuracy. However, in practical scenarios, the YOLOv8 model exhibits a large number of parameters and hyperparameters, and selecting different combinations can lead to variations in model performance. Therefore, it becomes imperative to thoroughly study and compare the performance of six models based on the YOLOv8 architecture for human pose detection.

2. Method

This section introduces the basic framework of YOLOv8 and discusses six pose models based on YOLOv8.

2.1. Introduction of YOLOv8

The YOLOv8 is an object detection algorithm that uses deep learning and Anchor-Free detection head to achieve fast detection and classification of objects in images [6]. YOLOv8 inherits the high-precision detection methods of the YOLO series models. Indeed, YOLOv8 models the object detection task as a regression problem, where the neural network directly predicts the bounding box coordinates and class of the detected objects. This streamlined approach simplifies the detection process and contributes to the model's high-speed and real-time capabilities, making it an efficient solution for object detection in various applications. Compared to previous models, YOLOv8 potentially achieves higher detection speed and more accurate results. The Network Structure of YOLOv8 is shown in figure 1.

DOI: 10.54254/2755-2721/32/20230200

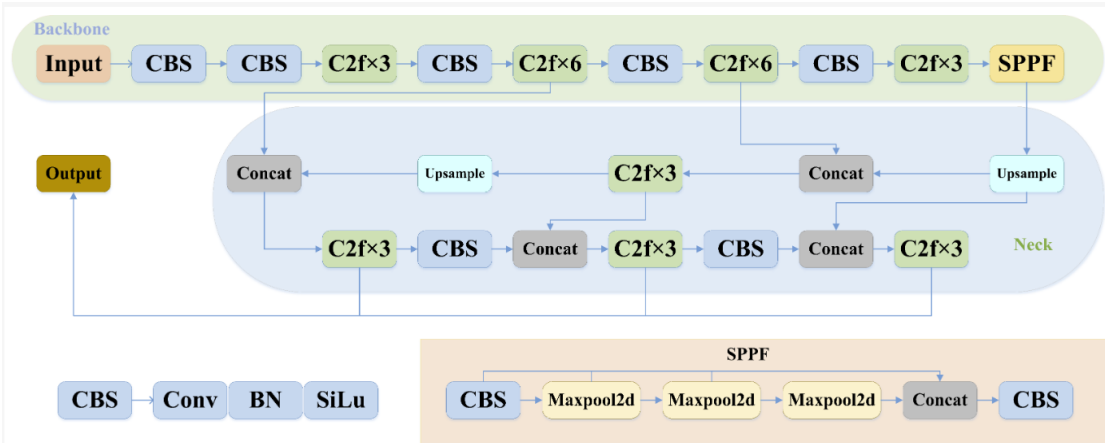


Figure 1. The network structure of YOLOv8 [7].

YOLOv8 uses the CSPDarkNet53 as its backbone network. The key highlight of the YOLOv8's backbone network is the implementation of a novel connection method known as Cross Stage Partial [8-9]. This technique optimally utilizes computational resources, ensuring both high accuracy and improved speed of the model. By intelligently leveraging these resources, YOLOv8 achieves a balance between accuracy and efficiency, making it a powerful tool for object detection tasks in real-world scenarios.

YOLOv8 also employs an Anchor-Free detection head, which directly predicts the object's bounding box from the feature map and adaptively adjusts the size and position of the bounding box [6]. This approach reduces the complexity of model training to some extent. In contrast to traditional detection heads, which adopt the Anchor-Based method and require predefined anchors of different sizes and positions for object localization and detection, the Anchor-Free detection head achieves better results, as adjusting anchor size and position in the Anchor-Based method can be complex [10].

In addition, YOLOv8 uses the CIOU and DFL loss functions [11]. The DFL loss function is defined as

$$DFL(S_i, S_{i+1}) = -\left((y_{i+1} - y)\log(s_i) + (y - y_i)\log(s_i + 1)\right) \quad (1)$$

In the DFL , s_i is the sigmoid function's output for the network. y_i and y_{i+1} are interval orders, and y denotes a label. This loss function handles overlapping between predicted bounding boxes and ground truth bounding boxes, improving the detection accuracy and robustness of small objects. The CIOU (Complete Intersection over Union) and DFL (Dynamic Focal Loss) loss functions in YOLOv8 are designed to measure the distance between predicted and ground truth bounding boxes, leveraging the IOU (Intersection over Union) value as a key metric. By incorporating a combination of cross-entropy loss and mean square error loss, these loss functions effectively optimize the model during training. This approach allows YOLOv8 to achieve higher precision and accuracy in detecting objects, further improving its performance in various object detection tasks.

2.2. Pose models based YOLOv8

In the YOLOv8 series, there are six pose estimation models (YOLOv8s-pose, YOLOv8n-pose, YOLOv8m-pose, YOLOv8l-pose, YOLOv8x-pose, YOLOv8x-pose-p6) that add the functionality of human pose estimation to the YOLOv8 model. These six models have different backbone network scales that balance accuracy and computational resources. Smaller models are suitable for resource-constrained environments, with lower computational and memory overhead, but may have limitations in pose estimation accuracy. Larger models offer higher pose estimation accuracy but require more computational resources and memory.

DOI: 10.54254/2755-2721/32/20230200

The pose estimation models in the YOLOv8 series combine the functionality of object detection and human pose estimation, exhibiting efficiency and accuracy. Absolutely, the selection of the appropriate model depends on the specific requirements of the task at hand. It involves striking a balance between computational resources and accuracy to achieve fast and precise object detection and pose estimation tasks. For scenarios where real-time performance is crucial, models like YOLOv8 that optimize speed while maintaining acceptable accuracy would be preferred. On the other hand, for applications where precision is of utmost importance and computational resources are less constrained, more sophisticated and accurate models might be chosen. The ultimate goal is to choose a model that best fits the practical needs and constraints of the given task.

The specific characteristics of each pose estimation model in the YOLOv8 series are as follows:

- YOLOv8s-pose

- Description of the backbone network: CSPDarkNet53.

- Characteristics of YOLOv8s-pose: small size, suitable for resource-constrained environments.

- Performance analysis: low computational and memory overhead, potential limitations in pose estimation accuracy.

- YOLOv8n-pose

- Description of the backbone network: CSPDarkNet53.

- Characteristics of YOLOv8n-pose: medium size, balance between accuracy and computational resources.

- Performance analysis: improved pose estimation accuracy compared to YOLOv8s-pose, increased computational and memory overhead.

- YOLOv8m-pose

- Description of the backbone network: CSPDarkNet53.

- Characteristics of YOLOv8m-pose: large size, better balance between accuracy and computational resources.

- Performance analysis: higher pose estimation accuracy compared to YOLOv8n-pose, increased computational resources and memory requirements.

- YOLOv8l-pose

- Description of the backbone network: CSPDarkNet53.

- Characteristics of YOLOv8l-pose: larger size, higher pose estimation accuracy.

- Performance analysis: improved accuracy and precision compared to YOLOv8m-pose, higher computational resources and memory requirements.

- YOLOv8x-pose

- Description of the backbone network: CSPDarkNet53.

- Characteristics of YOLOv8x-pose: largest size among the YOLOv8 series, highest pose estimation accuracy.

- Performance analysis: maximum computational resources and memory requirements.

- YOLOv8x-pose-p6

- Description of the backbone network: CSPDarkNet53.

- Characteristics of YOLOv8x-pose-p6: larger input resolution.

- Performance analysis: bette

3. Experimental analysis

3.1. Experimental details

Dataset: The coco128 dataset, widely used in computer vision, is utilized by the authors for training and evaluation purposes in this study. It includes images from 128 different categories, and each image may contain multiple object instances, thus possessing multi-label attributes.

Setup: The Google Colab experimental platform on GitHub is a cloud-based interactive computing environment that provides a free integrated Jupyter notebook for coding and running code directly in the browser. It supports GPU and TPU acceleration, seamlessly integrates with Google services, and has

DOI: 10.54254/2755-2721/32/20230200

seamless connectivity with GitHub, facilitating user sharing and collaboration. The platform comes pre-installed with popular data science and machine learning libraries, providing convenient access to common datasets. By leveraging Google's computational resources, users can fully exploit the potential of cloud computing for conducting experiments and projects in the fields of data science, machine learning, and deep learning.

3.2. Experimental results analysis

From YOLOv8n-pose to yolov8x-pose-p6, there is an increase in the complexity of the model, leading to longer testing time requirements. The following figure 2 presents the test results of six different YOLOv8 models with varying complexities on the same photo.

Figure 2. Experimental results in a low-light environment.

DOI: 10.54254/2755-2721/32/20230200

Overall, as the model complexity increases, the overall detection accuracy also improves. Nonetheless, when tested on images, the YOLOv8s-pose and YOLOv8m-pose models demonstrate a partial regression in accuracy. In particular, the accuracy of detecting the main subject on the left side decreased from 0.89 to 0.87, and the accuracy of detecting the person wearing a yellow shirt on the right side also decreased from 0.91 to 0.87. Additionally, the detection accuracy of the objects in the background, excluding the one in the middle, also experienced a decline. It is speculated that for the test targets, the results obtained by YOLOv8s-pose and YOLOv8m-pose are almost identical, with YOLOv8m-pose having slightly more layers and parameters than YOLOv8s-pose. Thus, within this range, increasing the model complexity does not necessarily lead to more accurate results in testing. On the contrary, it may result in longer model runtime, leading to repeated detections of a single target and consequently causing misidentification and decreased precision. Similarly, looking at the detection results of YOLOv8l-pose, a higher number of layers and parameters should ideally yield more accurate results, but it only detected eleven objects and had lower detection accuracy in the background.

Comparison of Metrics: In the provided results (Table 1), the preprocessing time refers to the duration taken to prepare and preprocess the images before feeding them into the model. The inference time represents the period consumed by the model to perform object detection predictions on the preprocessed images. Lastly, the post-processing time signifies the time spent on processing and organizing the model's outputs after the inference step, to present the final detected objects and their respective bounding boxes.

Table 1. Index comparison of six models.

Model Type	Number of Layers	Parameters (M)	Time Cost(ms)	Detected Persons	Preprocessing Procedure(ms)	Inference Process(ms)
YOLOv8n-pose	187	3.3	135.6	9	3.5	135.6
YOLOv8s-pose	187	11.6	375.4	13	2.4	375.4
YOLOv8m-pose	237	26.4	880.2	13	2.5	880.2
YOLOv8l-pose	287	44.5	1714.3	11	2.5	1714.3
YOLOv8x-pose	287	69.5	2555.4	14	2.6	2555.4
YOLOv8x-pose-p6	375	99.1	10373.2	16	12.6	10373.2

Analysis of Metrics: From the perspective of preprocessing and inference time, YOLOv8x-pose-p6 has the longest runtime. This indicates that it takes more time to handle issues such as scaling, cropping, and normalization due to the model's extensive layers and complex connections. Each input needs to pass through these layers, resulting in increased computational workload and correspondingly increased runtime. When performing higher-resolution scaling on images, employing more complex cropping strategies, or executing additional normalization steps, these additional preprocessing steps increase the model's processing time. Additionally, this model has the largest number of parameters, requiring more parameter calculations and storage during inference and training, leading to increased runtime. A larger number of parameters also increases the memory requirement, thereby slowing down the model's execution speed.

Increasing the number of layers in a model may imply greater complexity and expressive power, while parameters indicate more learning capacity but can also result in increased computational resource requirements. The above table displays the time required by each model for processing the photo. Among the models, YOLOv8n-pose is the fastest, with the highest inference speed and the smallest number of layers, but it does not yield the most accurate results. On the other hand, YOLOv8x-pose-p6

DOI: 10.54254/2755-2721/32/20230200

has the longest runtime, the most layers, and the slowest inference speed, but it delivers the most accurate results. The preprocessing, inference, and post-processing times represent the computational overhead at different stages of the model. The table clearly demonstrates that as the model type progresses to more advanced versions, both the model complexity and the number of parameters increase significantly. If the fastest inference speed is desired, YOLOv8n-pose can be selected. If higher detection accuracy is required, YOLOv8x-pose-p6 can be chosen.

4. Conclusion

To evaluate the results of different performance models on the same image, this paper analyzes six different pose models based on YOLOv8, discusses their principles, and focuses on analyzing the meanings of various parameter values after testing the six models on the same image. The analysis results indicate that increasing model complexity does not necessarily improve test accuracy and may instead lead to longer model runtimes, resulting in misidentification and decreased precision. Increasing the number of layers in a model implies increased complexity and expressive power, while increasing the number of parameters indicates increased learning capacity but can also result in higher computational resource requirements. Furthermore, the preprocessing, inference, and post-processing stages are associated with computational overhead, and different models handle these costs differently. The YOLOv8n-pose model is the fastest, with the highest inference speed, but it does not yield the most accurate results. On the other hand, the YOLOv8x-pose-p6 has the longest runtime, the most layers, and the slowest inference speed, but it achieves the highest accuracy. Indeed, from the aforementioned results, it is evident that both model complexity and the number of parameters increase as the model type advances.

In summary, as model complexity increases, detection accuracy generally improves, but within a certain range, further increasing model complexity may lead to longer runtimes, misidentification, and decreased precision. Selecting a suitable model requires considering inference speed and detection accuracy and striking a balance based on specific requirements. The preprocessing, inference, and post-processing stages also contribute to computational overhead, and for different application scenarios, it is possible to choose an appropriate model that achieves the optimal balance according to the specific needs.

References

[1] Toshev, A., & Szegedy, C. DeepPose: Human Pose Estimation via Deep Neural Networks. *In IEEE Conf. Comp. Vis. Patt. Recogn.* 2014, 1653-1660.

[2] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. Gradient-Based Learning Applied to Document Recognition, Proceedings of the IEEE, 1998, 86(11), 2278-2324.

[3] Zhang, H., Cui, Y., Zhang, L., Wu, S., & Zhang, H. CPM: A Large-scale Generative Chinese Pre-trained Language Model. arXiv preprint arXiv:2012.00413,2020.

[4] Newell, A., Yang, K., & Deng, J. Stacked Hourglass Networks for Human Pose Estimation. *Euro. Conf. Comp. Vis.*, 2020, 483-499.

[5] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *In IEEE Conf. Comp. Vis. Patt. Recogn.*, 2016, 779-788.

[6] Zhou, X., Wang, D., & Krähenbühl, P. Objects as Points. arXiv preprint arXiv:1904.07850, 2019.

[7] Haitong Lou, Xuehu Duan, Junmei Guo, DC-YOLOv8: Small-Size Object Detection Algorithm Based on Camera Sensor. *Electronics* 2023, 12(10), 2323.

[8] Y. Chen et al., "CSPDarkNet53: A Light-weight Convolutional Neural Network for Object Detection with a Comprehensive Evaluation," *IEEE Trans. Cir. Sys. Video Tech.*, 2020, 30 1, 1-14.

[9] He, K., Zhang, X., Ren, S., & Sun, J.. Deep Residual Learning for Image Recognition. In *In IEEE Conf. Comp. Vis. Patt. Recogn.*, 2021, 37-51.

[10] Ren, S., He, K., Girshick, R., & Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *In Adv. Neur. Inform. Proc. Sys.* 2015, 91-99.

[11] Zhu, X., He, C., & Zhang, J. Distribution Focal Loss for Dense Object Detection. *In IEEE Conf. Comp. Vis. Patt. Recogn.* 2019, 12814-12823.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230201

VGG16 based on dilated convolution for face recognition

Fanxing Meng

Electronic and Information Engineering, Beijing University Of Technology, Beijing, 100000, China

Mengfanxing@emails.bjut.edu.cn

Abstract. Face recognition has wide applications in fields such as security systems, biometrics, and human-computer interaction. However, traditional face recognition methods face challenges in capturing details and reducing model complexity. To address these issues, this paper proposes a new method based on VGG16, which improves recognition accuracy and reduces parameter quantity by introducing dilated convolution and parameter pruning. First, the hole convolution is introduced to expand the Receptive field and capture more details to enhance the ability of the model in distinguishing facial features. Next, parameter pruning is applied to reduce redundant parameters, optimize model structure, and improve computational efficiency. This article conducted experimental evaluation on the classic face recognition dataset CK+ dataset. The results show that the proposed method is significantly superior to the traditional VGG16 model in terms of recognition accuracy. At the same time, the use of pruning technology significantly reduces the number of parameters in the model and improves computational efficiency. The experimental outcomes conclusively validate the effectiveness and feasibility of the proposed method.

Keywords: VGG16, dilated convolution, face recognition.

1. Introduction

Facial expressions are one of the important nonverbal communication methods between people and the most direct manifestation of emotional transmission. By recognizing and understanding facial expressions, we can obtain information about a person's emotional state, emotions, and intentions. In the contemporary era, with the improvement of computer hardware, facial expression recognition technology has found application in numerous domains, including human-computer interaction, affective computing, intelligent security, and various other fields [1].

In traditional machine learning, features are extracted manually. Once the data volume is too large, feature extraction will be a very complex process. In deep learning, neural networks are mainly used for feature extraction, avoiding the tedious process of manual extraction, and the feature extraction effect is better [2]. This implies that it offers an effective and automated approach for analyzing facial expressions, thereby enhancing the performance and expanding the potential applications of facial expression recognition systems. This has great potential in fields such as social media analysis, personalized recommendations, facial animation, and emotional assistance therapy.

At present, deep learning based facial expression recognition has made significant progress. Researchers have proposed a series of deep neural network models, for example, prominent techniques

DOI: 10.54254/2755-2721/32/20230201

employed for extracting and learning expression features from face images or videos include Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and attention mechanisms. These advanced methodologies have demonstrated their effectiveness in capturing meaningful spatial information and modeling temporal dependencies, thereby facilitating accurate and robust facial expression recognition. At the same time, there are also a large number of publicly available datasets, such as FER2013, CK+, and FERG, which provide annotated data for facial expression recognition, promoting the development and evaluation of research.

In this paper, we use deep learning to recognize facial expressions in video. We mainly use Convolutional neural network to extract facial expression features from a single image. We use hole convolution to solve the problem of internal data structure loss and spatial hierarchical information loss caused by standard convolution, and on this basis, we conduct appropriate channel pruning, which has played a role in promoting facial expression feature extraction.

2. Method

2.1. Introduction to VGG16

To enhance the precision of facial expression recognition, this study employs VGG16 as the foundational neural network model [3]. VGG16, a deep Convolutional Neural Network (CNN), has been specifically chosen for its compatibility with facial expression recognition tasks. Its Deep structure and surface structure can extract rich and abstract features, and is sensitive to subtle expression differences. The VGG16 model has a smaller convolutional kernel size, which helps to extract more detailed information from images. VGG16 combines convolution and fully connected layers, possessing excellent feature representation and classification capabilities, comprehensively perceiving images and learning expression differences.

VGG16 was proposed by a research team at the University of Oxford in 2014. Its model has a 16-layer deep network structure, which includes 13 convolutional layers and 3 fully connected layers. Within the convolutional layer segment, there exist a total of 13 convolutional layers. The initial convolutional layers employ smaller 3×3 convolutional kernels, while the subsequent convolutional layers utilize larger kernels such as 3×3, 4×4, and 5×5. Following each convolutional layer, a ReLU activation function is applied to incorporate nonlinearity. To down sample the feature map and decrease the size and parameter count, a pooling layer is inserted between two adjacent convolutional layers. The most commonly employed pooling methods include 2×2 maximum pooling and mean pooling (Figure 1).

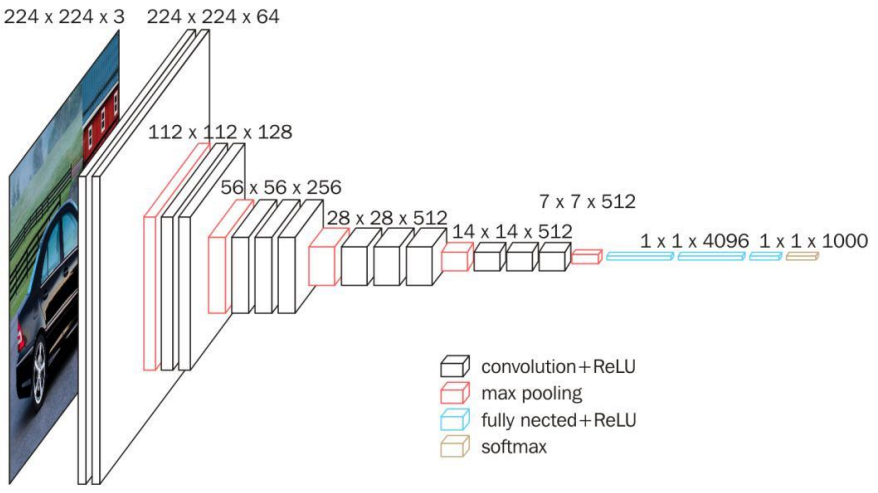


Figure 1. Structure diagram of vgg16 [3].

DOI: 10.54254/2755-2721/32/20230201

In the full connection layer, there are three hidden layers with 4096 neurons, and each hidden layer is followed by ReLU Activation function. The final output layer is a softmax layer with a number of output categories for multi category classification [4].

The VGG16 model extracts high-level features of images by stacking multiple convolutional layers and fully connected layers, thereby achieving the tasks of image classification and target recognition. The traditional VGG16 model has some shortcomings in facial expression recognition. Firstly, the VGG16 model is constructed based on traditional convolutional and pooling layers, which use relatively large convolutional kernels and pooling windows when processing inputs, resulting in the loss of some detailed information [5]. This may affect the model's ability to perceive subtle changes in facial features.

In addition, the convolution operation of the VGG16 model is continuous and does not consider the spatial relationships and contextual information between different regions. This is not ideal in facial expression recognition, as expressions often involve specific regions and local features, requiring the model to accurately capture this information. For these problems, introducing Dilated Convolution can solve them to some extent.

2.2. Implementation of dilated convolution

Hole convolution, also known as dilation convolution or hole convolution, is a special convolution operation used in Convolutional neural network. Compared to traditional convolution operations, hollow convolution has an additional hyperparameter called inflation rate, which defines the spacing between adjacent convolutional kernels. Traditional convolution operations scan the input feature map with a fixed filter size, while dilated convolution introduces gaps within the filter based on the expansion rate, resulting in voids in the convolution operation [6].

In this article, adding empty convolutions can help improve the accuracy of facial expression recognition. First, it expands the Receptive field by increasing the inflation rate, so that it can capture broader context information, which is particularly important for face recognition. In addition, cavity convolution can also by incorporating pooling layers between the convolutional layers, the number of parameters is reduced, thereby decreasing the model's overall complexity and computational requirements. This reduction is achieved while ensuring that the receptive field remains unchanged. Consequently, the model's efficiency is enhanced as it continues to capture the necessary information effectively [7]. Most importantly, hollow convolution avoids the use of pooling operations for down sampling, thereby preserving more detailed information and meeting the requirements of face recognition for high-resolution, detail rich images.

In the given scenario, the convolutional kernel size is 3 × 3. The figure showcases the empty convolution with an expansion rate of d. Figure 2 highlights that ordinary convolution is a particular case of empty convolution, precisely an empty convolution with an expansion rate of 1. Taking Figure 3 as an example, with an expansion rate of 2, the original 3 ×3 The convolution kernel of 3 increases the Receptive field to 5 × 5.

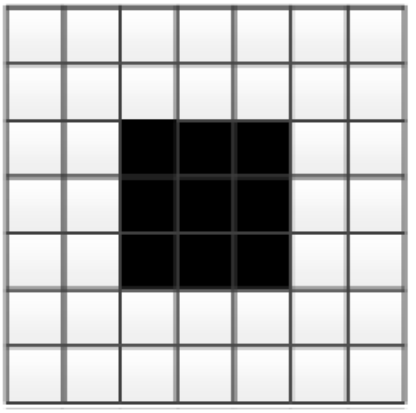


Figure 2. The ordinary convolution.

DOI: 10.54254/2755-2721/32/20230201

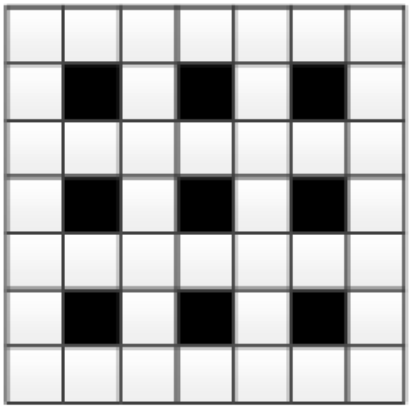


Figure 3. An example with an expansion rate of 2.

This article explores different scenarios for the application location and expansion rate of dilated convolution. After extensive experimentation, it was determined that the most optimal recognition rate was achieved by incorporating dilated convolution with an expansion rate of 2 in the final layer of the VGG-16 network.

2.3. Pruning

Neural network pruning is a technique that optimizes the structure of a neural network model by reducing redundant parameters. Pruning can reduce the size and computational complexity of the model by removing unimportant connections, neurons, or layers in the network, while maintaining model performance.

The VGG16 model consists of 13 convolutional layers and 3 fully connected layers, resulting in a total of approximately 140.6 million parameters. In this article, channel pruning is used to remove redundant and unimportant channels in the model to reduce the number of parameters. Firstly, we choose L1 regularization as the pruning algorithm [8]. L1 regularization punishes the channel weights of the convolutional layer and obtains the corresponding channel importance score. A smaller channel weight means that the contribution of the channel to the model is smaller, so it can be considered to delete these channels. Subsequently, we sort the channels based on the L1 norm score. After sorting, delete the channels with lower scores to achieve the desired pruning ratio. In this example, we choose a 30% pruning ratio [9]. After deleting the channel, fine tune or retrain the pruned model to recover the performance degradation caused by pruning.

Through pruning operations, this article achieves model compression and improves computational efficiency. Reduce the storage requirements of the model. This makes the deployment of the model on devices with limited memory more feasible and enables faster loading and execution.

3. Experiments:

3.1. Dataset introduction + Experimental details

The data set in the real scene generally has Confounding such as lighting, occlusion, low pixels, etc. It is relatively difficult to extract features. The CK+ data set used in this paper is the data set in this real scene. The CK+ dataset is a dynamic video dataset [10-11]. At the same time, it is a competition level dataset. Compared with the dataset recorded in the laboratory, it has added some Confounding, among which the Confounding mainly include occlusion, too low pixels, background changes, etc. Because of the existence of these Confounding, it is more realistic.

The CK+ (Cohn Kanade) dataset is a commonly used facial expression recognition dataset for training and evaluating facial expression recognition algorithms. This dataset was jointly created by researchers from Stanford University and the University of California, San Diego. The CK+ dataset contains facial expression sequences performed by a group of volunteers in a laboratory environment.

DOI: 10.54254/2755-2721/32/20230201

Each volunteer will exhibit a series of facial expressions, including seven basic expressions of happiness, sadness, anger, surprise, disgust, and fear, as well as a neutral expression. Each facial expression sequence consists of a series of consecutive facial images.

The CK+ dataset comprises 327 video sequences obtained from 123 subjects. The length of each sequence ranges from 10 to 60 frames. The dataset provides grayscale images, and each image has a resolution of either 640×480 or 640×490 pixels. It is a dynamic video dataset that forms a video frame dataset by performing frame truncation operations on the video, and each video frame has been pre labelled with emoticons.

3.2. Experimental details

In this article, it is required that the CNN network output multiple continuous facial features simultaneously. Therefore, the model inputs n facial images each time, and each facial image can share the CNN network weight for feature extraction. After setting the final n value to 8, 8 consecutive facial images are passed in at once. VGG16 adopts the VGG-16-FACE model with pre training weight, and the initial Learning rate is 0.01. The optimization algorithm is the random Gradient descent, the momentum parameter value is set to 0.9, and the training times are 200. For input, the image set included in the dataset was used, with each image size of 224 × 224. After the network training is completed, the test set is fed into the model and the test results are obtained. The experimental code was completed using PyCharm on the Ubuntu 20.04 system, with a virtual machine memory of 4GB and a host CPU model of R9-7945HX.

The F1 score, also referred to as the F-Measure, is a common metric employed in statistics and machine learning to assess the classification model's accuracy. It is calculated as the harmonic mean of Precision and Recall. By incorporating both Precision and Recall, the F1 score provides a comprehensive evaluation of the model's accuracy and ability to recall positive instances. This score is valuable in measuring the model's performance stability and generalization capabilities when making predictions.

Formula for calculating F1 score:

$$F1 = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{1}$$

$$\text{precision} = \frac{\text{TP}}{\text{TP+EP}} \tag{2}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{3}$$

Among these, True Positive (TP) denotes the number of accurate predictions made by the model. False Positive (FP) refers to the number of instances that were incorrectly predicted as belonging to a particular class. On the other hand, False Negative (FN) represents the number of instances that should have been predicted correctly but were mistakenly predicted as negative.

The calculation formula for accuracy:

$$\text{accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{4}$$

Among them, TN (True Negative) is the number that predicts other classes as correct.

3.3. Feature visualization

3.3.1. Confusion matrix. The Confusion matrix presents the relationship between the prediction outcomes of a classification model on the test dataset and the actual labels, represented in matrix form. It provides a visual representation of how well the model performs in classifying the data. Through the Confusion matrix, we can clearly understand the prediction of the classification model for each category, which is helpful to identify which categories the classification model predicts more accurately (Figure 4).

DOI: 10.54254/2755-2721/32/20230201

	Anger	Contempt	Fear	Happiness	Neutral	Sadness	Surprise
Anger	18	0	1	0	0	0	0
Contempt	0	10	0	0	1	0	0
Fear	1	0	12	2	0	0	0
Happiness	0	0	0	27	0	0	0
Neutral	0	1	2	0	12	2	1
Sadness	1	0	0	1	0	23	1
Surprise	0	0	0	0	0	0	23

Figure 4. CK+ dataset traditional VGG16 test results.

	Anger	Contempt	Fear	Happiness	Neutral	Sadness	Surprise
Anger	18	0	1	0	0	0	0
Contempt	0	10	0	0	0	0	0
Fear	1	0	12	1	0	0	0
Happiness	0	0	0	28	0	0	0
Neutral	0	1	2	0	13	1	0
Sadness	1	0	0	1	0	24	1
Surprise	0	0	0	0	0	0	24

Figure 5. CK+ dataset traditional VGG16+dilated Convolutional test results.

After adding hollow convolution, some test videos that were originally incorrectly recognized are now recognized correctly, and the original correct recognition remains correct. It can be seen that hollow convolution has an improvement in the accuracy of the original model's recognition (Figure 5).

	Anger	Contempt	Fear	Happiness	Neutral	Sadness	Surprise
Anger	18	0	1	0	0	0	0
Contempt	0	10	0	0	1	0	0
Fear	1	0	12	1	0	0	0
Happiness	0	0	0	28	0	0	0
Neutral	0	1	2	0	12	2	0
Sadness	1	0	0	1	0	23	1
Surprise	0	0	0	0	0	0	24

Figure 6. CK+ dataset traditional VGG16+dilated convolutional + pruning test results.

After channel pruning, as the number of parameters decreases, the recognition accuracy also decreases slightly, but the overall recognition rate remains at a high level (Figure 6).

3.4. Performance comparison

The table presents the test outcomes obtained from the CK+ dataset. When applied to video data, the standard VGG16 model attains an accuracy of 90.58%, precision of 83.33%, recall rate of 88.89%, and an F1 score of 0.86 in facial expression recognition. These findings suggest that the traditional VGG16 model has accomplished favorable results in recognizing facial expressions (Table1).

DOI: 10.54254/2755-2721/32/20230201

Table 1. The test outcomes obtained from the CK+ dataset.

	Accuracy	Precision	Recall	F1
VGG16	90.58%	83.33%	88.89%	0.86
VGG16+Dilated Convolutional	95.65%	93.33%	93.33%	0.93
VGG16+Dilated Convolutional+ Pruning	94.20%	89.36%	93.33%	0.91

To enhance the accuracy and performance of facial expression recognition even further, this research introduces dilated convolution and parameter pruning techniques to the VGG16 model. Through experimental testing, notable advancements in performance were observed. Upon the incorporation of dilated convolution and subsequent testing, the revised VGG16 model demonstrated an enhanced accuracy of 95.65% in facial expression recognition on the CK+ dataset. The model achieved an accuracy rate of 93.33%, a recall rate of 93.33%, and an F1 score of 0.93. These results exhibit the positive impact of integrating dilated convolution on improving the model's ability to recognize expressions. This is due to the fact that the hole convolution expands the Receptive field of the network, enabling it to better capture the details of facial expressions.

Subsequently, by applying parameter pruning techniques, this article successfully reduced the parameter count of the VGG16 model. As a result, the model's operations have become more efficient, leading to favourable performance on the CK+ dataset. Following the application of parameter pruning, the VGG16 model achieved an accuracy of 94.20% in facial expression recognition. The accuracy rate stood at 89.36%, the recall rate at 93.33%, and the F1 score at 0.91. Although the accuracy and F1 score have decreased due to the reduction of some parameters by pruning, the magnitude of the decrease is within an acceptable range, and the accuracy is still higher than traditional VGG16.

Table 2. Parameter quantity of VGG16.

Name	Parameter quantity before pruning	Parameter quantity after pruning
Convolutional Layer1	1,792	1,792
Convolutional Layer2	73,728	51,710
Convolutional Layer3	1,048,576	734,003
Convolutional Layer4	1,179,648	825,754
Convolutional Layer5	2,359,296	1,652,349
Convolutional Layer6	2,359,296	1,652,349
Convolutional Layer7	2,359,296	1,652,349
Convolutional Layer8	4,718,592	3,892,008
Convolutional Layer9	4,718,592	3,892,008
Convolutional Layer10	4,718,592	3,892,008
Convolutional Layer11	4,718,592	3,892,008
Convolutional Layer12	4,718,592	3,892,008
Convolutional Layer13	4,718,592	3,892,008
Fully connected layer1	102,760,448	62,949,474
Fully connected layer2	16,777,216	16,777,216
Fully connected layer3	4,194,304	4,194,304
Output layer	4,096	4,096
total	140,630,976	111,922,730

Through pruning, we observed a significant reduction in the parameter count of the VGG16 model. After 30% channel pruning, the parameter quantity of the model is approximately 111.9M, which is nearly 20% less than the original model's 140.6M. This accomplishment is obtained by applying pruning algorithms to remove a specific portion of channels, effectively reducing the number of parameters within the Convolutional Layer. Channel pruning technology plays a crucial role in decreasing model parameters, subsequently reducing the computational complexity of the model. Through the elimination

DOI: 10.54254/2755-2721/32/20230201

of certain channels, the computational complexity of multiplication and addition operations within the Convolutional Layer is reduced. This outcome translates to enhanced inference speed and higher computational efficiency (Table 2).

Simultaneously, reducing the parameter count has the added benefit of decreasing the storage space required by the model. Consequently, this simplifies the deployment and utilization of the model in resource-constrained environments.

4. Conclusion

This paper presents an optimized approach to enhance the accuracy of video facial expression recognition by building upon the traditional VGG-16 network framework. The optimization strategy entails two key techniques: hollow convolution (or dilated convolution) and parameter pruning. To begin with, the introduction of hollow convolution allows for the capturing of spatial information from facial features at various scales, without incurring additional computational costs. By adjusting the expansion rate of the convolution kernel, this method effectively expands the receptive field and enriches the model's ability to perceive intricate facial details. Furthermore, parameter pruning is employed to reduce the complexity of the model while maintaining its performance. Through precise analysis of network weights, unessential connections are identified and eliminated, resulting in a more compact and efficient model. By incorporating both hollow convolution and parameter pruning techniques, the proposed approach aims to enhance the accuracy of video facial expression recognition while optimizing computational costs and model efficiency. This article conducted extensive experiments on the CK+ dataset to evaluate the effectiveness of the proposed method. The results indicate that our method significantly improves accuracy compared to the benchmark VGG16 model without cavity convolution and pruning. The introduction of hollow convolution enables the model to capture fine-grained facial features, enhancing the ability to distinguish between individuals. In addition, parameter pruning effectively reduces the number of parameters in the network, improves computational efficiency without affecting performance.

This study contributes to the advancement of facial recognition technology by addressing the challenges of capturing fine-grained facial features and reducing model complexity. This method has broad application potential in fields such as biometric recognition systems, monitoring technology, and facial based authentication systems.

[1] Aderhold J, Davydov V Yu, Fedler F, Klausing H, Mistele D, Rotter T, Semchinova O, Stemmer J and Graul J 2001 J. Cryst. Growth 222 701

References

[1] Peng Z, Weiwei K, and Jinbao T. Face expression recognition based on multi-scale feature attention mechanism. *Comput. Eng. Appl.*, 2022,**58 (01)**: 182-189

[2] Xuan H. Research on face recognition methods in unrestricted scenes. *Sichuan Univ.*, 2021.

[3] Jinxiang L. Image recognition and target detection based on VGGNet. *Yanshan Univ.*, 2021.

[4] Jie Z. Research on the Application of Deep Learning Based Face Recognition Algorithms in Video Surveillance. *Electric. Des. Eng.*, 2023,**31 (13)**: 182-186.

[5] Jiehao W. Research on 3D facial expression recognition method based on deep learning. *Xi'an Univ. Tech.*, 2023.

[6] Qianqian L, Weixing W and Qin Y, etc. Research on audio-visual multimodal emotion recognition based on deep learning. *Comp. Dig. Eng.*, 2023,**51 (03)**: 695-699

[7] Shuyu D. Occlusive facial expression recognition based on deep learning. *Shandong Univ.*, 2022.

[8] Junling X. Research on Natural Scene Facial Expression Recognition Method Based on Deep Learning. *Chongqing Univ. Posts Telecomm.*, 2022.

[9] Huihua X, Ming L and Yan W, etc. Facial expression recognition based on DE Gabor features. *J. Nanchang Hangkong Univ. (Natural Science Edition)*, 2021,**35 (02)**: 82-91+124

[10] Ting Z, Research on facial expression recognition based on deep learning. *South China Univ. Techn.*, 2022.

DOI: 10.54254/2755-2721/32/20230201

[11] Dongdong Q, Lile H, Lin H. Improved Lightweight Face Recognition Algorithm. *J. Intel. Sys.*, 2023,**18 (03)**: 544-551

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230203

Review of artificial neural networks in first-person shooter games

Hongyu Chen

Computer Science and Technology, Northeastern University, Qinghuangdao, Hebei, China, 110819

Chenhongyu375@gmail.com

Abstract. More and more games have entered the market as computer processing power has improved, drawing in a sizable fan base. As a result, the video game industry has seen a rise in both revenue and the breadth of its product offering. Whether or not a game can generate enough revenue is dependent on a number of factors, including the game's ability to attract players, the quality of its gameplay, and the experience it provides to those that play it. This paper through methods of literature review and analysis will review the target detection, image recognition, and other artificial intelligence-related technology used in games, as well as provide suggestions for future development and a summary of the current state of the field.

Keywords: artificial neural network, first person shooter game, computer game.

1. Introduction

By 2022, the Chinese market is projected to account for approximately one-third of worldwide esports sales, according to Newzoo, a pioneer in global market research and predictive analytics for the gaming business. In terms of earnings, the Asian-Pacific region will represent about half of the world's gaming market [1]. During the anti-occupation years, the online gaming industry expanded into a whole supply chain to meet the demand from gamers of all ages. The exponential expansion of the gaming population can be directly attributed to the proliferation of online games during the past few years. The number of Chinese mobile online game users is expected to reach 600 million by 2020, and is expected to continue growing rapidly. FPS games account for as much as 47.13 percent of the whole gaming market, which has reached a record amount of 250 billion yuan [2]. Providing a satisfying gaming experience for customers is now a primary goal in order to maintain a growing player base.

Many first-person shooters have AI enemies for new players to face, as the 100 people who appear in Jedi Survival based on the player's level. However, these NPCs are not separated into tiers and cannot be on par with the player's level. As a result, the player does not have a sufficient feeling of accomplishment, and the implementation of AI is imperfect.

When it comes to player-versus-player combat or story, today's popular games are near-perfect. However, they don't provide players with artificial intelligences that can keep up with them as they level up. After logging in, most players will have a bad time due to hostile server environments and high-level gamers killing them. Applying AI to games has the potential to dramatically increase players'

DOI: 10.54254/2755-2721/32/20230203

interest, provide a sense of achievement in the battle with AI, and make players' operations more skilled, all of which contribute to player retention as AI technology advances.

In this paper, we will use the research methods of a literature review and literature analysis to summarize the target detection, image recognition, and other AI-related technologies used in the game, offer some recommendations, and discuss some possible directions for future study and development. There are a lot of games out there right now, but none of them use artificial intelligence-related techniques. As a result, the future of AI in games seems bright. This paper provides a wealth of algorithms and games paired with the idea, which can encompass the game involved in most of the principles of AI, target detection techniques, deep reinforcement learning, artificial neural networks, and so on. Therefore, this paper is useful as a point of reference for game developers and those working to advance artificial intelligence in the gaming industry.

2. Application of basic theory and related technology

2.1. Game development engine

A game development engine is a crucial part of any system for creating and editing computer games or any interactive real-image application. These frameworks equip game developers with the resources they require to create games rapidly and easily. Most gaming engines are cross-platform and can be used on a number of different computer systems. A few examples of game engines are the rendering engine, the physics engine, the collision detection system, the sound engine, the scripting engine, the computer animation engine, the artificial intelligence engine, the network management engine, and the scene management engine [3-4]. For the development of the engine competency determines whether the game is stable or not, the selection of a suitable development engine is one of the main aspects. The engine comprises a number of operations, such as creating new game projects, altering game animation game scenes, logic, etc. Games have entered the high-definition age with the arrival of Sony's PS3 and Microsoft's XBOX360, and the eight-core processor game consoles have vastly improved the quality and effects of games compared to older PC games. As a result, the "next generation" of gaming was initiated [5]. The Unreal Engine has become the most popular engine for making cutting-edge video games thanks to its innovative blueprint development method, comprehensive skeletal animation system, high-quality and precise lighting normal mapping, potent material mapping, rich and complete API calls, and support for multi-platform release [6].

2.2. Artificial neural network

It was around the year 1980 that the artificial neural network [4] was developed as a special mode of information communication that mimics the process of information transfer by modeling it after the neural networks in the human brain. The concept of an artificial neural network (ANN) has been gaining traction recently. An ANN is a mathematical model that mimics the way human neurons connect to one another in order to store and retrieve information (the "memory"). Fully connected networks, feed-forward networks, and convolutional networks are the three main types of artificial neural networks. Neural networks have great potential in many data-related applications. The usage of artificial neural networks allows for the efficient processing of a game's vast amounts of data, as well as the screening and training of non-playable characters to provide the player with the most relevant information possible.

2.3. Target detection algorithm and deep learning

Image categorization is taken to the next level with target detection algorithms, which not only identify target types but also locate and contextualize them inside images. Algorithms for detecting targets are one type of use of artificial neural networks. In this paper, we focus primarily on the One-stage algorithm (end-to-end) for target detection and recognition because, in comparison to reading the player's data directly from the game, the use of a target detection algorithm is more closely related to the player's operation, allowing the simulation of the player's aiming and shooting process via the target detection algorithm's configuration.

DOI: 10.54254/2755-2721/32/20230203

Because machine learning is essential to the development of artificial intelligence, its subfield, deep learning (also known as DL), is a promising new area of study. The study of artificial neural networks gave rise to the idea of deep learning; one type of deep learning structure is the multilayer perceptron, which consists of several hidden layers. When it comes to discovering high-level representations of attribute classes or features, deep learning excels because it combines low-level characteristics to generate more abstract high-level representations. The goal of deep learning research is to create neural networks that can learn analytically like the human brain [5]. These networks would then be able to understand data like sights, sounds, and texts in a manner similar to the human mind.

3. Realization and analysis of AI

3.1. Players' needs analysis and AI decision-making system

It is now necessary to consider the player's level of expertise in the game before attempting any one shot. Players that begin the game without the necessary operating skills or acquaintance with the map are frequently eliminated by other players. Therefore, players should be able to practice shooting against AIs of a similar level, and AI characteristics should be altered accordingly to accommodate players of varying skill levels. At the same time, players should be able to tailor the game's interface settings to their individual preferences. This is why it's important to implement player versus computer and character customization options in addition to the traditional player vs player action.

NPCs will be made at the point of resurrection and, once created, will make decisions based on data gathered from their immediate environment and from other NPCs, all in an effort to simulate the player's decision-making process. When a trigger is pulled in a first-person shooting game, the NPC will get a message and use the game's decision-making mechanism to determine how to respond. The game's decision-making algorithm decides how smart the AI is, and whether or not the character controlled by the AI can make the right call in the given situation and take the necessary action. The decision-making process is responsible for all of these results.

3.2. AI perception system

The perception system provides information for the AI decision-making system and is built with less complexity than other components. There are now three methods available to deliver information about the game world for AI characters that work together to form the game's perception system. The three techniques are called polling, event-driven, and trigger [6].

First, polling is the quickest and easiest approach to give the NPC player access to information about its environment. Polling is synonymous with querying. The system repeatedly checks the status of the sensor to see if a predetermined event has occurred, however this approach is flawed. It is tough to make modifications to the code when there are more players and NPCs since too much information is acquired in one polling and most of it is not obviously helpful for NPCs' activities. However, its useful features allow it to be used in less complex contexts.

In contrast to polling, an event-driven technique involves waiting for a certain time to pass before retrieving data and passing it on to other NPCs. To handle the player's reaction to the NPC's activity, event-driven time keeps track of potential future occurrences and reacts to the associated state whenever they occur. This method keeps an eye on the game's environment and reacts accordingly; this monitoring can be done at regular intervals to look for changes or in conjunction with triggers. After a certain amount of time has passed, a notification is sent to the AI character that was following the event [7].

Thirdly, we have "triggers," which function similarly to NPC-triggering devices in that they cause the relevant NPC (a non-player controlled intelligent character) to do an action when a certain event occurs. Set the login trigger when the player signs in, the scene trigger when they enter a new area, and the collision trigger when they collide with an object. The three methods are used together to enhance one another and create a more humane decision-making mechanism for the intelligent character in the game who is not controlled by the user.

DOI: 10.54254/2755-2721/32/20230203

3.3. Extensible state machine

Every choice the player makes in the game will alter the present situation and affect how the game evolves. The idea of state machines must be introduced if NPCs are to adapt to the player's actions. Due to their inability to scale, finite state machines are rarely a good choice for handling the game's many different scenarios. Instead, scalable state machines should be used [8]. By using scalable state machines, AI may improve its decision-making system to a larger extent, and avoid the challenges of scaling poorly. Each element t of T can be represented as a quintuple, source (t), target (t), event (t), condition (t), action (t)>, where S is the set of states, S0 is the accidental state, I is the set of input messages, V is the set of variables, O is the set of output messages, and T is an ensemble of state migrations. where source (t) is the migration's meta-state, target (t) is the destination state, event (t) is the incentive event on migration (t) consisting of a number of input variables or null, condition (t) is the precondition for t to be able to execute (or null), and action (t) is the action caused by migration (t) consisting of a series of variable assignment statements or output statements (or null). The result of migration t is represented by action(t), which may be a series of statements assigning values to variables, statements producing output, or nothing at all. To perform a migration, the EFSM must be in the target state target (t) and have received the incentive event event(t), all while the precondition condition(t) of the migration is True.

Starting from the game's initial state, the game's AI constructs a route for the NPC to follow throughout the game. In order to respond appropriately to the player's action, the non-player controlled intelligent character will transition to the desired state. This template represents the player's internal thought process in-game, as they assess the scenario and behave accordingly.

3.4. Target detection algorithms and deep reinforcement learning applications

When the intelligent, non-player character finally locates the player, it should begin firing at the player. A non-player controlled intelligent character should follow a similar procedure as the player during the operation, which entails finding the adversary, aiming at the enemy, and firing. For use in games, algorithms must be quick to react and light on processing resources. Therefore, a single-stage algorithm is more practical; his algorithm follows the principle of using a dedicated CNN model to accomplish end-to-end target detection; the computer image is sent to the CNN network in real-time for prediction; and finally, the detected target is processed based on the network's predictions. This mimics the action of a non-player character aiming at the player and firing once they have been identified as the target. The player's reaction time can be tested multiple times to get an average value, and then that value can be fed into a target detection algorithm to get the player image position after the target aiming time of the player, yielding a non-player controlled intelligent role with a level similar to that of the player.

Gaming can also benefit from the optimization of the decision-making system of AI characters controlled by the game's AI with the help of deep reinforcement learning. It will initially function in accordance with the pre-generated action line, keeping track of the current line's parameters and adding the reward process in accordance with the current performance. After being exposed to a large amount of data, the AI will be able to make decisions based on what it considers to be the best course of action in any given situation, much like a human player would. This includes summarizing relevant data, determining where threats are most likely to materialize, and making educated guesses about how to best respond.

4. Conclusion

The present study primarily functions as a comprehensive literature analysis on the utilization of ANNs within the gaming sector. This article undertakes an analysis of the requirements of gamers, taking into consideration their gaming experience and the prevailing market conditions, with the objective of integrating AI into games. Subsequently, the research proceeds to devise the decision-making system and perception system for the AI. The decision-making engine of the AI incorporates a scalable state machine to customize its actions according to the player's individual playstyle. In order to enhance the performance of AI, optimize the player's gaming experience, and improve the AI's capacity to fulfill

DOI: 10.54254/2755-2721/32/20230203

user requirements, the perception system grants the AI access to auditory and visual capabilities equivalent to those of the player.

This document has room for improvement as it mostly presents a conceptual framework for the game's design, without the practical implementation of the algorithms presented. Likewise, the availability of empirical evidence supporting the algorithm's feasibility is limited. Furthermore, the examination of the game's operation, as well as its continuous maintenance and optimization, lacks specific recommendations. However, further investigation and refinement in these domains are anticipated in the upcoming period, which will serve as substantiation for the program's feasibility.

Given the anticipated ongoing advancements and broadening applications of artificial neural networks, it is highly probable that their integration into video games will emerge as a prominent industry trend in the foreseeable future. The gaming experience for players will be further enhanced as improved algorithms are implemented alongside the upgrading or replacement of electronic components.

References

[1] Yin Yuhan, Yang Kaicheng. Comparison and analysis of e-sports and e-games [J]. The 14th National Academic Conference on Sports Information Technology, Chinese Society of Sports Science, 2022, 167-168.

[2] D. Tang. Design and realization of first person shooting game based on Unity3D engine [D]. University of Electronic Science and Technology, 2021.DOI:10.27005/d.cnki.gdzku.2021.002547.

[3] Xie Yangxiao. Research on next-generation game scene design under Unreal Engine [D]. Zhejiang University of Technology, 2018,33-35.

[4] Jian Zhang. Next-generation game engine design and implementation [D]. Beijing Jiaotong University, 2014,107-109.

[5] Zhang Jing. Research on the application of neural network in intelligent information processing [J]. Science and Technology Outlook, 2015, 25(24):5+7.

[6] Wei Jian,Liu Aijuan,Tang Jianwen. Research on alarm model of digital TV monitoring platform based on deep learning neural network technology[J]. Cable TV Technology,2017(07):78-82.DOI:10.16045/j.cnki.catvtec.2017.07.023.

[7] H.Y. Wang, M.Y. Chen, Y.N. Hua, Z.J. Shi. Unity3D artificial intelligence programming collection [M]. Tsinghua University Press, 2014: 141-142.

[8] Wang Yuan. Research on the application of artificial intelligence in shooting video games [D]. North China University of Water Conservancy and Hydropower, 2021.DOI:10.27144/d.cnki.ghbsc.2021.000001.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230204

A review of deep learning-based text sentiment analysis research

Wanlu She

School of Management and Economics, University of Electronic Science and Technology of China, Chengdu, Sichuan, China, 611731

sehhh1205@gmail.com

Abstract. The study of textual sentiments is a growing subfield of natural language processing. Research on deep learning-based text sentiment analysis approaches has received a lot of interest as machine learning technology has advanced. There are a variety of approaches to analyzing text for sentiment, but they may be broken down into three broad categories: those that rely on neural networks, those that introduce an attention mechanism, and those that rely on pre-trained models. In this paper, the author uses a literature review approach and CNKI as the search engine to examine previous studies on deep learning-based text sentiment analysis methods and models and to categorize the evolution of this field so as to aid in the development of similar studies in the future.

Keywords: text sentiment analysis, deep learning, neural networks, attention mechanisms.

1. Introduction

Sentiment analysis, also known as emotional disposition analysis or opinion mining [1], focuses on how to extract users' opinions, attitudes, and emotions, etc. from text, audio, or images, etc., among which textual sentiment analysis focuses on emotionally-charged text and mines its embedded emotional tendencies. Text sentiment analysis has grown in importance as a field of study within natural language processing during the past few years, necessitating the development of more sophisticated technical approaches. Previous related research identifies three basic approaches to text sentiment analysis: those that rely on a sentiment lexicon, those that use conventional machine learning, and those that use deep learning [2].

Deep learning is a significant subfield of machine learning theory that tries to model the human brain in order to more precisely and efficiently extract features from sample data for the purposes of machine recognition and analysis. Today, deep learning is widely used for several tasks, including speech recognition, image recognition, and NLP [3]. Deep learning for natural language processing was first explored in 2008 when Collobert et al. applied it to common NLP tasks like POS tagging using embeddings and multilayer one-dimensional convolutional structures [4]. This paper follows the framework established by the literature [2], which categorizes deep learning-based text sentiment analysis methods into four categories: analysis with a single neural network; analysis with a hybrid (combined, fused) neural network; analysis with the introduction of an attention mechanism; and analysis with a pre-trained model. This paper aims to examine deep learning's potential as a technical

DOI: 10.54254/2755-2721/32/20230204

solution for text sentiment analysis, to categorize the most important models of deep learning-based text sentiment analysis from the research literature available in CNKI, and to make suggestions for further study in this area.

2. Deep learning based approaches

2.1. Neural network

The interconnected neurons of a real neural network are mimicked in an artificial neural network. There are three layers to a neural network model: the input layer, the hidden layer, and the output layer. An input layer, where each neuron represents a feature and receives the input signal, is fully connected to a hidden layer, where the number of neurons varies based on the circumstance, and an output layer, where the number of neurons reflects the number of classifications, according to the literature [5]. These days, CNNs, RNNs, LSTMs, and Bi-LSTMs are the mainstays of neural network learning.

A convolutional neural network, or CNN, has a hidden layer made up of a convolutional layer, a pooling layer, and a fully connected layer. A convolutional group is made up of the convolutional layer and one or more of the pooling layers, which are selected in a random order and placed in an alternating pattern. The learning process of CNN proceeds layer-by-layer, from the local features to the global features, and ultimately achieves the classification through the fully connected layer. Experimental results on the COAE2014 Task 4 corpus showed that CNNs are effective for sentiment analysis of Chinese text [6], and researchers Liu Longfei, Yang Liang, Zhang Shaowu, et al. used word-level vectors of vocabulary as raw features for microblogging sentiment tendency analysis with CNN. Using character-level, vocabulary-level, and concept-level distributed representations as three-channel inputs, ICNNSTCM proposed by Gao Yunlong, Wu Chuan and Zhu Ming extracts their numerical features through convolutional operations, and adds sparse self-coding measure rate in the fully-connected layer to decrease the model's complexity while simultaneously increasing its generalization ability and experimental results [7].

Recurrent neural networks, or RNNs, are a type of neural network that are typically employed to handle sequences of varying lengths. In contrast to a CNN, an RNN may pick up on contextual semantics and use previously memorized information to better understand the context of the current event. Due to its susceptibility to gradient vanishing and explosion when processing lengthy text sequences, Hochreiter et al. proposed LSTM as a more robust structure of RNN [8]. Based on this, Niu Chengming, Zhan Guohua, and Li Zhihua proposed a Word2Vec and LSTM improved RNN model for Chinese microblogging sentiment analysis [9]. This involves replacing each hidden layer of RNN with a cell with memory function, which has a better processing capability for time series and linguistic text sequences. The Bi-LSTM model consists of two LSTM networks, each of which is fed into a separate LSTM neural network in forward and reverse order for feature extraction, and then the resulting feature vector is spliced together. Since a lot of linguistic knowledge and sentiment resources are underutilized in sentiment analysis tasks, Li Weijiang and Qi Fang proposed a multi-channel Bi-LSTM based sentiment analysis model called Multi-Bi-LSTM [10], which models the existing linguistic knowledge and sentiment resources in the sentiment analysis task, generates different feature channels, and allows the model to fully learn the sentence's sentiment information, it achieving better performance than Bi-LSTM, CNN combined with sentiment sequence features, and traditional classifiers in experiment.

Since traditional RNNs are unable to recall information for a long time and a single CNN cannot accurately represent the contextual semantics of text, Yang Yunlong, Sun Jianqiang, and Song Guochao proposed a sentiment analysis model GCaps that integrates GRU and capsule features [11], which captures the contextual global features of the text through GRU to obtain the overall scalar information, and then iteratively processes the captured information via a dynamic routing algorithm at the initial capsule level to obtain the vectorized feature information that represents the overall attributes of the text, and finally the combination between features is carried out in the main capsule part to obtain more accurate text attributes and analyze the sentiment polarity of the text according to the intensity of each feature. Experimental results on the benchmark dataset MR demonstrate that compared to CNN+INI

DOI: 10.54254/2755-2721/32/20230204

and CLCNN, GCaps achieves a 3.1% and a 0.5% improvement in classification accuracy. Liang Zhijian, Xie Hongyu and An Weigang proposed a text classification method based on Bi-GRU and Bayesian classifier [12]. They use Bi-GRU to extract text features, assign weights by TF-IDF algorithm, and use Bayesian classifier to discriminate classification, which improves the shortcomings of GRU's insufficient dependence on post-text, shortens the training time of the model, and improves the efficiency of text classification. Comparative simulation experiments are carried out on two types of text data, and the experimental results show that the classification algorithm can effectively improve the efficiency and accuracy of text classification compared with the traditional RNN.

Due to the fact that each neural network method has its own set of benefits and drawbacks, many researchers have turned to hybrid methods while doing text sentiment analysis. With the gradient explosion problem of RNN limiting the accuracy of text classification, and the existence of backward and forward dependency in the structure of natural language, Li Yang and Dong Hongbin proposed a CNN and Bi-LSTM feature fusion model that uses CNN and Bi-LSTM to extract the local features of text vectors and the global features related to the text [13]. Experimental results demonstrate that the suggested feature fusion model significantly boosts text categorization accuracy over competing methods. To address the shortcomings of previous approaches to text sentiment analysis—namely, inaccurate results, lengthy processing times, and a lack of relevant features—Zhao Hong, Wang Le, and Wang Weijie developed a Bi-LSTM-CNN serial hybrid model [14]. This model first employs Bi-LSTM to extract the text's context, then CNN to extract local semantic features from the extracted features, and finally Softmax to derive the text's emotional tendency. This model improves the overall evaluation index F1 by 1.86% compared to LSTM-CNN and by 0.76% compared to Bi-LSTM-CNN parallel feature fusion. A text sentiment analysis model that combines CNN and Bi-GRU was proposed by Miao Yalin, Ji Yichun, Zhang Shun et al. to address the issue of heavy workload caused by conventional sentiment analysis methods and the disregard of network training speed by most deep learning approaches [15]. Extracted via a CNN and Bi-GRU, the text's local static and sequential features are coupled to a GRU layer for further dimensionality reduction before being passed to Sigmoid for sentiment classification. Experiments conducted on the self-created Douban movie and TV review dataset demonstrate that, in comparison to the CNN-BLSTM model of the same complexity, this model improves classification accuracy by 2.52% and training rate by 41.43%.

2.2. Attention mechanisms

The first to introduce the attention mechanism into the field of natural language processing was Bahdanau et al. [16], who applied it to the field of machine translation. The nature of the attention mechanism is a set of weight values distribution, which is manifested in the field of natural language processing as words with higher weights are more important throughout the text and play a greater role in the overall classification task [17]. The attention mechanism can extend the ability of neural networks to approximate more complex functions to focus on specific parts of the input.

Feng Xingjie, Zhang Zhiwei, and Shi Jinchuan combined a convolutional neural network (CNN) with an attention model (AM) for text sentiment analysis, and their experiments demonstrated significant improvements in accuracy [18], recall, and F1 measure compared to both traditional machine learning methods and pure AM methods. Using deep learning for sentence-level sentiment analysis tasks as a starting point, Guan Pengfei, Li Baoan, Lyu Xueqiang, and others suggested an attention-enhanced Bi-LSTM model [19]. The model employs Bi-LSTM to learn the semantic information of the text, which improves the classification effect through parallel fusion, and the attention mechanism to learn the weight distribution of each word on the sentiment tendency of a sentence directly from the basis of word vectors. Experimental results on the NLPCC2014 sentiment analysis corpus show that this model performs better than others at classifying the tone of individual sentences. Shi Lei, Zhang Xinqian, and Tao Yongcai et al. built the SAtt-TLSTM-M model by combining the self-attention mechanism with the introduction of Maxout neurons at the output of the Tree-LSTM model in order to address the issues of information memory loss [20] and the negligence of the correlation between context-discontinuous words, and gradient dispersion in RNN models. Using the COAE2014 evaluation dataset, the accuracy

DOI: 10.54254/2755-2721/32/20230204

of the model was shown to be higher than that of the standard SVM, MNB, and LSTM models for sentiment analysis by a margin of 16.18%, 15.34%, and 12.05%, respectively. Zhang Jin, Duan Liguo, Li Aiping, and others suggested a text sentiment classification model that combines Bi-GRU-Attention and a gating mechanism to perform aspect-level fine-grained sentiment analysis based on user comments [21]. They combine the negation dictionary and lexical information to increase the user evaluation sentiment knowledge, and use the user evaluation sentiment knowledge as the user review sentiment feature information to integrate the existing sentiment resources. The seed sentiment dictionary is the HOWNET evaluation sentiment dictionary, and the SO-PMI algorithm is used to expand the user review sentiment dictionary. Next, they use Bi-GRU to do deep feature extraction on the text by introducing the word feature and sentiment feature information, and then combining these as model inputs. At the output layer, text sentiment analysis is conducted, and the final sentiment polarity is obtained via Softmax. This process begins with the acquisition of information about a text's aspect words and continues with the extraction of contextual sentiment features related to those words using the gating mechanism and the attention mechanism. Improved experimental outcomes are achieved by testing the model on the Chinese dataset of AIchallenger2018's fine-grained sentiment analysis, where it achieves a MacroF1score value of 0.7218.

Some researchers have also implemented the attention mechanism on the basis of hybrid neural networks, expanding on what has previously been done with single neural networks. A text categorization technique based on a hybrid model of LSTM-Attention and CNN was proposed by Teng Jinbao, Kong Weiwei, and Tian Qiaoxin et al. to address the limitation of standard LSTM and CNN in not being able to express the importance of each word in the text while extracting features [17]. After CNN has extracted the relevant local information from the text, the whole text semantics can be integrated. The attention mechanism is then introduced after LSTM to extract the attention score of the output information, and LSTM is utilized to extract the text context features. Finally, CNN's output is fused with LSTM-Attention's, enabling the focus of attention to be directed on key words via efficient feature extraction. The model's accuracy is 90.23%, and the F1 measure is 90.12%, according to experimental results on three public datasets; these numbers are higher than those of LSTM and CNN. Since the information between neurons in the same layer of traditional CNN cannot be transmitted to each other, unable to make full use of the feature information in the same level, and the lack of the representation of sentence system features leads to the limited feature learning ability of the model, Wang Liya, Liu Changhui, and Cai Dunbo et al. proposed a model based on the introduction of the attention mechanism of the joint CNN-Bi-GRU network [22]. The model employs CNN to extract deep phrase features, Bi-GRU for serialized information learning to obtain sentence system features, strengthened linking of CNN pooling layer features, and attention mechanism to complete effective feature screening. Experiments comparing the model to multiple groups on the dataset reveal that it is able to enhance text categorization accuracy for minimal effort and time investment. The deviation of weight allocation is caused by the fact that both the traditional self-attention mechanism and the Bi-GRU network disregard the local dependencies that exist between the texts, leading to inaccurate predictions. SAttBiGRUMCNN was proposed by Chen Kejia and Liu Hui [23]; it is a text categorization model built on Bi-GRU with an enhanced self-attention mechanism and a multi-channel CNN. Bi-GRU is utilized to obtain the text's contextual semantic information at the global level, and then the text's local features are extracted using the optimized multi-channel CNN. Based on this, they incorporate a position weight parameter and redistribute the self-attention weight probability value based on the text vector training location before using Softmax to classify the sample labels. The experimental findings on two standard datasets demonstrate the model's accuracy of 98.95% and 88.1%, respectively, outperforming FastText, CNN, RCNN, and other classification methods.

Furthermore, Gao Jiaxi and Huang Haiyan suggested a text sentiment analysis model based on the TF-IDF and multi-head attention Transformer model to deal with the issue that current computational methods cannot appropriately handle text datasets with high complexity and confusion [24]. The TF-IDF algorithm is used in the text preprocessing stage to initially screen the words that affect the text's sentiment tendency to a greater extent, while ignoring the common stop words and other proper nouns

DOI: 10.54254/2755-2721/32/20230204

that have less influence. Finally, the Transformer model encoder trained with multi-head attention is employed for feature extraction to further understand and generalize the text's semantics. When applied to the Ec_60k dataset, for instance, the model has an accuracy of 98.17%.

2.3. Pre-training model

In NLP, pre-trained models are deep network architectures that have been trained on a large, unlabeled text corpus. Word2vec, a static pre-training model, has been the most popular text representation method since the advent of NLP technology; however, it has the flaws of learning only a shallow representation of the text and being context-independent, so it has little bearing on the improvement of subsequent tasks. After much investigation, researchers created dynamic pre-training models (ELMo and BERT, primarily) to address the aforementioned issues. Since the introduction of the BERT model ushered in a new age, numerous more pre-training models have been developed, most of which may be categorized as either better BERT-based models or XLNet [25].

The ELMo model is made up of a forward and backward Bi-LSTM language model and an objective function to maximize the likelihood of the model. In contrast to more standard word vector models, this one allows each word to map to exactly one word vector. To address the limitation that word embedding methods like Word2Vec and GloVe only generate a single semantic vector for polysemous words, Zhao Yaou, Zhang Jiachong, Li Yibin, et al. presented an ELMo-MSCNN hybrid model for sentiment analysis [26]. The model learns the pre-trained corpus with ELMo, generates contextually relevant word vectors, and initializes the embedding layer of ELMo with the pre-trained Chinese character vectors, which accelerates the training speed and improves the training accuracy; then the features of word vectors are extracted twice with MSCNN, and feature fusion is performed to generate the overall semantic representation of the sentence; finally, the classification of textual sentiment tendency is realized after Softmax excitation function. Experiments are run on two publicly available datasets (hotel reviews and NLPCC2014 task2), with the results showing an improvement in model accuracy of 1.08% compared to the attention-based Bi-LSTM model on the hotel reviews dataset and an improvement in accuracy of 2.16% compared to the LSTM-CNN model on the NLPCC2014 task2 dataset. To get around the fact that CNN is unable to directly extract bidirectional semantic features of sentences and that conventional word embedding methods are unable to effectively represent the multiple meanings of a word, Zhao Yaou, Zhang Jiachong, Li Yibin et al. proposed a hybrid model based on ELMo and Transformer [27]. This model employs the ELMo model to produce word vectors, incorporates contextual features of the sentence into the word vectors, and produces distinct semantic vectors for the various semantics of polysemous words; the obtained ELMo word vectors are then fed into the Transformer model with modified Encoder and Decoder structures for sentiment classification. Since the model can extract semantic properties of sentences from multiple perspectives, the resulting semantic information is completer and more nuanced. The experimental results show that compared with the current mainstream methods, this one improves the classification accuracy by 3.52% on the NLPCC2014 Task2 dataset, and by 0.7%, 2%, 1.98%, and 1.36%, respectively, on the four subdatasets of hotel reviews. It was proposed by Wu di, Wang Ziyu, and Zhao Weichao that they use a model called ELMo-CNN-Bi-GRU [28]. The model uses ELMo and Glove pre-training models to generate dynamic and static word vectors, respectively, generates input vectors by stacking and embedding the two word vectors, and adopts the self-attention mechanism to process the inputs; then, the internal word dependencies are calculated to construct a dual-channel neural network structure fusing CNN and Bi-GRU, and the text local features and global features are acquired simultaneously; finally, the dual-channel neural network structure is trained to predict the Compared to the H-Bi-GRU model, which excels among similar sentiment classification models, the experimental findings reveal that the model improves accuracy and F1 value on the IMDB, yelp, and sentiment140 datasets.

The BERT model, first presented by Google, takes into account the contextual semantics of words by using WordPiece embedding as word vectors in the input, as well as position vectors and sentence tangent vectors, and the bidirectional transformer mechanism for the language model [29]. Chen Zhiqun and Ju Ting used the BERT model to extract semantic feature representations of microblog comment

DOI: 10.54254/2755-2721/32/20230204

texts, and then fed the acquired word semantic features into the Bi-LSTM model for propensity classification to address the issue of word polysemy that cannot be solved by traditional language models in word vector representations [30]. Experiments conducted on Sina Weibo comment data demonstrate that this model achieves an F1 value of 91.45%, placing it above other widely used propensity analysis models. In order to address the imbalance in evaluation, Liu Ji and Gu Fengyun proposed the model M2BERT-BiLSTM [31], which combines BERT and Bi-LSTM. The model first converts the sequences in the hidden layer of the BERT model into vectors, then splices them along the dimensions of sentence length according to the pooling of the mean and maximum values, and finally inputs the semantic features of the spliced words into the Bi-LSTM for the textual sentiment analysis in order to alleviate the imbalance in the evaluation. The experimental findings demonstrate that the model provides a more accurate assessment of the indicators. To address the limitations of sentiment characteristics derived by conventional pre-training models, Huang Zemin and Wu Yinggang suggested a BERT-CNN(LRN)-BiSRU model for text sentiment classification [32], which combines BERT with a convolutional bidirectional simple recurrent network. In order to obtain dynamic word vectors that incorporate sentence context, the model is pre-trained with BERT; the word vector features are then extracted twice with a multi-granularity convolutional neural network; the pooled and injected into the local response normalization layer LRN to normalize the feature map; and finally, the two-way simple loop unit is used to further learn the contextual semantic information. The model achieves a high F1 value of 91.27% in experiments, demonstrating its usefulness in real-world settings. The BERT model has been further improved by the addition of the attention mechanism proposed by a number of academics. An integrated BERT-TCN and attention mechanism model for text sentiment analysis was proposed by Zhang Jian [33]. The transformer-based BERT model is used first to obtain the text word vectors containing contextual semantics; the TCN model is then used to further extract the contextual semantic features of the text word vectors; the attention mechanism is then introduced to focus on the crucial sentiment features in the context; and finally, the Softmax classifier is used to perform the sentiment classification. The experimental results suggest that the model outperforms the BERT-TCN model in terms of precision, recall, and F1 score.

The XLNet model is an improvement over the BERT model in that it can be trained in a bidirectional context without having to mask some of the predicted words. Li Dongjin, Shan Rui, and Yin Liangkui et al. proposed a text sentiment analysis model combining a generalized autoregressive pretrained language model, using XLNet to represent text features, extracting local features of text vectors through CNN, then using Bi-GRU to extract deeper contextual information of the text, introducing an attention mechanism to assign different weights to the features according to their importance, and performing a text sentiment polarity analysis [34]. The simulation experiment contrasts the accuracy of this model to that of five commonly used sentiment analysis models, and the results demonstrate that this model is significantly more accurate.

3. Conclusion

Now more than ever before, people use the Internet to share their opinions and emotions with the world. Interaction via text not only replicates occurrences but also conveys feelings. Given this context, text sentiment analysis becomes a powerful tool for understanding the state of affairs and making informed predictions about the future, with potential uses in areas as diverse as public opinion research, user profiling, and more. This paper examines the use of deep learning in text sentiment analysis technology by reviewing relevant CNKI articles and providing an explanation of how it works. Three primary technique models are introduced, and their potential enhancements are discussed.

Deep learning-based approaches for text sentiment analysis start with CNN and to improve the problem of accessing contextual semantics, RNN is used to apply memorized content. In addition to CNN and RNN, the researchers suggested LSTM and GRU for model optimization, and then suggested Bi-LSTM and Bi-GRU by rearranging the input order. Scholars are also interested in studying and applying hybrid neural network models, with the majority of this work focusing on CNNs paired with LSTMs or GRUs. Researchers have built on this foundation by adding attention processes to the model

DOI: 10.54254/2755-2721/32/20230204

or by employing pre-trained models to enable horizontal model extension. When training on a large scale with the dataset, the pre-training model learns contextual semantics and the semantics of polysemous words, which greatly improves the classification effect. The attention mechanism uses the importance of words in the text to better capture the contextual information, leading to more accurate text classification. Text sentiment analysis is a promising area of study. Researchers in this area have developed numerous approaches and models for use in real-world settings, and it is believed that this field will have richer research results in the future.

References

[1] Zhong Jiawa, Liu Wei, Wang Sili, et al. Review of Methods and Applications of Text Sentiment Analysis [J]. Data Analysis and Knowledge Discovery, 2021, 5(6): 1-13.

[2] WANG Ting, YANG Wenzhong. Review of Text Sentiment Analysis Methods [J]. Computer Engineering and Applications, 2021, 57(12): 11-24.

[3] Yu Kai, Jia Lei, Chen Yuqiang et al. Deep Learning: Yesterday, Today, and Tomorrow [J]. Journal of Computer Research and Development, 2013, 50(09): 1799-1804.

[4] Collobert R, Weston J, Bottou L, et al. Natural language processing (Almost) from scratch [J]. Journal of Machine Learning Research, 2011, 12: 2493-2537

[5] ZHOU Fei-Yan, JIN Lin-Peng, DONG Jun. Review of Convolutional Neural Network. A review of convolutional neural network research [J]. Chinese Journal of Computers, 2017, 40(06): 1229-1251.

[6] LIU Longfei, YANG Liang, ZHANG Shaowu et al. Convolutional Neural Networks for Chinese Micro-blog Sentiment Analysis [J]. Journal of Chinese Information Processing, 2015, 29(06): 159-165.

[7] GAO Yunlong, WU Chuan, ZHU Ming. Short Text Classification Model Based on Improved Convolutional Neural Network [J]. Journal of Jilin University (Science Edition), 2020, 58(04): 923-930.DOI: 10.13413/j.cnki.jdxblxb.2019422.

[8] Hochreiter S, Schmidhuber J. Long short-term memory [J]. Neural Computation, 1997, 9(8) : 1735-1780.

[9] NIU Cheng-Ming, ZHAN Guo-Hua, LI Zhi-Hua. Chinese Weibo Sentiment Analysis Based on Deep Neural Network [J]. Computer Systems & Applications, 2018, 27(11): 205-210.DOI: 10.15888/j.cnki.csa.006645.

[10] LI Weijiang, QI Fang. Sentiment Analysis Based on Multi-Channel Bidirectional Long Short Term Memory Network [J]. Journal of Chinese Information Processing, 2019, 33(12): 119-128.

[11] YANG Yunlong, SUN Jianqiang, SONG Guochao. Text sentiment analysis based on gated recurrent unit and capsule features [J]. Journal of Computer Applications, 2020, 40(09): 2531-2535.

[12] LIANG Zhi-jian, XIE Hong-yu, AN Wei-gang. Text classification based on bidirectional GRU and Bayesian classifier [J]. Computer Engineering and Design, 2020, 41(02): 381-385. DOI: 10.16208/j.issn.1000-7024.2020.02.013.

[13] LI Yang, DONG Hongbin. Text sentiment analysis based on feature fusion of convolution neural network and bidirectional long short-term memory network [J]. Journal of Computer Applications, 2018, 38(11): 3075-3080.

[14] ZHAO Hong, WANG Le, WANG Weijie. Text sentiment analysis based on serial hybrid model of bi-directional long short-term memory and convolutional neural network [J]. Journal of Computer Applications, 2020, 40(01): 16-22.

[15] MIAO Ya-lin, JI Yi-chun, ZHANG Shun et al. Application of CNN-BiGRU Model in Chinese Short Text Sentiment Analysis [J]. Information Science, 2021, 39(04): 85-91.DOI: 10.13833/j.issn.1007-7634.2021.04.012.

[16] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to alignand translate [C] //Proc of International Conferenceon Learning Representations. 2015.

DOI: 10.54254/2755-2721/32/20230204

[17] TENG Jinbao, KONG Weiwei, TIAN Qiaoxin et al. Text Classification Method Based on LSTM-Attention and CNN Hybrid Model [J]. Computer Engineering and Applications, 2021, 57(14): 126-133.
[18] Feng Xingjie, Zhang Zhiwei, Shi Jinchuan. Text sentiment analysis based on convolutional neural networks and attention model [J]. Application Research of Computers, 2018, 35(05): 1434-1436.
[19] GUAN Pengfei, LI Bao'an, LV Xueqiang et al. Attention Enhanced Bi-directional LSTM for Sentiment Analysis [J]. Journal of Chinese Information Processing, 2019, 33(02): 105-111.
[20] SHI Lei, ZHANG Xin-qian, TAO Yong-cai et al. Sentiment Analysis Model with the Combination of Self-attention and Tree-LSTM [J]. Journal of Chinese Computer Systems, 2019, 40(07): 1486-1490.
[21] ZHANG Jin, DUAN Li-guo, LI Ai-ping et al. Fine-grained Sentiment Analysis Based on Combination of Attention and Gated Mechanism [J]. Computer Science, 2021, 48(08): 226-233.
[22] WANG Liya, LIU Changhui, CAI Dunbo et al. Chinese text sentiment analysis based on CNN-BiGRU network with attention mechanism [J]. Journal of Computer Applications, 2019, 39(10): 2841-2846.
[23] CHEN K J, LIU H. Chinese text classification method based on improved BiGRU-CNN [J]. Computer Engineering, 2022, 48(5): 59-66, 73.
[24] GAO Jiaxi, HUANG Haiyan. Text Emotion Analysis Based on TF-IDF and Multihead Attention Transformer Model [J/OL]. Journal of East China University of Science and Technology: 1-8[2023-07-09]. DOI: 10.14135/j.cnki.1006-3080.20221218002.
[25] LI Zhou-jun, FAN Yu, WU Xian-jie. Survey of Natural Language Processing Pre-training Techniques [J]. Computer Science, 2020, 47(03): 162-173.
[26] ZHAO Ya'ou, ZHANG Jiachong, LI Yibin et al. Sentiment analysis using embedding from language model and multi-scale convolutional neural network [J]. Journal of Computer Applications, 2020, 40(03): 651-657.
[27] ZHAO Yaou, ZHANG Jiachong, LI Yibin et al. Sentiment Analysis Based on Hybrid Model of ELMo and Transformer [J]. Journal of Chinese Information Processing, 2021, 35(03): 115-124.
[28] WU D, WANG Z Y, ZHAO W C. ELMo-CNN-BiGRU dual-channel text sentiment classification model [J]. Computer Engineering, 2022, 48(8): 105-112.
[29] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: Association for Computational Linguistics, 2019: 4171-4186.
[30] Chen Z Q, Ju T. Research on Tendency Analysis of Microblog Comments Based on BERT and BLSTM [J]. Information studies: Theory & Application, 2020, 43(08): 173-177. DOI: 10.16353/j.cnki.1000-7490.2020.08.026.
[31] Liu Ji, Gu Fengyun. Unbalanced Text Sentiment Analysis of Network Public Opinion Based on BERT and BiLSTM Hybrid Method [J]. Journal of Intelligence, 2022, 41(04): 104-110.
[32] Huang Zemin, Wu Yinggang. TEXT EMOTION ANALYSIS BASED ON BERT AND CNN-BISRU [J]. Computer Applications and Software, 2022, 39(12): 213-218.
[33] ZHANG Jian. Chinese Text Sentiment Analysis Model Based on BERT-TCN and Attention Mechanism [J]. Information & Computer, 2022, 34(22): 77-82.
[34] LI Dongjin, SHAN Rui, YIN Liangkui et al. Sentiment analysis of Chinese text based on XLNet [J]. Journal of Yanshan University, 2022, 46(06): 547-553.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230205

A study on the key technologies and existing challenges in the development of autonomous vehicles

Kunhua Su

Beijing Institute of Technology, Zhuhai, 519088, China

3275420748@qq.com

Abstract. As an emerging means of transportation, autonomous driving has great development potential and application prospects. This paper investigates relevant issues of autonomous vehicles, aiming to explore the technical principle and implementation method of autonomous driving and evaluate its performance and effect in road traffic. By analyzing the existing related research results and practical experience, this paper conducts an in-depth study on key technologies of autonomous vehicles, including sensor and sensing systems as well as decision-making and control systems. Besides, the paper also summarizes the ethical and safety problems in the course of autonomous driving development and puts forward a deep learning-based control strategy for autonomous driving. This strategy has high accuracy and stability in different scenarios and can effectively improve the safety and performance of autonomous vehicles. To sum up, this paper carries out in-depth research on autonomous vehicles and tries to propose innovative solutions in key technical fields. This study provides reliable theoretical and technical support for the development and application of autonomous vehicles and promotes progress and innovation in the field of transportation.

Keywords: autonomous vehicles, technical principle, deep learning, control strategy, security.

1. Introduction

With the rapid development of science and technology and the increasing demand for traffic safety and travel efficiency, autonomous vehicles are regarded as an important representative of future transportation modes. By introducing advanced sensor and sensing systems as well as decision-making and control systems, autonomous vehicles can drive autonomously, having comprehensive environmental perception, decision-making reasoning, and precise control capabilities [1].

This paper aims to systematically study the related technical principles and implementation methods of autonomous vehicles and evaluate their performance and effect in road traffic. Through the comprehensive analysis of the existing related research results and practical experience, this paper studies sensor and sensing systems as well as decision-making and control systems of autonomous vehicles, and on this basis, puts forward the existing challenges and corresponding solutions to autonomous vehicles in the development process.

This study has important theoretical and practical significance. First of all, as a new technology, autonomous driving is of great significance to be further studied in order to promote innovation and development in the field of transportation. Secondly, an in-depth study of the technical principles and implementation methods of autonomous vehicles will help improve their performance, thus

DOI: 10.54254/2755-2721/32/20230205

contributing to the realization of the goals of intelligent transportation systems. Finally, by evaluating the safety and performance of autonomous vehicles, this study can provide a scientific basis for the government and enterprises to formulate relevant policies and strategies.

2. The development of autonomous vehicles

The development of autonomous vehicles can be traced back to the early 1950s. At that time, scientists began to explore automatic driving technology. Early automatic driving technology mainly relied on mechanical and electronic equipment, such as mechanical trackers and electronic compasses.

With the continuous development of science and technology, autonomous vehicles have developed rapidly in recent decades. In the 1960s, engineers began to research and develop vehicle prototypes that can drive automatically. However, due to the limitation of computing power and sensing technology at that time, these prototypes still had many difficulties and challenges.

In the 1980s, with the rapid development of computer technology, autonomous driving technology made a great breakthrough. Researchers began to introduce systems based on computer vision and sensor technology to provide necessary information and real-time data for autonomous vehicles. These systems can detect and identify traffic signs and vehicles on the road, and process and analyze them through algorithms, so as to realize autonomous navigation and driving of vehicles.

As time goes by, there are more and more experiments and research and development (R&D) of autonomous vehicles. Many large technology companies and automobile manufacturers have joined the competition of autonomous driving technology. They have invested a lot of time and money in the R&D of autonomous driving algorithms and systems. At the same time, experimental vehicles are also tested in various road environments and scenes to verify the feasibility and practicability of these technologies.

In the experimental and R&D stage, researchers and engineers have done a lot of testing and verification. First, they use the simulation platform and the actual environment to simulate and evaluate the performance and safety of autonomous vehicles. In this way, researchers can find and solve potential problems in advance to ensure the stability and reliability of autonomous vehicles on the actual road.

At the same time, experiments and R&D also involve a large number of test vehicles and test sites. Researchers choose different road conditions and weather conditions to conduct various tests on autonomous vehicles, including complex scenes such as straight driving, turning, and overtaking. Through these experiments, researchers can collect a large amount of data to analyze and optimize the decision-making and control system of autonomous vehicles. In addition, they will also conduct experiments on sensors and sensing systems to ensure that they can accurately perceive the surrounding environment.

Cooperation is very important in the experimental and R&D stages. Scientific research institutions, automobile manufacturers, and technology companies have established close cooperative relations to jointly promote the progress of autonomous vehicles. By sharing resources and experience, they can speed up the process of technology R&D and commercialization.

In short, the initial exploration of autonomous driving technology is a process full of challenges and opportunities. By introducing computer vision and sensor technology, researchers make autonomous vehicles more intelligent and reliable. The experiment and R&D stage of autonomous vehicles is very important, and lays the foundation for the development of autonomous driving technology. Through a lot of testing and verification work, researchers can continuously optimize the performance and safety of autonomous vehicles and solve the challenges that arise. With the continuous progress of technology, it is believed that autonomous vehicles will become an important part of the future transportation system.

DOI: 10.54254/2755-2721/32/20230205

3. The technical principle of autonomous vehicles

3.1. Sensors and sensing systems

The sensors and sensing systems of autonomous vehicles are the key components to realize autonomous navigation and environmental awareness. The sensor system includes laser radar, camera, radar, and other equipment, which is used to collect the information around the vehicle [2]. In autonomous vehicles, laser radar is a commonly used sensor, which calculates the distance and shape of an object by emitting a laser beam and measuring the time it takes for the laser to reflect from the object. The camera is used to shoot images on the road and detect and identify road signs, vehicles, and pedestrians through image recognition technology. Radar can help vehicles perceive static and dynamic obstacles around them. Sensors play a vital role in the perception system of autonomous vehicles. They acquire and process information about the surrounding environment of vehicles through different technologies.

The sensing system is another important part of autonomous vehicles, which uses the data provided by sensors to understand and perceive the surrounding environment. By analyzing and processing the sensor data, the sensing system can determine the road geometry, identify different types of traffic participants, and predict their behaviors and intentions. For example, when an autonomous vehicle drives into an intersection, the sensing system can judge the state of traffic lights by analyzing the images provided by the camera, and then decide the driving direction and speed of the vehicle. The perception system can also detect and predict the trajectories of other vehicles and pedestrians by analyzing radar data, so as to make corresponding safety responses.

The accuracy and reliability of sensors and sensing systems directly affect the driving safety and effect of autonomous vehicles. Therefore, R&D personnel constantly strive to improve the sensor technology and the accuracy and performance of the sensor, in order to cope with various complex road conditions and environmental changes. For example, laser radar can improve the recognition accuracy and tracking performance of surrounding objects by increasing the density of the laser beam and improving the algorithm. At the same time, the algorithm of the sensing system is constantly optimized and improved to improve the ability of understanding and predict the environment.

In a word, sensors and sensing systems are important components of autonomous navigation and environmental awareness for autonomous vehicles. They provide key sensing functions for autonomous vehicles by collecting and analyzing information around vehicles. It is one of the important technical challenges to realize the commercial application of autonomous vehicles to continuously improve the accuracy of sensors and the reliability of sensing systems. With the continuous development and progress of technology, it is believed that the performance of sensors and sensing systems will continue to improve, providing strong support for realizing safer and smarter autonomous vehicles.

3.2. Decision-making and control system

The decision-making and control system of autonomous vehicles is the core technology to realize its autonomous driving function. The system continuously collects, analyzes, and processes the environmental information obtained by sensors and monitors the vehicle status in real time, so as to make decisions and control driving according to the current road conditions. Among them, the decision-making module is mainly responsible for real-time road analysis and target detection according to the data provided by the perception system, judging obstacles, traffic signs, and vehicles in the surrounding environment, and predicting their behavior. The control module controls the steering wheel, brake, and throttle of the vehicle according to the output of the decision-making module, so as to realize accurate lateral and longitudinal control.

Road analysis is one of the key steps in the decision-making module. Through deep learning and computer vision analysis of the environment in front of the vehicle, the system can identify information such as lane lines, road signs, and road boundaries, and then determine the driving path and vehicle driving planning. At the same time, the target detection module can identify vehicles,

DOI: 10.54254/2755-2721/32/20230205

pedestrians, and obstacles in the surrounding environment, providing important references for vehicle safety decision-making. These decision results will be passed to the control module to perform corresponding operations.

After obtaining the output of the decision-making module, the control module changes the motion state of the vehicle by adjusting the steering wheel angle and braking pressure and throttle opening of the vehicle in real time. Among them, the lateral control is mainly realized by the rotation of the steering wheel. According to the deviation between the target path and the vehicle position, autonomous vehicles can carry out accurate road maintenance, including lane maintenance and lane change. The longitudinal control mainly adjusts the speed and acceleration of the vehicle by controlling the throttle and brake, so as to ensure a safe distance from the preceding vehicle and realize a smooth acceleration and deceleration [3].

The key to decision-making and control systems is real-time and reliability. In order to achieve rapid decision-making and precise control, autonomous vehicles usually use high-performance computing platforms and real-time operating systems to process massive perceptual data and execute complex algorithms. At the same time, in order to ensure the reliability of the system, the decision-making and control system should have multi-level redundancy and fault-tolerant mechanism to deal with sensor failures, algorithm errors, and unexpected situations.

Generally speaking, the decision-making and control system of autonomous vehicles is the core technology to realize its autonomous driving function. Through continuous optimization and development, the decision-making and control system will provide better protection for the safety, reliability, and comfort of autonomous vehicles, and further promote the wide application and market development of autonomous driving technology.

4. Challenges of autonomous vehicles and corresponding solutions

4.1. Legal and moral issues

The development of autonomous vehicles has brought a series of legal and moral problems, which is a challenging task. First of all, a major legal and moral issue is the definition of responsibility [4]. When a traffic accident happens, who should bear the responsibility is always something people will discuss. Some agree that it is the responsibility of the vehicle manufacturer while others believe that it is the responsibility of the vehicle owner or the driver himself. Nowadays, drivers need to obey traffic rules and be responsible for their actions. However, when the vehicle changes from manual operation to automatic driving, the responsibility problem becomes complicated. Since self-driving vehicles have the ability of independent decision-making and operation, responsibility is a complex problem to be solved when there is a traffic accident.

Secondly, autonomous vehicles also involve issues of user privacy and data security. In order to achieve autonomous driving, vehicles need to collect a large amount of data and process and analyze it through the cloud. These data include internal and external information about the vehicle, which may involve the driver's personal privacy. Therefore, how to ensure the security and privacy of these data is an urgent problem to be solved. At the same time, how to prevent these data from being abused and misused is also an important legal and moral consideration.

In addition, autonomous vehicles have also caused a series of moral and ethical problems. For example, in an emergency, autonomous vehicles may need to make choices, for instance, whether to hit obstacles to avoid hitting pedestrians or to hit pedestrians to protect passengers' safety. These selective decisions involve ethical and moral standards. How to ensure that the automatic driving system can make these decisions fairly and follow social values has become a problem that needs attention.

Therefore, it is urgent to solve the legal and moral problems of autonomous vehicles. This requires the joint efforts of the government, automobile manufacturers, technical experts, and all walks of life. The government needs to clearly stipulate the responsibility, data security, and privacy protection of autonomous vehicles in legislation; automobile manufacturers need to establish a reliable automatic

DOI: 10.54254/2755-2721/32/20230205

driving system and ensure its compliance; technical experts need to strengthen research to make the automatic driving system more reliable and safe; all sectors of society need to have full discussions and debates to reach a consensus to ensure the rational and responsible development of autonomous driving technology. Only by fully considering and responding to legal and moral issues can the sustainable development of autonomous vehicles be promoted.

4.2. Safety and reliability issues

The safety and reliability of autonomous vehicles is also a very critical issue. The instability of technology and loopholes in the system may lead to serious safety accidents and traffic chaos [5].

In order to improve the safety and reliability of autonomous vehicles, researchers have adopted many innovative solutions. First of all, they constantly improve and update sensors and sensing systems to improve the vehicle's perception of its surrounding environment. The update of the sensor system ensures that vehicles can accurately identify and understand various objects on the road, including other vehicles, pedestrians, and obstacles. The development of this advanced perception system enables autonomous vehicles to better adapt to the complex traffic environment and respond accordingly.

Secondly, the improvement of decision-making and control systems is also an important step to improve the safety and reliability of autonomous vehicles. Using advanced algorithms and technologies, researchers have designed a system that can make decisions quickly and accurately and a control system that can accurately control the running of vehicles. These improvements ensure that autonomous vehicles can run more stably and reliably in various complex traffic scenes and reduce accidents.

In addition, strengthening the research on the legal and moral issues of autonomous vehicles can provide better guidance for the improvement of safety and reliability. These issues include the interaction between autonomous vehicles and other vehicles and pedestrians, the distribution of responsibilities and privacy protection. Researchers actively cooperate with legal and ethical experts to discuss and formulate relevant laws and ethical standards to ensure the normal operation and social acceptance of autonomous vehicles.

In a word, the improvement of the safety and reliability of autonomous vehicles is inseparable from the continuous improvement of sensors and sensing systems, the optimization of decision-making and control systems, and the full study of legal and moral issues. The implementation of these solutions will lay a solid foundation for the commercial application of autonomous vehicles and provide a positive impetus for the development of intelligent transportation systems in the future.

5. Conclusion

After in-depth research and exploration on the related issues of autonomous vehicles, this paper summarizes the technological exploration and research and development process of autonomous vehicles, and deeply studies the sensors and sensing systems as well as decision-making and control systems of autonomous vehicles, but these key technologies still need to be further optimized and improved. For example, in the complex and changeable traffic environment, the perception ability and decision-making ability of autonomous vehicles are still facing challenges, and more accurate and reliable algorithms and models are needed to improve their performance. Secondly, legal and moral issues are important considerations for the development of autonomous vehicles [6]. At present, the laws, regulations, and ethics of autonomous vehicles are not perfect. It is necessary to strengthen the research and formulation of relevant laws, regulations, and ethics while developing technology to ensure the safety and legitimacy of autonomous vehicles.

Based on the above research, future research can be carried out from the following aspects. First of all, the sensors and sensing systems as well as decision-making and control technology of autonomous vehicles can be further optimized and enhanced to improve their performance and safety. Secondly, the research and formulation of laws, regulations, and ethics of autonomous vehicles can be strengthened to provide a more reliable and legal environment for the application and development of

DOI: 10.54254/2755-2721/32/20230205

autonomous vehicles [6]. In practice, the practical application and verification of autonomous vehicles can be further developed to verify and evaluate the performance and feasibility of autonomous vehicles. At the same time, it can also strengthen the collaborative research between autonomous vehicles and other means of transportation and infrastructure, and promote the seamless docking and optimization of autonomous vehicles and transportation systems.

In short, as a new mode of transportation, autonomous vehicles have great development potential and application prospects. This study provides an important reference and basis for promoting the development and application of autonomous vehicles by deeply exploring the technical principles and implementation methods of autonomous vehicles. Future research and practice will further improve the key technologies, laws, regulations, and ethics of autonomous vehicles, and promote the safety, reliability, and popularization of autonomous vehicles.

References

[1] Zheng, H. Y., Shang, M. Y. and Zhou, B. L. (2020). Driverless taxis will reshape cities. Encyclopedia Forum.

[2] Liu, T., Wang, X., Xing, Y., Gao, Y., Tian, B. and Chen, L. (2019). Parallel driving system and application based on digital quadruplets. Journal of Intelligent Science and Technology, (01), 40-51.

[3] Liu, Z. Q. (2020). Design and implementation of path planning and control system for unmanned low-speed electric vehicles (Master's Thesis, University of Electronic Science and Technology of China).

[4] Zhai, J. Y. (2019). A brief discussion on the subject of tort liability of autonomous vehicle. Chizi, 000(011). doi: 218.10.3969/j.issn.1671-6035.2019.11.194.

[5] Shi, Y. X. (2021). Research on autonomous driving decision algorithm based on LSTM and grasshopper optimization algorithm (Master's Thesis, Jilin University).

[6] Peng, Z. H. (2020). Research on ethical risks in the development of autonomous vehicles (Master's Thesis, Wuhan University of Technology).

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230206

Ethical research on the artificial intelligence training system for preschoolers under the guidance of child-centered theory

Xi Liu

Faculty of Humanities, The Education University of Hong Kong, 10 Lo Ping Road, Tai Po, New Territories, Hong Kong, China

286274726@qq.com

Abstract. In recent years, with the rapid development of artificial intelligence technology, early childhood artificial intelligence training systems have gradually been applied in the field of education. However, there is still a lack of systematic research and exploration on the ethical issues of artificial intelligence training systems for young children. This study is guided by the child-centered theory and aims to explore the ethics of artificial intelligence training systems for young children, and propose corresponding solutions. Firstly, the development status and ethical issues of artificial intelligence training systems for young children were analyzed through literature review. Then, based on the principle of child centeredness, an ethical evaluation was conducted on the design and use of artificial intelligence training systems for young children. Finally, specific suggestions were proposed to protect children's rights and promote their development in the early childhood artificial intelligence training system, and future research was prospected.

Keywords: child-centered theory, early childhood artificial intelligence training system, ethical research, children's rights, children's development.

1. Research background and purpose

1.1. Child-centered theory and its application in education

Child-centered theory is a theoretical perspective based on the rights and needs of children, emphasizing the centrality of children in education and social development, and promoting their comprehensive development by respecting their rights and paying attention to their needs and interests. In education, child-centered theory guides child-centered educational practices, including the application of the following aspects: firstly, respecting children's subjectivity. Children are considered to have the ability to think independently, express themselves, and participate in decision-making. Educators establish equal cooperative relationships with children, fully respecting their opinions and ideas. Secondly, pay attention to the developmental characteristics of children. The physical and mental development of children has certain patterns and differences. Educators need to understand the cognitive, emotional, and social development characteristics of children, teach them according to their aptitude, and provide suitable learning environments and resources. Thirdly, protect the rights and interests of children. Children enjoy a series of rights in the education process, such as receiving equal educational opportunities, being respected, and being protected from abuse and discrimination.

DOI: 10.54254/2755-2721/32/20230206

Educators should take measures to ensure that children's rights are effectively protected. Finally, cultivate children's awareness of participation and provide personalized support and assistance. Children are encouraged to participate in school and community affairs by participating in decision-making, expressing opinions, and showcasing their abilities, cultivating their sense of participation and responsibility, and enhancing their self-confidence and social adaptability [1].

1.2. Concept and development of artificial intelligence training systems for young children

The early childhood artificial intelligence training system refers to an educational tool or platform designed and developed specifically for young children using artificial intelligence technology, aiming to provide a personalized, interactive, and autonomous learning environment, promoting the cognitive, language, emotional, and social development of young children. With the rapid development of artificial intelligence technology, artificial intelligence training systems for young children are constantly emerging and widely used. The artificial intelligence training system for young children can provide personalized learning content and paths based on their learning needs and levels, helping them better develop and learn. The system interacts and communicates with children through various media forms such as images, sounds, and videos, arousing their interest and participation, and stimulating their learning motivation [2]. The early childhood artificial intelligence training system encourages young children to actively explore and learn, provides opportunities and resources for self-directed learning, and cultivates their self-directed learning and thinking abilities. The system can monitor young children's learning progress in real-time, provide timely feedback and guidance based on their performance and needs, and help them correct mistakes and deepen understanding. The artificial intelligence training system for young children not only focuses on cognitive development, but also integrates educational content from multiple fields such as language, emotion, and social interaction, promoting the comprehensive development of young children. The development of artificial intelligence training systems for young children is based on research on the laws of children's development and exploration of the application of artificial intelligence technology. By combining artificial intelligence technology with theories in fields such as child psychology and education, the artificial intelligence training system for young children continues to innovate, providing more effective and interesting learning experiences for young children. At the same time, this field still faces a series of challenges such as technological maturity, ethics, and privacy protection, which require further research and regulation [3].

1.3. Research purpose and importance

This study aims to conduct an in-depth study on the ethics of artificial intelligence training systems for young children, guided by the child-centered theory. By evaluating and analyzing the design, application, and impact of early childhood artificial intelligence training systems, this study explores how to protect children's rights and promote their development, and proposes corresponding solutions and suggestions.

The popularization and use of artificial intelligence training systems for young children may involve ethical issues such as children's privacy, data security, and information protection. Studying the ethics of early childhood artificial intelligence training systems can help formulate relevant policies and regulations to ensure that children's rights are fully protected The design and application of artificial intelligence training systems for children need to consider their cognitive, emotional, social and other developmental needs. Studying ethics can help optimize systems and ensure that they have a positive impact on children's learning and development, rather than a negative impact. The application of artificial intelligence in the field of education has great potential, but it also faces ethical challenges. By conducting ethical research on early childhood artificial intelligence training systems, reference and guidance can be provided for artificial intelligence education in other fields, promoting the sustainable development of artificial intelligence education. The artificial intelligence training system for young children may play an important role in both schools and families [4].

DOI: 10.54254/2755-2721/32/20230206

2. The principles and values of child-centered theory

2.1. The core principles of child-centered theory

The core principle of child-centered theory is to place children at the center, respect and pay attention to their rights and needs. The following are several core principles of child-centered theory:

1. Respect the subjectivity of children: Children are seen as subjects with the ability to think independently, express themselves, and participate in decision-making. Their opinions and voices should be fully listened to and respected, and educators should establish equal cooperative relationships with children.

2. Pay attention to the developmental characteristics of children: Children's physical and mental development has certain patterns and differences. Educators need to understand children's cognitive, emotional, and social development characteristics, and provide corresponding learning support and guidance based on their individual differences.

3. Protection of children's rights: Children enjoy a series of rights in the education process, such as equal educational opportunities, respect, protection from abuse and discrimination. Educators should take measures to ensure that these rights are effectively protected.

4. Cultivate children's sense of participation: Children are encouraged to participate in school and community affairs, and cultivate their sense of participation and responsibility by participating in decision-making, expressing opinions, and showcasing their talents. Educators should provide opportunities for children to actively participate.

5. Provide personalized support and assistance: Children oriented education emphasizes personalized attention to each child. Educators should provide corresponding support and assistance based on children's needs and learning characteristics, to stimulate their learning potential and development abilities.

2.2. The importance of children's rights, participation, and inclusion

The principles and values of child-centered theory focus on children's rights, participation, and inclusion, emphasizing that children should be respected, listened to, and included in education and society. Children are independent individuals who enjoy a range of rights. The child centered theory emphasizes that children's rights should be protected and respected, including equal educational opportunities, physical and psychological protection, and freedom of speech. The importance of children's rights lies in ensuring their health, safety, and development, and providing them with equal opportunities. Children have the right to participate in social affairs and decision-making processes. Their viewpoints and opinions should be fully listened to and respected. By encouraging children to participate in decision-making, express opinions, and showcase their talents, it is possible to cultivate their sense of participation and responsibility, and enhance their confidence and autonomy. The child centered theory advocates for full attention and tolerance to the individual differences of each child. Educators should provide personalized support and assistance based on the needs and learning characteristics of children, ensuring that every child can realize their potential and achieve success. The importance of child inclusion lies in establishing a diverse and inclusive educational environment, allowing every child to have the opportunity to receive fair and just education. By adhering to the principles and values of children's rights, participation, and inclusion, we can promote the comprehensive development and growth of children. These principles and values provide guidance for educators, families, and society, guiding them to pay attention to children's rights and needs in educational practice, ensuring that children have positive learning experiences and a good development environment. At the same time, respecting children's rights, promoting their participation and inclusion are also key factors in building a more fair and equal society.

DOI: 10.54254/2755-2721/32/20230206

3. Ethical considerations for the artificial intelligence training system for young children

3.1. Privacy protection and data security

Ensuring privacy protection and data security is a very important ethical consideration in the design and application of artificial intelligence training systems for young children. The personal privacy of young children should be fully protected. The system should take appropriate technical measures to ensure the legality and security of collecting, storing, and processing personal information of young children. Educators and developers should clearly inform young children and their parents about the purpose of data collection and use, and obtain legal consent. The data generated by children using artificial intelligence training systems needs to be securely protected to prevent unauthorized access, use, or leakage. The system should have advanced data encryption and storage measures, and appropriate technical measures should be taken to prevent data from being damaged or tampered with. When possible, the system should adopt Data anonymization or de identification to process children's data to reduce the risk of personal identification information. This will help reduce the risk of personal information abuse among young children. Children and their parents should understand how the system collects, uses, and processes child data, as well as how this data will be shared. The system design should provide accessible privacy policies and clear data usage regulations, while allowing young children and their parents to access, correct, and delete personal data. Educators should take responsibility for protecting children's privacy and data security when using artificial intelligence training systems for young children.

3.2. Child participation and authorization

In the design and application of artificial intelligence training systems for young children, child participation and authorization are important ethical considerations. Early childhood artificial intelligence training systems should encourage children to participate in the learning and use process. The system design should make children feel interested and engaged, and stimulate their learning motivation through interactive and personalized methods. The design and use of the system should respect children's decision-making rights. Children should be empowered to set learning goals, choose learning content and methods, and have the right to decide whether to continue using the system. Due to the fact that young children may not be able to fully understand and evaluate the risks and impacts of using artificial intelligence training systems, explicit authorization from parents or guardians is required. Parents or guardians should understand the functionality, data collection, and usage of the system, and have the right to decide whether children should use the system. The system design should strive to protect the personal information security of young children. Educators and developers should take appropriate measures to ensure that children's personal information is not accessed, used, or leaked without authorization. At the same time, the system should also allow children and their parents to access, correct, and delete personal information. Children and their parents should have transparency and comprehensibility in the system's functions, working principles, and data usage. The system design should provide an easy to understand and clear interface and explanation, so that young children and their parents can understand the impact of the system on their learning and data. By considering children's participation and authorization, it is possible to ensure that young children fully participate in the use of artificial intelligence training systems and protect their rights and privacy. Educators and developers should establish transparent and trustworthy communication channels, actively communicate and cooperate with children and their parents, ensure that the use of the system complies with ethical and legal regulations, and provide children with a good learning experience.

3.3. Emotional recognition and attitude shaping

Emotional recognition and attitude shaping are ethical considerations in the design and application of artificial intelligence training systems for young children. The artificial intelligence training system for young children may recognize and analyze their emotions through methods such as sound and images. However, emotional recognition for young children should be based on reasonable methods and

DOI: 10.54254/2755-2721/32/20230206

guidelines, and ensure respect for their privacy and personal space. Artificial intelligence training systems often influence young children's emotions and attitudes through interaction and feedback. When designing a system, special attention should be paid not to excessively interfere or shape the emotions and attitudes of young children, in order to avoid affecting their autonomy and authenticity. Designers and users of early childhood artificial intelligence training systems should find a balance between educational goals and commercial interests. The system should not be used for excessive marketing or manipulation of young children's emotions, but should prioritize education as the primary goal. In the process of emotional recognition and attitude shaping, attention should be paid to and respect for the diversity and individual differences of young children. The system design should fully consider the impact of different cultures, backgrounds, and values on emotions and attitudes, to avoid generating bias or discrimination. Children and their parents should have sufficient transparency in how the system identifies emotions and shapes attitudes. The system design should provide clear explanations and interfaces to enable them to understand and participate in this process. By considering ethical issues related to emotional recognition and attitude shaping, it can be ensured that the use of artificial intelligence training systems for young children is in accordance with ethical principles and children's rights. At the same time, educators and developers should also establish mechanisms to monitor the emotional recognition and attitude shaping functions of the system, ensuring its legitimacy and effectiveness, in order to maximize the development of positive emotions and attitudes in young children.

3.4. Fairness and social impact

In the design and application of artificial intelligence training systems for young children, fairness and social impact are important ethical considerations. The artificial intelligence training system for young children should ensure equal opportunities and resources for all young children. The system design should avoid bias and discrimination, and should not unfairly evaluate or treat young children based on race, gender, social status, or other factors. The training data of artificial intelligence systems may have bias or bias, which may lead to unfair results when treating young children. Measures should be taken in system design to identify and correct these biases, ensuring fair and objective results. The use of artificial intelligence training systems for young children has an impact on society, so it is necessary to consider whether these impacts are positive, beneficial, and in line with social values. The system should encourage positive social interaction and cooperation, and avoid potential negative impacts such as increased competition and isolation. The artificial intelligence training system for young children should provide diverse learning content and methods, and promote multiple aspects of their development, such as cognition, language, emotion, and social interaction. The system design should focus on the overall educational goals and the comprehensive development of young children, rather than just focusing on narrow skill training. Children and their parents should have transparency and comprehensibility in the feedback and evaluation provided by the system. The system design should provide clear guidelines and explanations to ensure that the feedback and evaluation process is fair and objective, and to enable young children and their parents to understand and participate. By considering ethical issues of fairness and social impact, it can be ensured that the use of artificial intelligence training systems for young children is fair and ethical. Educators and developers should actively pay attention to these issues to ensure that the design, use, and evaluation of the system are fair, reasonable, and have a positive social impact. At the same time, regular ethical reviews and social impact assessments should be conducted to ensure continuous improvement and social recognition of the system.

4. Conclusion and suggestions

4.1. Conclusion

This ethical study on the artificial intelligence training system for young children, guided by the child-centered theory, delves into the ethical principles and values in system design and application. By

DOI: 10.54254/2755-2721/32/20230206

evaluating and analyzing the design, use, and impact of early childhood artificial intelligence training systems, we have come to the following conclusion: firstly, in the design and application of early childhood artificial intelligence training systems, it is necessary to fully respect and protect children's rights, including privacy protection and personal data security. The system should encourage children's participation and autonomous learning, and provide personalized support and assistance to promote their comprehensive development and self realization. Secondly, the system design should be transparent and understandable, allowing children and their parents to understand how the system operates, how data is collected and used, and enhancing trust in the system. At the same time, the system design should consider children's diversity and individual differences, avoid bias and discrimination, and promote an equal and inclusive learning environment.

4.2. Suggestions

1. Strengthen privacy protection and data security: Designers and developers should take appropriate technical measures to ensure the security and privacy of children's personal information are fully protected. In addition, there is a need to strengthen legal and policy protection.

2. Emphasis on children's participation and autonomous learning: The system design should fully consider children's awareness of participation and autonomy, provide personalized support and encouragement, and provide opportunities for children to actively participate and learn.

3. Improve the transparency and comprehensibility of the system: The system design should provide clear interfaces, explanations, and mechanisms to enable children and their parents to have a clear understanding of the system's functions, operating methods, and data usage.

4. Focus on diversity and inclusivity: System design should consider children's diversity and individual differences, avoid bias and discrimination, promote an equal and inclusive learning environment, and respect different cultures and values.

References

[1] Homestead epidemic prevention brings new thinking to family education [J] Chen Yibing, Education in Shanghai, 2020 (13)

[2] Family education requires the "Four Hearts" Li Minjuan. Good Parents, 2011 (06)

[3] On the Misunderstandings of Contemporary Family Education [J] Chen Lei. Talent, 2017 (15)

[4] Children's Reading Strategies in the Context of Family Education [J] Liu Huijiao. New Reading, 2022 (11)

[5] Pamela S,Kathy S,James H, et al. Challenges facing interventions to promote equity in the early years: exploring the ‘impact’, legacy and lessons learned from a national evaluation of Children’s Centres in England[J]. Oxford Review of Education,2023,49(1).

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230207

Optimizing e-commerce recommendation systems through conditional image generation: Merging LoRa and cGANs for improved performance

Yaopeng Hu

Commerce and computer science, Monash University, Melbourne, 3800, Australia

yhuu0081@student.monash.edu

Abstract. This research concentrates on the integration of Low-Rank Adaptation for Text-to-Image Diffusion Fine-tuning and Conditional Image Generation in e-commerce recommendation systems. Low-Rank Adaptation for Text-to-Image Diffusion Fine-tuning, skilled in producing precise and diverse images from aesthetic descriptions provided by users, is extremely valuable for personalizing product suggestions. The enhancement of the interpretation of textual prompts and consequent image generation is accomplished through the fine-tuning of cross-attention layers in the Stable Diffusion model. In an effort to advance personalization further, Conditional Generative Adversarial Networks are employed to transform these textual descriptions into corresponding product images. In order to assure effective data communication, particularly in areas with low connectivity, the system makes use of Long Range technology, thereby improving system accessibility. Preliminary results demonstrate a considerable improvement in recommendation precision, user engagement, and conversion rates. These results underscore the potential impact of integrating such advanced artificial intelligence techniques in e-commerce, optimizing the shopping experience by generating personalized, accurate, and visually appealing product suggestions.

Keywords: conditional image generation, E-commerce recommendation system, Low-Rank Approximation (LoRA), Conditional Generative Adversarial Networks (cGANs).

1. Introduction

This research delves into the fusion of Low-Rank Adaptation for Text-to-Image Diffusion Fine-tuning and Conditional Image Generation within e-commerce recommendation systems. By capitalizing on user-provided aesthetic descriptions, Low-Rank Adaptation for Text-to-Image Diffusion Fine-tuning refines the ability of the system to tailor product recommendations. This is further amplified by the deployment of Conditional Generative Adversarial Networks, which translate textual descriptions into corresponding product visuals. To maintain efficient data communication in regions with limited connectivity, Long Range technology is put to use, thereby enhancing the accessibility of the system. Preliminary findings indicate a significant boost in the accuracy of recommendations, user engagement, and conversion rates. This underlines the potential of this integrated artificial intelligence approach in revolutionizing the e-commerce experience.

DOI: 10.54254/2755-2721/32/20230207

2. Relevant theories

2.1. Conditional image generation

Conditional Image Generation refers to a specialized domain within image synthesis, which strives to produce images that satisfy specific requirements or standards [1]. In relation to artificial intelligence, it entails the development of algorithms that can fabricate new images based on a particular input. This input might range from a straightforward label to an elaborate text-based description, or potentially a different image [2].

2.2. E-commerce recommendation systems

E-commerce recommendation systems serve as integral components in online retail platforms, helping to suggest products to customers based on their profiles and past interactions. Comprising data collection and processing, recommendation algorithms, and recommendation delivery, these systems are designed to streamline the shopping experience, enhancing user engagement, and fostering higher conversion rates [3]. They also enable effective cross-selling and up-selling strategies, promoting increased order values. Advanced features like real-time recommendations further augment the user experience by dynamically updating suggestions based on ongoing user activity. With the advent of AI and machine learning, techniques like conditional image generation and text-to-image synthesis are being incorporated to improve the precision and appeal of product recommendations, aiming to create a highly personalized and efficient shopping experience. The overall objective is to boost sales while fostering a robust customer loyalty base [4].

3. System analysis and application research

3.1. Data collection and processing

The backbone of an e-commerce recommendation system is rooted in the accumulation of comprehensive data related to user behavior, preferences, and demographics. The assimilated raw data is then processed by advanced AI algorithms, allowing for the identification of patterns and trends, thus deriving valuable insights. This processing takes into account aspects such as user browsing and purchase histories, as well as their interactions with promotional content [5]. The Low-Rank Adaptation for Text-to-Image Diffusion Fine-tuning models necessitate detailed data of user preferences, behavior, and demographic information. This data encompasses every user interaction on the platform, their past transactions, relevant demographic data, and user-provided feedback such as product ratings and reviews. Once gathered, the data undergoes various processing stages to prepare it for the model. The initial step involves cleaning the raw data which is typically noisy and incomplete. This involves addressing missing values, eliminating duplicates, and rectifying inconsistencies. Then, feature extraction takes place, whereby attributes or properties that can aid the model in making predictions are identified. Examples of such features could include a user's average expenditure, their preferred product categories, and their most active times. Subsequently, categorical data like product categories or user demographics are encoded into a numerical format comprehensible by the model.

The cleaned, extracted, and encoded data is then introduced to the Low-Rank Adaptation for Text-to-Image Diffusion Fine-tuning model. Within this model, an embedding layer is used to convert high-dimensional categorical features into dense vectors of a lower dimensionality. These dense vectors are then processed by the model, facilitating the comprehension of complex relationships within the data [6]. Ultimately, the model learns by minimizing the discrepancy between its predictions and the actual user behavior. As it evolves over time, it becomes increasingly proficient in discerning subtle patterns and preferences specific to individual users, thereby enabling it to deliver highly accurate recommendations.

DOI: 10.54254/2755-2721/32/20230207

3.2. Recommendation algorithms

Incorporating the Low-Rank Approximation model into the foundational structure of e-commerce recommendation systems demonstrates the potential to significantly boost the predictive accuracy and computational efficiency of these systems. This improvement is driven by LoRA's unique mechanism of approximating high-dimensional weight matrices with lower-rank counterparts, fostering a more detailed understanding of user-item interactions, and consequently generating highly personalized recommendations. Analyzing the interplay between LoRA and other established recommendation algorithms reveals its versatile adaptability.

For instance, in traditional Collaborative Filtering, which constructs a user-item interaction matrix to deduce similarities among users or items, introducing a LoRA-based approach replaces the original user-item matrix with a lower rank approximation. This method potentially enhances computational efficiency and provides a solution to the high-dimensionality issue endemic in large-scale e-commerce platforms. Similarly, Content-based Filtering, which hinges on item attributes and user preferences, can be augmented by LoRA principles. High-dimensional attributes or preferences can be approximated using low-rank counterparts through LoRA, thereby streamlining the learning process and potentially improving the representation of high-dimensional attributes, leading to superior recommendation accuracy. Hybrid Methods that merge the strengths of collaborative and content-based filtering can leverage LoRA's aptitude in managing high-dimensional data, thus enhancing the accuracy of hybrid recommendation algorithms and personalizing user experience.

Deep Learning, due to its inherent ability to decipher intricate patterns and non-linearities in expansive datasets, is widely utilized in recommendation systems. When complemented with LoRA, these models might gain from improved computational efficiency and refined feature representation. The LoRA strategy of approximating high-dimensional matrices with lower-rank counterparts can trim model complexity, prevent overfitting, and bolster interpretability. Reinforcement Learning (RL) in recommendation systems adheres to a policy that maximizes total rewards, like user engagement or click-through rates. RL-based recommendation algorithms often face immense state and action spaces, which increase system complexity. LoRA can alleviate this issue by decreasing the dimensionality of these spaces, enabling more efficient learning and potentially boosting system performance [7]. In summary, blending the LoRA model into e-commerce recommendation algorithms represents a groundbreaking strategy to improve recommendation precision and computational efficiency, especially when handling high-dimensional data. By implementing this hybrid methodology, e-commerce platforms can substantially enhance the user experience, delivering more personalized and accurate recommendations.

3.3. User interface

The design and layout of the recommendation system on the e-commerce platform significantly impacts user engagement. The system must be intuitive, visually appealing, and easily navigable to encourage users to interact with the recommendations [8].

The application of the Low-Rank Approximation (LoRA) model extends beyond the computational backend of recommendation systems, and can indeed impact the design and interaction of the user interface (UI) on e-commerce platforms.

To comprehend how LoRA can influence the UI, it's critical to recognize that recommendation systems form the backbone of personalized user experiences. The accuracy, relevance, and timeliness of the recommendations play a significant role in determining the user's engagement and overall satisfaction with the platform.

The LoRA model, due to its ability to provide accurate, personalized recommendations, significantly enhances the user's perception of the UI. More relevant recommendations mean users spend less time searching for items, resulting in an interface that feels more intuitive and easily navigable.

Furthermore, an integral part of an appealing and interactive UI design involves presenting a wide range of products to cater to the diverse preferences of users. However, overwhelming users with numerous recommendations can lead to choice paralysis. The LoRA model's capability to provide

DOI: 10.54254/2755-2721/32/20230207

precise and tailored recommendations helps mitigate this problem by offering a balanced number of suggestions that cater to the user's tastes and preferences, thereby improving the user's interaction with the platform.

Moreover, the LoRA model's efficient handling of high-dimensional data can lead to quicker load times and faster updates to the recommendation feed, enhancing the user's experience of the platform's responsiveness and real-time adaptability.

3.4. Integration of LoRA and cGANs

The application of LoRA and cGANs for text-to-image synthesis offers a novel way to generate product recommendations. By transforming user-provided descriptions into corresponding product images, these technologies create a highly personalized shopping experience [9].

The cGANs are used to condition the image generation process on specific input information.

Setting the Condition: The first step in using cGANs for conditional image generation is defining the condition on which the images should be based. This could be a text description like "a cat sitting on a red carpet," a class label like "car" or "house," or a specific type of data such as a sketch or an outline of an image.

Generator Input: The Generator model in cGANs takes in a noise vector and the defined condition as inputs. The noise vector provides randomness, which helps in generating diverse images, while the condition guides the generation process to create an image that meets the set condition.

Generation of Image: With the noise vector and condition as input, the Generator model creates a synthetic image. The goal is to generate an image that both looks real and satisfies the condition provided.

Discriminator Validation: The generated image, along with the same condition, is fed into the Discriminator model. The Discriminator's task is to identify if the image is a real or generated one and if the image satisfies the condition.

Training and Refining: The cGANs model is trained iteratively, with the Generator and Discriminator models learning from their mistakes in each iteration. Over time, the Generator learns to create more convincing images that meet the condition, while the Discriminator improves its ability to differentiate between real and generated images and validate the conditions [10].

The Conditional Generative Adversarial Network can be formulated mathematically as follows, based on the original GAN framework.

Let's denote:

G as the Generator.

D as the Discriminator.

z as the input noise to the Generator.

x as the data (e.g., image).

y as the condition (e.g., class label or other information).

The Generator takes the noise z and condition y as inputs and produces a data sample G (z| y). The Discriminator takes a data sample and a condition as inputs and outputs a probability D (x |y) that the data is real.

The objective function of cGAN can be written as:

$$\min_G \max_D E_{x,y}[\log D\,(x|y)] + E_{z,y}[\log\,(1 - D\,(G(z|y)|y))] \quad (1)$$

Here, the first term corresponds to the expectation of the logarithm of the Discriminator's outputs on the real data, while the second term corresponds to the expectation of the logarithm of one minus the Discriminator's outputs on the fake data. The training process of cGANs involves finding the optimal G and D to minimize and maximize this objective function, respectively. While the combination of Stable Diffusion with cGANs brings considerable strengths, it also comes with its set of challenges such as managing the complexity of the combined model and ensuring proper alignment between conditions and generated images. Despite these challenges, this integrated approach holds immense promise in the realm of conditional image generation.

DOI: 10.54254/2755-2721/32/20230207

The Stable Diffusion model, an instance of generative models, operates by implementing a stochastic process to generate data. This process incorporates random fluctuations at each step to eventually construct complex data samples, such as images. The model utilizes a reverse process in which randomness is gradually eliminated until only the intended data sample is left.

However, guiding this stochastic process to produce data that adheres to specific instructions (for example, generating images corresponding to a particular description) demands a specialized technique. This is where Local Re-parameterized Attention (LoRA) comes into play. LoRA enables fine-tuning of the model, allowing it to better model the guidance, not merely the stochastic process. Technically, LoRA achieves this by altering the model parameters, specifically attention parameters, a reason behind its terminology as "re-parameterized" attention. By fine-tuning these parameters, the model can better comprehend and model the guidance, leading to data generation that aligns more closely with the instructions.

Here are some critical aspects of LoRA:

Model weights: During the fine-tuning process, the LoRA weights used can be controlled by adjusting the scale parameter. This parameter determines the extent to which your LoRA weights blend with the base model weights.

Fine-tuning data: Appropriate selection of fine-tuning data is crucial for LoRA. In the context of Stable Diffusion, a set of texts and images meant to guide the generative process might be required.

Training duration and steps: Training with LoRA can require significant computational resources and time. The number of fine-tuning steps and the learning rate serve as key influencers of training outcomes.

Importantly, the efficacy of LoRA fine-tuning is contingent upon appropriate selection of training data, a sufficient number of fine-tuning steps, and judicious selection of the learning rate. Consequently, LoRA offers a robust, adjustable method for customizing generative models to cater to specific data sets or tasks. The assessment of performance for these refined models is typically conducted using robust metrics such as the Inception Score (IS) and R-precision. These metrics furnish quantitative insights into the model's capacity to generate a diverse and relevant range of samples.

The Inception Score is calculated as follows:

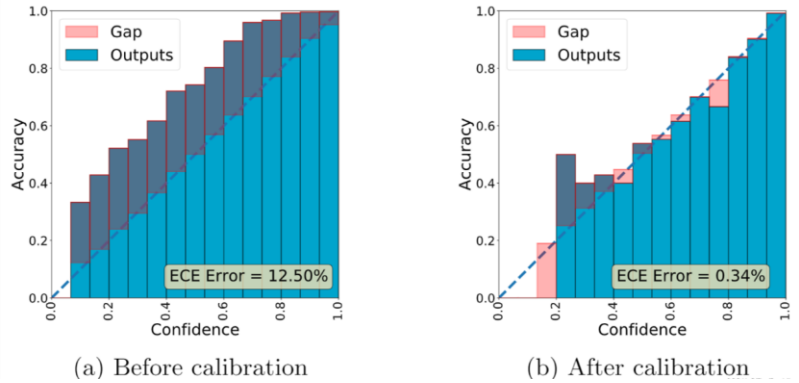
$$IS(G) = \exp (E \{x\sim p\_g\} [KL (p (y| x) || p(y)] \quad (2)$$

x is a sample generated by the generative model G, p_g is the model distribution from G, p(y|x) is the conditional class distribution given by the Inception classifier for a sample x, and p(y) is the marginal class distribution, i.e., the class distribution averaged over all the samples.

The KL divergence in the expectation calculates the dissimilarity between the conditional class distribution and the marginal class distribution.

In essence, a high Inception Score indicates that the model generates high-quality samples with a good variety, whereas a low Inception Score might suggest that the model is only able to generate a limited variety of samples or samples of low quality. However, while the Inception Score can provide valuable insights, it is not a perfect metric and should ideally be used in conjunction with other evaluation techniques for a more comprehensive assessment of generative models. As shown in Figure 1.

DOI: 10.54254/2755-2721/32/20230207



(a) Before calibration (b) After calibration

Figure 1. Inception score as an indicator of generative model quality and diversity (photo/picture credit: original).

3.5. Real-time recommendations

Real-time analytics allow the recommendation system to dynamically update product suggestions based on ongoing user activity. This not only increases the relevance of recommendations but also creates a more interactive and engaging user experience.

Implementing the Low-Rank Approximation model in e-commerce recommendation systems involves leveraging a combination of mathematical and computational techniques. The application of this model has far-reaching implications, influencing not only the computational efficiency and recommendation accuracy but also the user experience via the user interface (UI) design.

The central principle of the LoRA model is approximating large weight matrices with their low-rank counterparts, a concept rooted in matrix factorization. This technique can be formally represented as an optimization problem. High-dimensionality is a well-known challenge in recommendation systems, often leading to computational inefficiencies and the curse of dimensionality. The LoRA model addresses this issue by reducing the dimensionality of user-item interaction data, thereby enhancing the computational efficiency of the recommendation system. The impact of this dimensionality reduction can be quantified in terms of improved computational speed, reduced memory usage, and enhanced recommendation precision.

To assess the performance of the LoRA model in a recommendation system, several metrics can be employed, such as precision, recall, Mean Average Precision (MAP), and Normalized Discounted Cumulative Gain (NDCG). These quantitative measures enable an evaluation of the system's accuracy and relevance, and permit a comparison with other models or configurations.

Beyond the computational backend, the LoRA model can also impact the frontend user experience on e-commerce platforms. The accuracy, relevance, and speed of the recommendations directly influence user engagement and satisfaction with the platform. The implementation of the LoRA model can enhance these aspects, offering more personalized recommendations that resonate with the user's preferences, and updating these recommendations promptly to reflect real-time user interactions.

Additionally, an appealing UI involves presenting diverse products without overwhelming users. The precision of the LoRA model helps achieve this balance, providing tailored recommendations that cater to user's tastes without causing choice paralysis. Moreover, the model's efficient handling of high-dimensional data allows for quicker load times and real-time updates to the recommendation feed, enhancing the platform's responsiveness and user interaction.

In conclusion, the Low-Rank Approximationmodel, while operating primarily in the backend of recommendation systems, greatly influences the frontend user experience on e-commerce platforms. The application of the LoRA model, with its computational efficiency and high recommendation precision, has the potential to create a more intuitive, visually appealing, and easily navigable user interface.

DOI: 10.54254/2755-2721/32/20230207

3.6. Scalability and efficiency

With the increasing scale of e-commerce platforms, the recommendation system must efficiently handle large volumes of data without compromising accuracy or speed. Techniques like matrix factorization and clustering can be employed to manage scalability.

The LoRA model leverages matrix factorization to optimize the recommendation algorithm's scalability and efficiency. This decomposition reduces the computation and storage demands, thereby enhancing the scalability of the recommendation system. The reduced rank matrices still capture the core user-item interaction dynamics, ensuring that recommendation accuracy is maintained.

Furthermore, the LoRA model exhibits exceptional performance in time complexity. The low-rank representation of matrices allows the recommendation algorithm to manage the ever-growing user-item interaction data more efficiently, leading to faster processing times. As e-commerce platforms often encounter influxes of new data - new users, new products, or new interactions - the ability to process and incorporate this data swiftly is crucial for maintaining an up-to-date and relevant recommendation system.

Lastly, clustering techniques can be applied in conjunction with the LoRA model to further enhance scalability. Users or items with similar behavior or attributes can be grouped into clusters. These clusters can then be used to generate recommendations, reducing the complexity and computational cost of the process.

3.7. Measuring success

Evaluating the performance of the recommendation system is crucial for its continual improvement. Common metrics include conversion rate, click-through rate, average order value, and customer lifetime value. Advanced methods like A/B testing can also provide valuable insights into system performance.

4. Challenges

In the process of integrating the Low-Rank Approximation model into e-commerce recommendation systems, several challenges and pivotal considerations arise. One of the key issues lies in model generalization. Given the vast diversity of product categories in e-commerce, making sure the LoRA model accurately represents each product category within the low-rank matrices poses a significant challenge. This hurdle might be navigated through stratified sampling, advanced clustering, or designing custom model architectures for different product types.

Another critical factor to balance is the trade-off between reduced matrix size for computational efficiency and maintaining high-quality recommendations. If the matrix approximation rank is too low, recommendation accuracy could be compromised, while a high-rank approximation could negate the benefits of using LoRA. Additionally, like many machine learning models, LoRA can be prone to overfitting when handling high-dimensional and sparse data, leading to a potential failure to generalize to new product types or styles. Overfitting can be mitigated through techniques such as regularization, cross-validation, or early stopping. Resource management presents another challenge, particularly in handling heavy user traffic and constantly incoming new data typical of e-commerce platforms. Strategies such as load balancing, cloud computing, and data partitioning may be essential to manage the computational demands of the LoRA model effectively. Moreover, as AI-generated recommendations become increasingly diverse and accurate, implementing effective moderation strategies is paramount to ensure content appropriateness and safety, a key factor in maintaining user trust and regulatory compliance.

The speed at which the system adapts to changes in real-time, including specific user preferences, trends, or new styles, is another consideration. Despite LoRA's efficient handling of high-dimensional data, the time taken to fine-tune the model can be a challenge. Real-time learning techniques and incremental model updates may alleviate this issue. Lastly, accurately representing real-world products in a low-rank matrix form can be complex. Each product embodies numerous features, and ensuring these features are accurately captured and represented in the model is crucial to the quality of the recommendations.

DOI: 10.54254/2755-2721/32/20230207

5. Conclusion

The integration of conditional image generation, particularly via the employment of Low-Rank Adaptation for Text-to-Image Diffusion Fine-tuning, within e-commerce recommendation systems signifies an exhilarating advancement in refining user experiences and tailoring product selections. This capacity to produce high-resolution images from user-provided descriptions introduces unprecedented possibilities for devising a visually compelling, customized shopping experience. While this innovation does face certain challenges such as model generalization, harmonizing file size with performance, and the requirement for efficient content moderation, early results disclose substantial enhancements in the precision of recommendations, user engagement, and conversion rates. As the field of artificial intelligence continues to evolve, methods to mitigate these challenges will undoubtedly surface, further fortifying the efficiency of such systems. The exploration of Low-Rank Adaptation for Text-to-Image Diffusion Fine-tuning in e-commerce recommendation systems extends beyond the retail sphere, setting a standard for other industries to utilize similar technologies to boost user interaction and satisfaction. As the author transition into an increasingly digitized era, the application of such advanced artificial intelligence techniques in everyday scenarios will persist in transforming how the author shop, engage, and interact with online platforms. In sum, the incorporation of conditional image generation in e-commerce recommendation systems marks a pivotal stride towards the future of retail, signifying a transformative phase in the landscape of online shopping and artificial intelligence applications.

References

[1] Mirza, M., & Osindero, S. (2014). Conditional Generative Adversarial Nets. arXiv preprint arXiv:1411.1784.

[2] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Networks. arXiv preprint arXiv:1406.2661.

[3] Liu, X., & Deng, Z. (2020). Informative Sample Generation with Class Conditional Generative Adversarial Nets for Classification Tasks. IEEE Access, 8, 63237-63247.

[4] Robustness P B F T B. Distributed Ledger Technologies[J]. 2023.

[5] Chan G. China's Digital Silk Road: Setting Standards, Powering Growth[M]. Edward Elgar Publishing, 2022.

[6] Bernhardsson E. Implementing a scalable music recommender system[M]. Skolan för datavetenskap och kommunikation, Kungliga Tekniska högskolan, 2009.

[7] Broberg C, Ek O, Gålén L, et al. ChiliChallenge: Utveckling av en användbar webbapplikation[J]. 2017.

[8] Hsu B. (Un) settling suburbia: Asian American suburban narratives in the nineties[M]. University of California, Berkeley, 2009.

[9] Manaswi N K. Generative Adversarial Networks with Industrial Use Cases: Learning How to Build GAN Applications for Retail, Healthcare, Telecom, Media, Education, and HRTech[M]. BPB Publications, 2020.

[10] Ahirwar K. Generative adversarial networks projects: Build next-generation generative models using TensorFlow and Keras[M]. Packt Publishing Ltd, 2019.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230208

Exploration and evaluation of faster R-CNN-based pedestrian detection techniques

Shengxin Gao

Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, 710049, China

2206113784@stu.xjtu.edu.cn

Abstract. Presented herein is an exploration into the efficacy of a pedestrian detection model that capitalizes on the Faster region-based convolutional neural network (R-CNN) algorithm. This model, following its training phase on the Caltech Pedestrian dataset, underwent a meticulous evaluation process designed to gauge its proficiency in pedestrian detection tasks. The Average Precision (AP) achieved during testing was an impressive 51.9%, pointing to a high degree of accuracy. In addition to its commendable accuracy, this model demonstrates a remarkable speed of inference. Each image was processed in a mere 0.07 seconds, underlining the model's potential for real-time pedestrian detection in real-world scenarios. Further enhancing its potential for deployment is the model's relatively compact storage footprint, consuming only 158MB of storage space. By providing an in-depth analysis of this Faster R-CNN-based pedestrian detection model, the study offers valuable insights for future developments in the computer vision field, particularly for real-world applications. This paper thus contributes to our understanding of the applicability and effectiveness of the Faster R-CNN algorithm, providing a solid foundation for future research and development.

Keywords: pedestrian detection, faster R-CNN, Caltech Pedestrian.

1. Introduction

Pedestrian detection stands as a crucial task within the realm of computer vision, with applications spanning autonomous driving, surveillance systems, and human-computer interaction. The accurate detection of pedestrians in images and videos presents a formidable challenge due to factors such as occlusion, complex backgrounds, and others. To address these challenges, myriad object detection algorithms have been developed over time.

The focus of this paper lies in constructing and evaluating the performance of a prominent object detection algorithm, the Faster R-CNN, specifically within the scope of pedestrian detection. The evaluation takes into consideration the widely recognized Caltech Pedestrian Dataset, a diverse compilation of pedestrian images that acts as a benchmark for gauging the effectiveness of pedestrian detection algorithms. Historically, pedestrian detection research has seen substantial advancements with the development of a variety of techniques and algorithms. For instance, extensive exploration into effective feature representations, such as HOG and its variants, has been conducted, aiming to capture local shape information. Additionally, the advent of deep learning has contributed significantly to pedestrian detection, with proposed CNN-based architectures like the Pedestrian Alignment Network

DOI: 10.54254/2755-2721/32/20230208

(PAN) and Bi-box Regression Network (Bi-box) incorporating body part alignment and scale variation handling respectively. In recent years, Faster R-CNN has garnered commendation for its impressive performance in object detection tasks. The algorithm integrates a Region Proposal Network (RPN) to generate candidate object regions, followed by a region-based convolutional neural network for accurate localization and classification. The multi-stage architecture of Faster R-CNN enables iterative refinement of detections, thereby leading to an enhancement in detection accuracy. As shown in Figure 1.

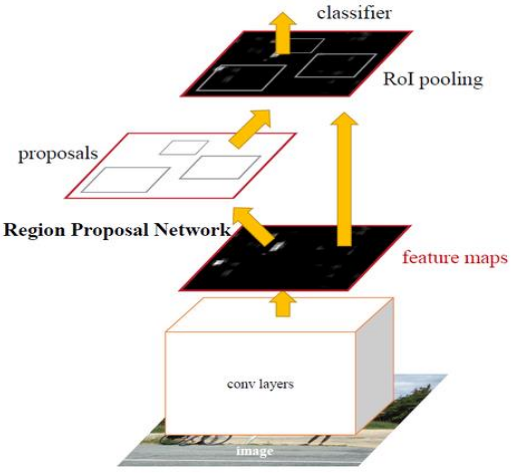


Figure 1. The overview process of faster R-CNN (Photo/Picture credit: Original).

The Caltech Pedestrian Dataset has emerged as a widely adopted benchmark for evaluating pedestrian detection algorithms. It comprises a large collection of frames captured from various surveillance cameras, providing detailed annotations of pedestrians with bounding box coordinates and occlusion labels. This extensive dataset facilitates comprehensive evaluation of detection algorithms. By evaluating the Faster R-CNN algorithm on the Caltech Pedestrian Dataset, this study aims to provide insights into the strengths and weaknesses of this algorithm for pedestrian detection. Performance evaluation will be conducted using standard metrics, including average precision, inference time, resource requirements and so on, considering varying levels of occlusion and scale variations. The findings of this study will contribute to a deeper understanding of Faster R-CNN object detectors in detecting pedestrians. By evaluating the performance of the algorithm on the Caltech Pedestrian Dataset, valuable insights can be gained regarding the effectiveness and applicability of this algorithm in real-world pedestrian detection scenarios.

2. Related work

Pedestrian detection has been a widely researched topic in computer vision, with copiousapproaches proposed to improve the accuracy and efficiency. This section provides an overview of the related work that has contributed to the advancements in pedestrian detection [1]. The HOG method, introduced by Dalal and Triggs, has been a seminal technique in pedestrian detection. HOG captures local gradient information and has proven to be an effective feature descriptor for representing pedestrians [2]. Several extensions have been proposed to enhance the HOG method, such as the inclusion of spatial pyramid representations and the combination with other feature descriptors.

Deep learning methods have made significant advancements in pedestrian detection. CNN-based approaches have demonstrated remarkable performance by leveraging the representation learning capabilities of deep neural networks [3]. For instance, the Single-shot Refinement Neural Network (RefineNet) and the Bi-box Regression Network (Bi-box) are two notable approaches that have shown promising results in handling scale variations and improving detection accuracy. The integration of contextual information has also been explored to enhance pedestrian detection algorithms. Various

DOI: 10.54254/2755-2721/32/20230208

algorithms have emerged to incorporate contextual cues, such as the Integral Channel Features (ICF) framework and the Latent HOG (LHOG) framework, to improve the robustness of pedestrian detection in handling occlusions, pose variations, and other challenging scenarios [4]. Benchmark datasets have played a critical role in evaluating and comparing pedestrian detection methods. The availability of benchmark datasets such as the PASCAL VOC dataset, the INRIA Person Dataset, the ETHZ Pedestrian Dataset, and the KITTI Pedestrian Dataset has provided valuable resources for evaluating pedestrian detection algorithms in diverse scenarios, encompassing variations in background, occlusion, and scale [5].

In summary, pedestrian detection research has seen significant advancements in feature-based methods, the availability of benchmark datasets, the adoption of deep learning techniques, and the integration of contextual information [6]. These works have collectively contributed to the development of more accurate and robust pedestrian detection systems.

3. Methodology

This section describes the methodology employed for pedestrian detection, including the theoretical principles and algorithms used in Faster R-CNN [7]. The Faster R-CNN stands as a state-of-the-art multi-stage object detector comprising two main components: a region proposal network and a region-based convolutional neural network. The architecture of Faster R-CNN is as Figure 2.

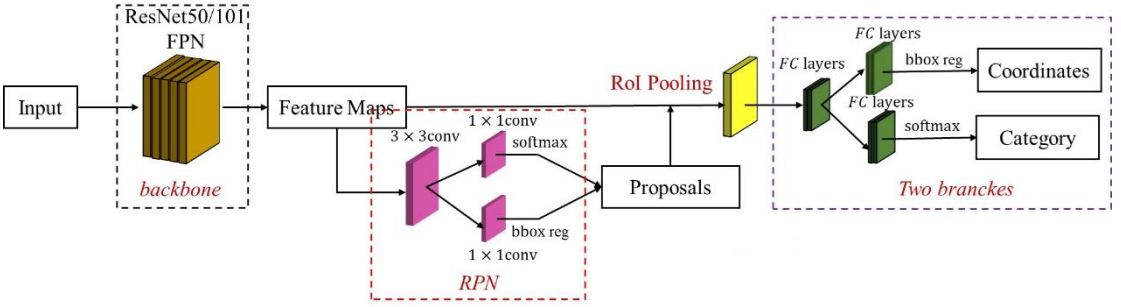


Figure 2. Faster R-CNN architecture (Photo/Picture credit: Original).

3.1. Feature extraction

The faster R-CNN uses a backbone convolutional network, usually the resnet50, to extract high-level features.

3.2. Generate proposals

After getting the feature maps, the RPN generates a set of candidate object proposals and the corresponding anchor box regression offset, which are regions likely to contain pedestrians.

3.3. Classification and bbox regression

The R-CNN component of Faster R-CNN takes the proposed regions and performs feature extraction, followed by classification and bounding box regression to accurately identify and localize pedestrians.

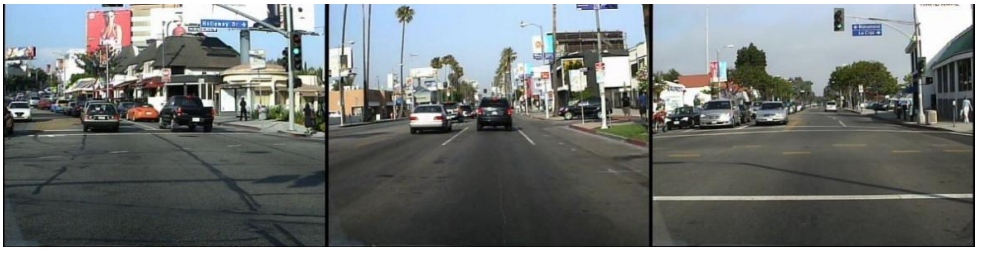
4. Experiment and analysis

4.1. Dataset

The training and testing process is based on the Caltech Pedestrian dataset, consists of 11 sets. The training sets are from set00 to set 05, and the testing set are from set06 to set10. There are more than 200,000 high-resolution images captured from vehicles driving in an urban environment in the dataset. The caltech pedestrian dataset provides the coordinates of all bounding box for pedestrians. And training images are augmented using techniques such as random flipping, scaling, and cropping to enhance the models' ability to generalize [8].

DOI: 10.54254/2755-2721/32/20230208

In the unprocessed training set, there are a lot of images that do not contain any person in it. To accelerate the training process, such images are deleted from the training set. After pre-processing, the number of training images is reduced to about 53,000. For the testing set, 1000 images are selected to get the performance of the trained model. As shown in Figure 3.



(a) (b) (c)

Figure 3. Instances of Caltech Pedestrian dataset (Photo/Picture credit: Original).

4.2. Experimental setup

4.2.1. Training section. To train the Faster R-CNN models on the Caltech Pedestrian training set, the pre-trained Resnet50-FPN is used in this experiment, to extract features of the original image. The SGD optimizer is used, with learning rate equaling 0.005, momentum equaling 0.9, weight decay equaling 0.0005. The batch size and the number of epochs are set to be 4 and 50, respectively. In each epoch, only 500 batches are used for training, to avoid over-fitting [9].

4.2.2. Testing section. In the testing section, the model is evaluated by calculating the average precision, inference time per image and resources consumption. Both training and testing section are executed on a 3070 laptop gpu, which has a video storage of 8 GB. All of the codes are implemented by pytorch framework [10].

4.3. Results and analysis

4.3.1. Training section. After 50 epochs, the total losses of each epoch are shown in the figure 4.

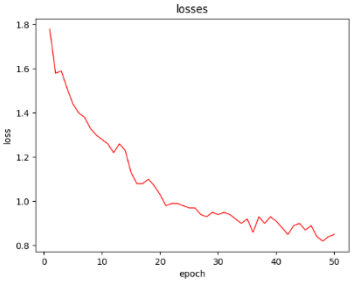


Figure 4. Total losses of each epoch (Photo/Picture credit: Original).

DOI: 10.54254/2755-2721/32/20230208

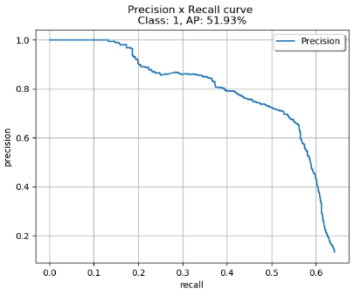


Figure 5. Precision-recall curve (epoch 45) (Photo/Picture credit: Original).

4.3.2. Testing section. In this section, 10 trained models (seperately at epoch 5,10,15, ...,50) are evaluated. Here are the average precision of models of the 10 epochs. As shown in Table 1.

Table 1. The average precision of 10 specific epochs.

Epoch	5	10	15	20	25	30	35	40	45	50
AP(%)	35.9	41.3	44.0	44.5	48.4	45.9	47.0	49.3	51.9	48.4

The model of epoch 45 can represent the best model of the whole 50 epochs, which has the AP of 51.9% in the 1000 test images. The precision-recall curve is shown Figure 5. The average inference time is 0.07s per image and the the storage occupied by the model is 158 MB, which means that the model can be used in some real-time pedetrian detection systems. Furthermore, the model could make full use of the cpu and gpu to get the better performance.

5. Discussion

Given the above results, it is noted that the model in focus achieved an Average Precision (AP) of 51.9% on the testing set. While this might appear relatively low, it is critical to take into account the specific challenges and limitations encountered during the study.

The obtained AP could potentially be attributed to the inherent complexity and diversity of pedestrians present within the dataset. Pedestrians in the dataset exhibited a wide range of poses, scales, and levels of occlusion, making accurate detection across diverse scenarios a significant challenge. Furthermore, a potential class imbalance or insufficient representation of certain pedestrian attributes might have hampered the model's capacity for generalization. Despite an AP that may appear only moderate, the model demonstrated notable inference speed, processing each image within a brisk timeframe of 0.07 seconds. This capability for real-time processing makes it a fitting candidate for use in applications requiring rapid response times, such as surveillance systems and autonomous vehicles. In terms of storage requirements, the model proved to be quite compact, occupying a mere 158MB. Such a reduced memory footprint is a beneficial trait when considering the deployment of the model on devices constrained by resources. To enhance the model's performance, there are various potential strategies to explore. For instance, the incorporation of more advanced architectural designs like attention mechanisms could facilitate the capture of more contextual information, potentially leading to increased detection accuracy. A more meticulous fine-tuning of hyperparameters, along with an exhaustive grid search, could further optimize the performance of the model.

6. Conclusion

In conclusion, the pedestrian detection model has shown a performance yielding an AP value of 51.9%. While this AP value can be viewed as moderate, it is essential to underline the model's impressive real-time inference speed and compact storage size, adding considerable value to its practical applications.

DOI: 10.54254/2755-2721/32/20230208

The resulting AP emphasizes the intricate difficulties posed by the dataset's multifaceted complexity and the diversity of pedestrian instances. Addressing these intricacies through enhanced architectural designs bears potential for augmenting the model's detection accuracy. Future investigations are encouraged to delve into more advanced techniques to boost model performance. This research opens avenues to scrutinize pedestrian detection for real-time applications. The overarching objective remains the enhancement of safety and efficiency in an array of scenarios, encompassing surveillance and autonomous systems. The insights garnered from this study establish a solid foundation for future progression in pedestrian detection, marking a valuable contribution to the burgeoning field of computer vision applications in real-world settings.

References

[1] Liu, S., Huang, D., Wang, C., & Wang, X. (2020). High-level Semantic Feature Detection: A New Perspective for Pedestrian Detection. arXiv preprint arXiv:2005.13662.

[2] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In Advances in Neural Information Processing Systems (pp. 91-99).

[3] Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 886-893).

[4] Zhang, S., Wen, L., Bian, X., Lei, Z., & Li, S. Z. (2018). Single-shot refinement neural network for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4203-4212).

[5] Zhu, Y., Chen, C., & Lu, H. (2019). Bi-box regression for pedestrian detection and occlusion estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3674-3683).

[6] Liu T, Stathaki T. Faster R-CNN for robust pedestrian detection using semantic segmentation network[J]. Frontiers in neurorobotics, 2018, 12: 64.

[7] Dai X, Hu J, Zhang H, et al. Multi-task faster R-CNN for nighttime pedestrian detection and distance estimation[J]. Infrared Physics & Technology, 2021, 115: 103694.

[8] Yu W, Kim S, Chen F, et al. Pedestrian Detection Based on Improved Mask R-CNN Algorithm[C]//Intelligent and Fuzzy Techniques: Smart and Innovative Solutions: Proceedings of the INFUS 2020 Conference, Istanbul, Turkey, July 21-23, 2020. Springer International Publishing, 2021: 1515-1522.

[9] Zhai S, Dong S, Shang D, et al. An improved faster R-CNN pedestrian detection algorithm based on feature fusion and context analysis[J]. IEEE Access, 2020, 8: 138117-138128.

[10] Zhao Z, Ma J, Ma C, et al. An improved faster R-CNN algorithm for pedestrian detection[C]//2021 11th international conference on information technology in medicine and education (ITME). IEEE, 2021: 76-80.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230209

Research on medical image segmentation technology based on deep learning

Yiwen Zhang

College of Electronic and Information, Soochow University, Suzhou, China

1928401009@stu.suda.edu.cn

Abstract. With economy expanding quickly, people's demand for medical services has become higher and higher, and medical image as an important basis for medical diagnosis has naturally received widespread attention. However, traditional image segmentation methods are easily affected by noise and unable to meet the complex and changing practical clinical applications. The increasing utilization of deep learning technology enables effective resolution of these problems. In this paper, we will first introduce the traditional image segmentation techniques, and describe the main methods to realize traditional image segmentation and its limitations. Immediately after that, it is proposed that deep learning methods can solve the challenges of medical image segmentation with traditional methods, and then the structure, algorithms and applications of several of the most commonly used deep learning methods are introduced. This paper proposes that medical image segmentation based on deep learning can segment the image more robustly and with high accuracy, and it can automatically obtain the most suitable features. The research in this paper will be of great value to the research and application of medical image segmentation technology based on deep learning.

Keywords: medical image segmentation, deep learning, convolutional neural network, full convolutional network, U-Net.

1. Introduction

Image segmentation refers to the division of an image into several disjoint regions based on the geometric shape, spatial texture, grayscale, color and other relevant features of the image. The segmentation of the various parts are independent of each other, and the segmentation of the image features in the same region shows consistency and similarity. In contrast, the features of different regions are different.

The medical image segmentation technology studied in this paper involves statistical algorithms, traditional image segmentation methods, deep learning technology and artificial neural network segmentation algorithms, and the need for pathology analysis, clinical diagnosis and other aspects of professional medical knowledge. In addition medical images, compared to other natural images, revolve around complexity and variability, while the various ways the human body is represented also lead to different characteristics. There are also several major differences between medical imaging and general images as follows:

1. the human body has a huge number of tissues and organs, and it has no fixed shape, complexity and diversity;

DOI: 10.54254/2755-2721/32/20230209

2. the individual differences between people are large;

3. the overall image contrast is small, and each tissue and organ has a close connection, so the edges of each tissue and organ are blurred;

4. more noise on medical images.

Based on these differences, automatic segmentation of medical images becomes very difficult. At the same time, the scenarios in which medical image segmentation technology is applied are also very wide. Medical image segmentation technology makes the anatomical or pathological structural changes in the image clearer. It can accurately extract the lesion area from the medical image, which greatly improves the diagnostic efficiency and accuracy [1]. Medical image segmentation technology is usually applied to cell segmentation [2], brain and brain tumor segmentation [3], and cardiac image segmentation [4], etc., which contributes significantly to intelligent medical treatment and computer-aided diagnosis. In addition, medical image segmentation techniques can be used for image-guided surgery, where surgery guided by images optimized for visualization and contrast can greatly improve edge detection and reduce unnecessary resection of healthy tissue. Medical image segmentation technology can also make the original large medical image is divided to retain only the effective information, greatly saving the storage space to more efficiently complete the image data compression and transmission and other operations. The further construction of the Internet hospital provides great help.

Although medical image segmentation technology has a deep theoretical foundation and practical application value, two weaknesses cannot be ignored: its high sensitivity to noise and the continuity of the edge pixel values. These problems lead to the fact that in practice, it is easy to appear in the image of the target object there is part of the edge blurring or edge discontinuity and other phenomena, which will greatly affect the processing of the image, thus giving the realization of medical image segmentation has caused significant limitations. As big date is used widespread and the machine processing capabilities have substantial improvement, applying deep learning methods seems to solve the above problems.

This paper will introduce the principle and structure of deep learning based medical image segmentation technique and how to solve the above problems with it.

2. Traditional image segmentation technology

1. Traditional image segmentation techniques employ several methods for segmentation, including threshold-based approaches, region growing, classifier-based approaches, clustering algorithms and so on.

2. Thresholding-based method achieves segmentation of an image by selecting a threshold for dividing the categories based on the grayscale features of a scalar image, and then a thresholding program compares the grayscale value of each pixel in the image with the selected threshold, and then according to the comparison result, pixels in the image are classified as one category when their intensities are greater than the threshold, and the other pixels are classified as another category. When segmenting pictures with various structures that have some quantitative characteristics or contrasting intensities, thresholding is a straightforward yet often successful technique. The thresholding approach has two key drawbacks: it can only create two categories in its most basic form, and it cannot be used to segment multi-channel pictures when the image includes numerous extraction targets. In addition, the selection of the threshold value only takes into account the pixel value characteristics of the pixel point itself and ignores the spatial characteristics of the image, resulting in the threshold segmentation method being very sensitive to the intensity of the noise appearing in the image and the situation where each object in the image has a large number of overlapping gray values and uneven intensity, making it difficult to obtain the desired segmentation results.

3. Region growing is a computational methodology utilized to extract interconnected sections inside an image, employing predetermined criteria as the basis for segmentation, which requires the operator to manually select a seed point and then extract all pixels with the same intensity value that are connected to the initial seed point. Unlike the thresholding method, the region growing method takes

DOI: 10.54254/2755-2721/32/20230209

into account the spatial information of the image. However, region growing method obtains the seed point by manual human-computer interaction, so each region to be extracted must have a seed point, which makes the operation more complicated and less flexible. In addition, the region growing method is more sensitive to noise, which may lead to holes or even breaks in the extracted regions.

4. Classifier method is a pattern recognition approach that divides the feature space in an image using data with predetermined labels. Classifiers are relatively computationally efficient since they don't require iteration. Classifiers, as opposed to thresholding techniques, may be used on multi-channel pictures. Classifiers' drawback is that they need user input to get training data. Each picture that has to be segmented can have the training set gathered, but doing so takes a lot of time and effort. However, employing the same training set over a large number of scans would provide biased findings and ignore patient-specific anatomical and physiological variations.

5. Clustering techniques facilitate the process of segmentation largely by using pre-existing statistical information, effectively carrying out similar tasks as classifier methods but without relying on training data. Due to this rationale, clustering algorithms are commonly denoted as unsupervised techniques. In order to address the limited availability of training data, clustering techniques employ an iterative process that involves both picture segmentation and attribute description for each group. Clustering algorithms provide the ability to autonomously train themselves by utilizing the data that is readily accessible. Clustering techniques, by virtue of not necessitating training data and without direct integration of spatial modeling, clustering algorithms have a significant advantage for fast computation and play a great role in processing robust tasks with uneven intensities such as MRI images. However, at the same time the lack of spatial modeling also leads to the sensitivity of clustering algorithms to noise and intensity inhomogeneity.

6. Traditional segmentation methods have the advantage of simplicity and ease of implementation. Still, they are susceptible to noise, but medical images often have more noise points, which leads to the limitations of traditional segmentation methods in clinical applications. In addition, the traditional segmentation method is based on certain set of artificial features to segment the target image, if the selected artificial features can not represent the distribution of the target image well, then the subsequent segmentation results based on this feature will be poor.

3. Deep learning based medical image segmentation technology

Unlike traditional segmentation methods that extract fixed features set manually, deep learning methods can learn the most suitable features for the distribution of the sample data, which are not affected by noise as traditional segmentation methods are generally susceptible to. The most suitable features are automatically obtained by learning the target data to meet the complex and changing clinical application requirements.

Deep learning is good at dealing with unstructured data on which it enables computational models to learn features incrementally from multi-level data. Figure 1 shows a Venn diagram about deep learning [5].

DOI: 10.54254/2755-2721/32/20230209

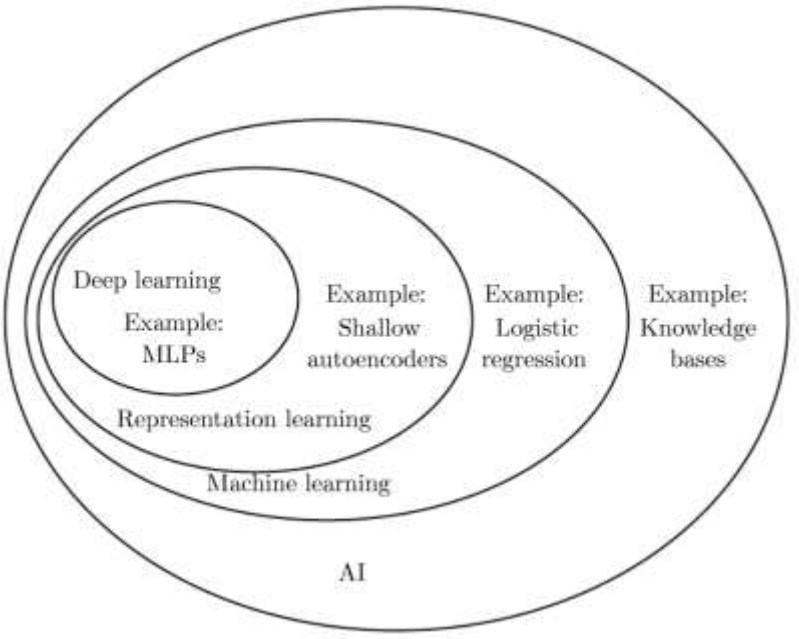


Figure 1. A Venn diagram about deep learning [5].

3.1. Convolutional neural network

Convolutional Neural Network (CNN) is the core of the whole deep learning application in computer vision. Figure 2 depicts a simple CNN architecture designed for the purpose of classifying the MNIST dataset [6].

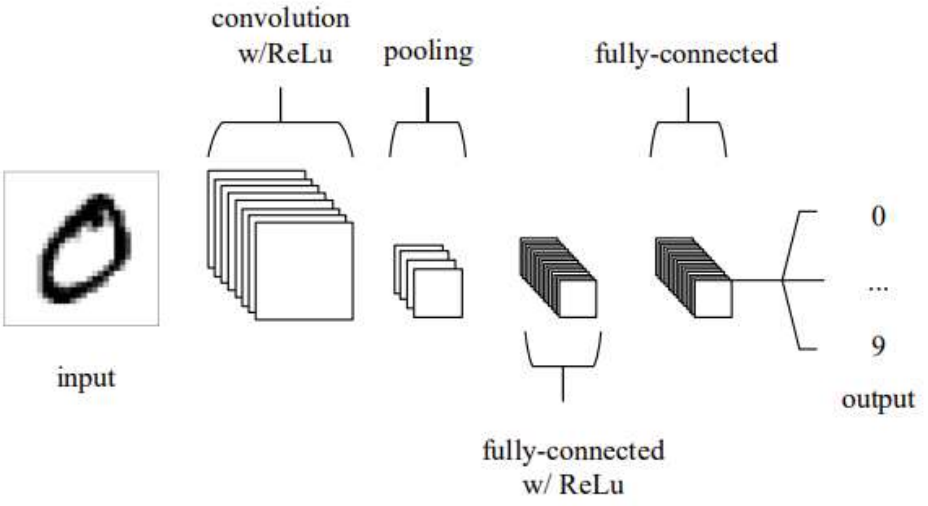


Figure 2. A simple CNN architecture [6].

Convolutional layers are present for feature extraction from the image. Each pixel is convolved with trainable weight filters to produce a new feature map, which is then fed to the activation function. The activation function receives the output from the upper layers and passes it to the math function. It contains two commonly used activation functions: Sigmoid and Relu. The main role of the activation function is to help the artificial neural network to learn and understand the very complex and nonlinear data.

DOI: 10.54254/2755-2721/32/20230209

Pooling layers employ a filter to systematically examine the entirety of the input, but the filter utilized in the pooling operation does not include any weights. The essence of the pooling operation is a kind of downsampling, which completes the features' compression and downsampling. Pooling operations are also performed on the input image in a sliding fashion, where representative values are selected and output according to some rule for the neuron activation values in the region covered by each pooling kernel. In most CNNs, the pooling layer is maximum, which means that the maximum activation values in the region of the pooling kernel are selected as the output values for downsampling. Maximum pooling can increases the invariance of the local movement of objects in the input image.

A fully connected layer, which is in general found at the end of the CNN, contains neurons with directly connections to neurons in both neighboring layers, not connected to either of them, and its role is to purify and synthesize the multidimensional features and pass them to the subsequent regression analysis layer or classifier to accomplish the corresponding tasks.

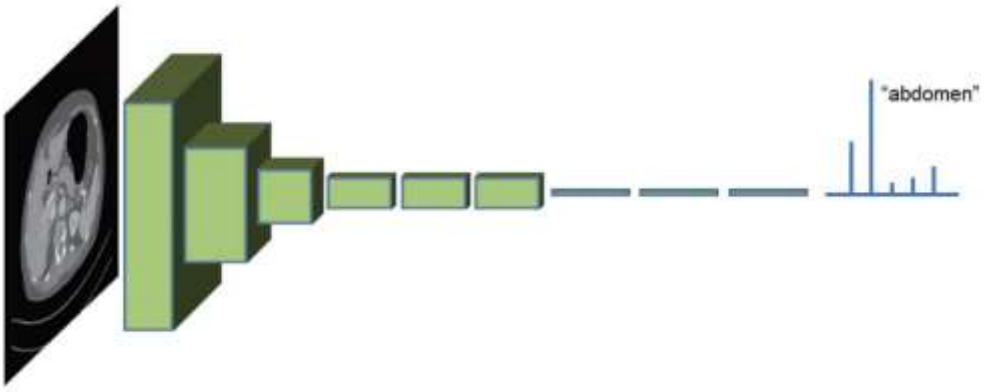


Figure 3. Convolutional neural network (CNN) [7].

Figure 3 shows an overview schematic of a representative CNN architecture which is capable of producing predictions for individual images through the utilization of softmax outputs for the purpose of multi-class categorization [7]. CNNs have been successfully applied to many medical image segmentation tasks. Kayalibay et al. demonstrated a CNN-based medical image segmentation method for bone and tumor segmentation tasks for hand and brain MRI, respectively [8].

However, the utilization of the CNN model for image classification leads to the compression of the 2D matrix information in the original image by the fully connected layer. Consequently, this compression results in the loss of spatial information inside. Given the significance of spatial information in semantic segmentation tasks, its influence on the use of CNN models for picture segmentation is noteworthy. Deep learning image segmentation algorithms were mainly realized by sliding image blocks in the early days, i.e., a fixed-size image block is intercepted around the target pixel and fed into the CNN. The classification result obtained is the category to which the current pixel belongs. The image block sliding method has many repetitive computation operations and low efficiency, and the segmentation accuracy is directly limited by the image block size, which has certain limitations.

3.2. Full convolutional network

Long et al. [9] proposed a Full Convolutional Network (FCN) to overcome the limitations of CNN. FCN designs an end-to-end, pixel-to-pixel encoding and decoding structure that solves the semantic segmentation problem of CNN and reduces the loss of spatial information. Meanwhile, FCN can take inputs of any size and, using effective inference and learning, generate outputs of the same size, eliminating the limitation of CNN that has limitations on image size. Fig. 4 depicts the standard configuration of an FCN for the semantic segmentation of image slices, acquired by using computed tomography (CT) with ConvNets.

DOI: 10.54254/2755-2721/32/20230209

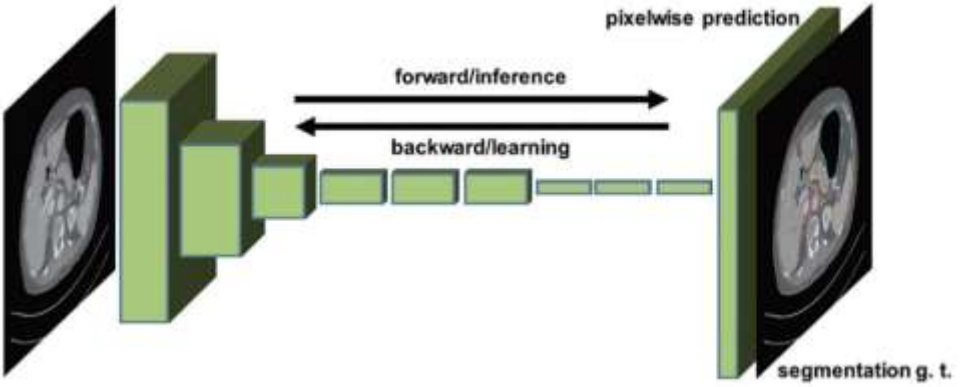


Figure 4. Fully convolutional network (FCN) [7].

FCN networks add jump-linking devices of different layer degrees to the part of the coding layer, which is implemented by summing the convolution with the corresponding anti-convolution for feature fusion. In FCN, a transposed convolutional layer replaces the final densely connected layer of the CNN in order to upsample the network's low-resolution feature maps. While performing semantic segmentation, this process restores the input image's original spatial dimension, thus solving the semantic segmentation problem of CNNs and reducing the loss of spatial information. FCNs only need to compute the softmax at each pixel of the final feature maps. In this way the shallow representational information is used to complement the spatial details of the deeper semantic information, thus achieving a more efficient and efficient feature fusion. information to complement the spatial details of the deeper semantic information to achieve more accurate results and ensure the network's robustness.

FCN can effectively learn to make dense predictions for per-pixel tasks. The proposed FCN eliminates the limitation of input image size, simplifies the preprocessing process, and reduces the loss of spatial information. However, the results obtained by a simple upsampling operation are still not fine enough, and the segmented output maps are still blurry, smooth, and insensitive to the details in the image.

3.3. U-Net

Based on FCN, U-Net has made further improvements. Figure 5 shows the structure of U-Net [10]. The difference between U-Net and FCN is that the jump link operation of U-Net is a superposition operation carried out together with the deconvolution simultaneously, and the shallow representational information gives the deep semantic information. U-Net also uses mirroring operation for the edge processing of convolution to ensure that the corresponding encoder and decoder have the same size. Gordienko et al. performed lung segmentation experiments on chest X-ray images using the U-Net network, and the results show that U-Net network can perform medical image segmentation quickly and accurately.

DOI: 10.54254/2755-2721/32/20230209

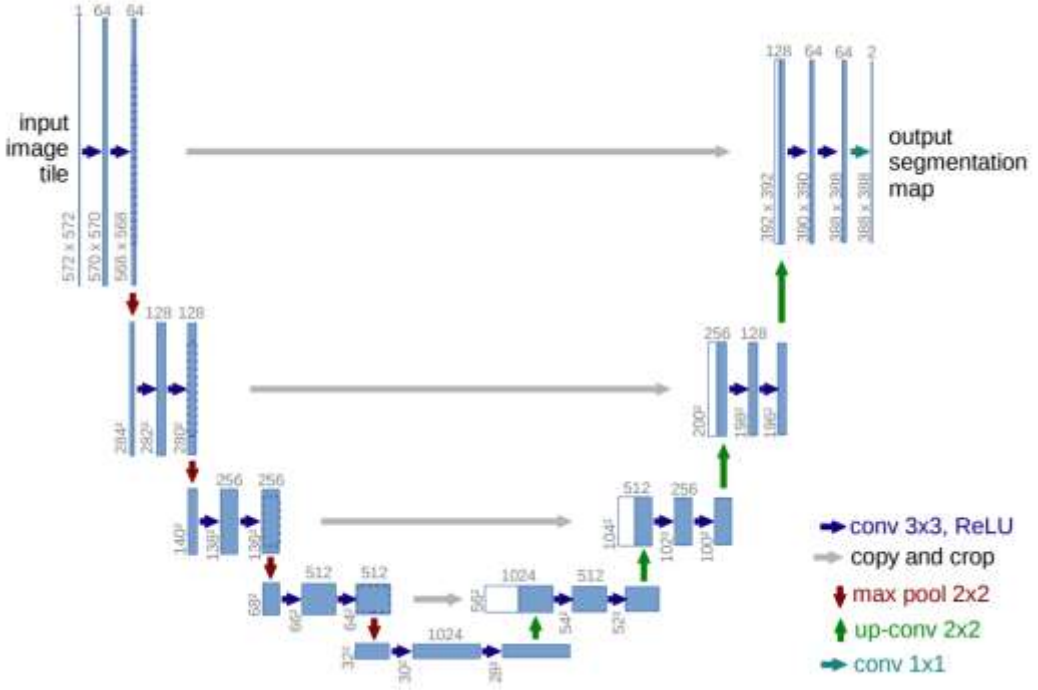


Figure 5. U-Net architecture [10].

Since U-Net is proposed for medical image segmentation, it has attracted extensive attention from research scholars as soon as it was proposed. In 2016, Cicek et al. extended the original U-Net network architecture so as to establish a 3D U-Net network architecture [11]. The authors proposed that U-Net was originally designed for cellular segmentation of 2D images, whereas much of the medical image data is actually volumetric data in 3D. Although it is possible to split the volumetric data into 2D image sequences for processing, this approach ignores the positional relationships between the different layers and often the images at different positions differ significantly, which is not conducive to the network learning generic features. The authors found that biomedical images are very rich in volumetric data. The computer screen can only show 2D slices, making it challenging to mark the segmentation labels immediately on the 3D level. However, neighboring 2D slices often contain approximate picture information, so 3D U-Net can learn to generate high-density volumetric segmentation by simply training on sparsely labeled 2D images. 3D U-Net input 3D volume and processes it by replacing U-Net's original 2D operations with corresponding 3D operations with relatively good experimental results.

In addition, Zhou et al. further optimized U-Net in 2018 and proposed UNet++ to meet the demand for more accurate segmentation [12]. UNet++ adds more jump connection paths and up-sampled convolutional blocks to the original UNet network architecture, and the intermediate hidden layer deep supervision is used, that spans the semantic divide between encoder and decoder, and solves the challenges of gradient vanishing during UNet++ network training and reduces the inference time of the model. The experimental results in Figure 6 demonstrate that UNet++ performs better than U-Net.

DOI: 10.54254/2755-2721/32/20230209

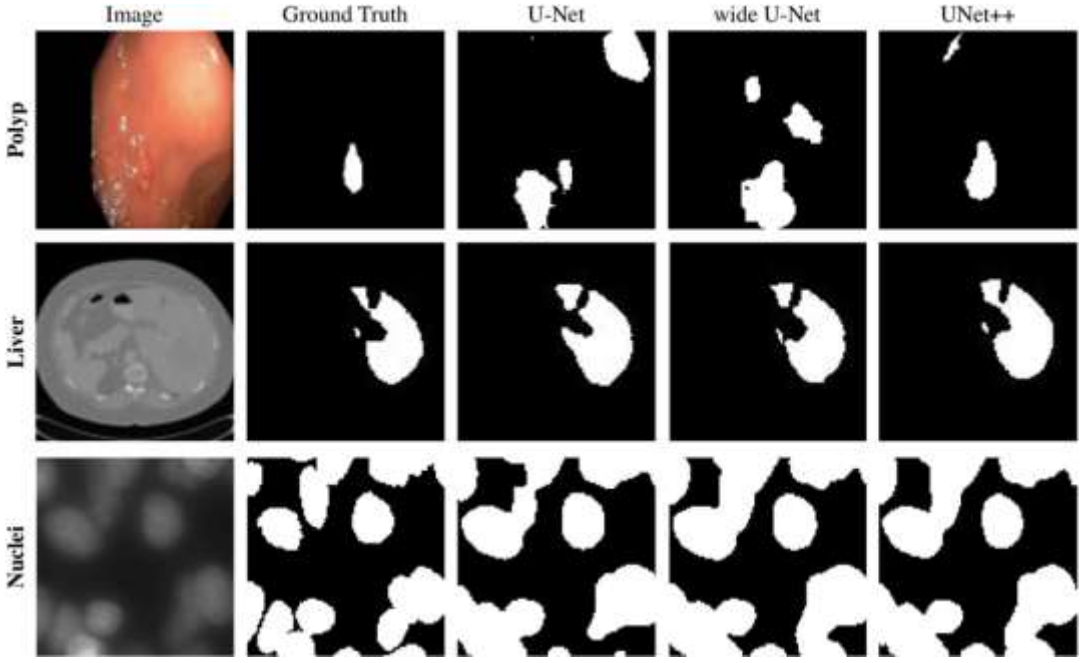


Figure 6. Segmentation results based on U-Net and UNet++ [12].

In recent years, more and more improvements and variants based on U-Net have been continuously proposed. Huang et al. proposed Unet 3+ in 2020, which further proposes a hybrid loss function to improve the borders and receive better segmentation outcomes; Cao et al. proposed Swein-unet, which extract context features using hierarchical SwinTransformer with shifted windows as the encoder. These researches have promoted the development of medical image segmentation technology and also expanded new ideas for future development.

4. Conclusion

This paper provides an overview of the importance and research background of medical image segmentation technology and explains the task of it. This paper introduces the principles of traditional image segmentation methods and their limitations. This paper reviews deep learning based medical image segmentation technology, introduce CNN, FCN, U-Net and its variants, and their respective principles, structures and applications are described.

Based on the research of this paper, it is known that the traditional medical image technology is easily interfered by noise and less flexible, while deep learning-based one is able to segment the image with more robustness and high accuracy, and can automatically obtain the most suitable features, which makes up for the shortcomings of the traditional ones. Deep learning-based medical image segmentation has brought a qualitative leap for medical image processing compared with traditional image segmentation techniques, and still has great potential for future development. For example, it can realize real-time medical image segmentation by compressing the model while ensuring the accuracy and stability; for example, uncertainty analysis algorithms can be added to allow the model to give the segmentation results while pointing out uncertain segmentation, so that the doctor can intervene to correct the segmentation, and ensure the quality of the segmentation and the results in the actual clinical application; for example, it can leave the support of large-scale high-quality labeled datasets, improve the weakly supervised learning under sparse labeling and data enhancement under small datasets, and expand the training set for the deep model, and so on, all of which are the directions for future development and research. Using the deep learning method, we can accomplish many complex medical image segmentation that are difficult to realize now, with great research significance and practical application value.

DOI: 10.54254/2755-2721/32/20230209

References

[1] Wang R, Lei T, Cui R, Zhang B, Meng H and Nandi A K 2022 Medical image segmentation using deep learning: A survey *IET Image Process.* **16** 1243–67

[2] Wang A, Zhang Q, Han Y, Megason S, Hormoz S, Mosaliganti K R, Lam J C K and Li V O K 2022 A novel deep learning-based 3D cell segmentation framework for future image-based disease detection *Sci. Rep.* **12** 342

[3] Zheng P, Zhu X and Guo W 2022 Brain tumour segmentation based on an improved U-Net *BMC Med. Imaging* **22** 199

[4] Fu Z, Zhang J, Luo R, Sun Y, Deng D and Xia L 2022 TF-Unet:An automatic cardiac MRI image segmentation method *Math. Biosci. Eng.* **19** 5207–22

[5] Bengio Y, Goodfellow I, Courville A 2017 Deep learning *Cambridge,MA, USA: MIT press*

[6] O'Shea K and Nash R 2015 An Introduction to Convolutional Neural Networks

[7] Roth H R, Shen C, Oda H, Oda M, Hayashi Y, Misawa K and Mori K 2018 Deep Learning and Its Application to Medical Image Segmentation **36**

[8] Kayalibay B, Jensen G and van der Smagt P 2017 CNN-based Segmentation of Medical Imaging Data

[9] Long J, Shelhamer E and Darrell T Fully Convolutional Networks for Semantic Segmentation

[10] Ronneberger O, Fischer P and Brox T 2015 U-Net: Convolutional Networks for Biomedical Image Segmentation

[11] Çiçek Ö, Abdulkadir A, Lienkamp S S, Brox T and Ronneberger O 2016 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation

[12] Zhou Z, Rahman Siddiquee M M, Tajbakhsh N and Liang J 2018 UNet++: A Nested U-Net Architecture for Medical Image Segmentation *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* 3–11

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230212

Research on image classification leveraging deep convolutional neural networks and visual cognition

Chen Liu

College of Software Engineering, Sichuan University, Chengdu, 610065, China

2020141060049@stu.scu.edu.cn

Abstract. The field of image classification has experienced remarkable improvements with the advent of deep learning techniques, especially Deep Convolutional Neural Networks. The present study provides an extensive exploration of the junction where image classification based on Deep Convolutional Neural Networks meets human visual cognition. Utilizing the inherent ability of these networks to automatically learn hierarchical features from raw pixel data, this research examines their potential in classifying images from diverse complex datasets, emphasizing predominantly on the extensively utilized ImageNet dataset. The initial aspect of this study involves training and evaluating models based on Deep Convolutional Neural Networks on the ImageNet dataset, which comprises millions of labeled images spanning across thousands of categories. Well-established network architectures such as AlexNet, VGGNet, GoogLeNet, and ResNet are employed, and their performance in the challenging task of image classification is assessed. Rigorous experiments highlight the strengths and weaknesses of each model while emphasizing the prospects of transfer learning and fine-tuning. Following this, the interpretability of Deep Convolutional Neural Networks is explored by using visualization techniques to comprehend the learned feature representations. By visualizing activation maps and class-specific saliency maps, valuable insights are gained into the regions of interest that guide the decision-making of these models. Moreover, the correlation between the features extracted by these models and human visual attention mechanisms is examined to shed light on the focus of attention of the models. The study also addresses the difficulties that adversarial attacks, data bias, and generalization capabilities present to Deep Convolutional Neural Networks. Strategies to enhance the robustness and adaptability of the models across various domains are examined, linking these observations to human cognitive behavior.

Keywords: cognitive science, computer vision, feature extraction, semantic, image classification.

1. Introduction

The research of depth convolutional neural network network (DCNN) in the field of image classification has far-reaching theoretical and practical significance. First, the need for automatic image recognition and classification has exploded in various fields, such as autopilot, security monitoring, medical diagnosis, space exploration and so on. These applications require high accuracy and efficiency of image classification, while the traditional rule-based or statistical image processing methods can process complex, large-scale and high-noise image data, the results are often poor.

DOI: 10.54254/2755-2721/32/20230212

As a deep learning model simulating human neural network, DCNN has shown excellent performance in image recognition task. By learning a large amount of image data, DCNN can automatically extract and learn the hierarchical features of the image, and carry out effective classification. However, despite the impressive performance of DCNN, its internal working mechanism is not clear, especially the process of how to extract and integrate image features, which leads to its performance in some specific scenarios, for example, the classification of novel objects, the classification of images with complex background or subtle changes, there are still challenges. Therefore, the research of DCNN in image classification can not only promote the progress of image processing technology and solve the problems in practical application, but also can compare and imitate the visual processing mechanism of human brain, to deepen our understanding of human visual cognitive processes and promote the cross-study of cognitive science, neuroscience and artificial intelligence.

2. Relevant theories and techniques

2.1. Feature integration theory

The Feature Integration Theory is a cognitive psychology concept developed by Anne Treisman and Garry Gelade in 1980. This theoretical framework provides an understanding of how different characteristics are combined during the perception and recognition of objects.

According to this theory, the initial stages of the perceptual process involve a primary analysis of external stimuli, including color, shape, and motion. These attributes undergo separate processing, then are relayed to the brain's early visual processing stage, referred to as the preattentive stage. Attention plays a crucial role within this theory. Controlled by attention, the information corresponding to diverse features is amalgamated, thereby forming a complete object perception. This integration occurs within the later stages of the visual system, recognized as the focused attention stage.

Feature Integration Theory underscores the importance of two stages: the independent processing of characteristics and their subsequent integration. It suggests that when there are substantial differences between the features of external stimuli, they can be easily consolidated into a comprehensive object during the integration phase. Conversely, if these features exhibit similarity, the integration procedure might be disturbed, potentially leading to errors in perception or recognition difficulties.

One classic experiment related to Feature Integration Theory is the "Feature Search" experiment proposed by Anne Treisman and Garry Treisman. In this experiment, participants are asked to find a specific target object within a group of objects with the same features, for example, finding a green circle among a group of red circles. The results show that the search task is easier when there is a significant feature difference between the target and the background. As the feature similarity between the target and the background increases, the search task becomes more difficult.

Feature Integration Theory is significant for understanding the mechanisms of human perception and visual cognition. It reveals the roles of feature analysis and integration in visual processing, as well as the modulation of attention in the integration process. This theory has broad applications in cognitive psychology, computer vision, and human-computer interaction.

2.2. Principles of human-computer interaction

The Principles of Human-Computer Interaction (HCI) provide guidelines and concepts for designing user-friendly and effective interfaces between humans and computer systems. These principles aim to enhance the usability, efficiency, and overall user experience of interactive systems. Here are some key principles:

User-Centered Design: The design process should revolve around the needs, goals, and abilities of the users. It involves understanding the users' tasks, behaviors, and preferences, and incorporating their feedback throughout the design and development stages.

Visibility and Feedback: The system should provide clear and immediate feedback to users, informing them about the system's state and the outcome of their actions. Visual cues, progress

DOI: 10.54254/2755-2721/32/20230212

indicators, and error messages are examples of providing feedback to enhance user understanding and control.

Consistency: The interface elements and interactions should be consistent throughout the system to minimize cognitive load and facilitate learning. Consistency includes using standardized design patterns, terminology, and navigation structures to create a predictable and familiar user experience.

Simplicity: The interface should be kept simple and intuitive, avoiding unnecessary complexity. This principle emphasizes removing unnecessary elements, reducing cognitive overload, and presenting information and functionality in a clear and understandable manner.

Flexibility and Efficiency: The system should provide flexibility for different user preferences and support various workflows. It should allow users to customize settings, provide shortcuts, and streamline repetitive tasks to enhance efficiency and productivity.

Error Prevention and Recovery: The interface should be designed to prevent errors through clear instructions, proper affordances, and validation mechanisms. Additionally, it should provide users with the ability to undo or recover from errors and provide meaningful error messages to guide users in resolving issues.

Accessibility: The interface should be accessible to users with different abilities, including those with disabilities. Design considerations such as proper contrast, keyboard accessibility, alternative text for images, and assistive technology support are essential for inclusive design.

Learnability: The system should be easy to learn and navigate, enabling users to quickly understand its functionality and features. Clear instructions, onboarding processes, and intuitive interactions contribute to the learnability of the system.

These principles, along with user research and iterative design processes, guide HCI professionals in creating interfaces that meet users' needs and enhance their overall interaction with computer systems.

2.3. Deep convolutional neural network

Deep Convolutional Neural Networks (DCNNs) have revolutionized the field of computer vision, particularly in image classification tasks. Several popular DCNN architectures have been developed over the years, including AlexNet, VGGNet, GoogLeNet, and ResNet. Here's a brief overview of these architectures: As shown in Figure 1.

Figure 1. Deep neural network and its development [1].

AlexNet: AlexNet, proposed by Alex Krizhevsky et al. in 2012, played a pivotal role in popularizing deep learning for image classification. It was the winner of the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) in 2012. AlexNet consists of eight layers, including five convolutional layers and three fully connected layers. It introduced the concept of using Rectified Linear Units (ReLU) as activation functions, local response normalization, and dropout regularization. As shown in Figure 2.

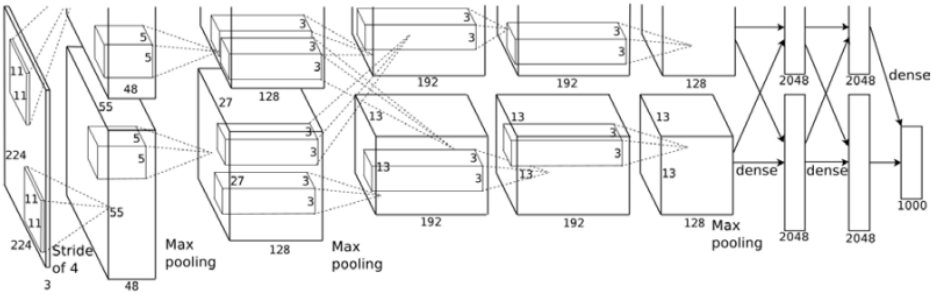


Figure 2. Alex Network structure diagram [2].

DOI: 10.54254/2755-2721/32/20230212

VGGNet: VGGNet, developed by the Visual Geometry Group at the University of Oxford in 2014, is known for its simplicity and depth. It is characterized by its uniform architecture, where 3x3 convolutional filters are stacked repeatedly, and max-pooling is performed after every two or three convolutional layers. VGGNet achieved excellent performance in the ILSVRC 2014 competition and is widely used as a baseline network for many computer vision tasks. As shown in Figure 3.

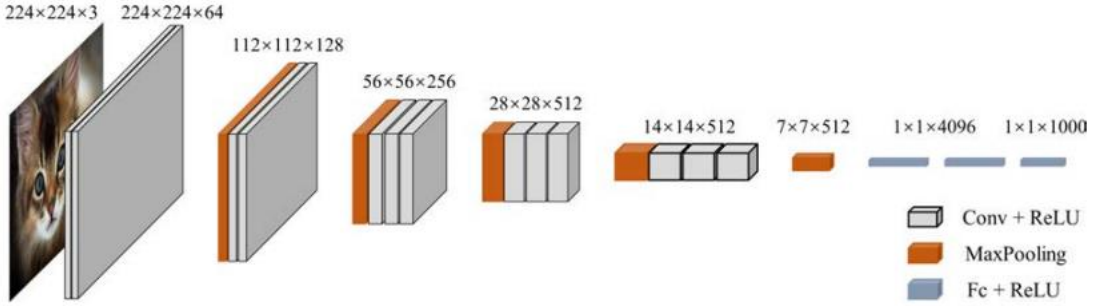


Figure 3. VGG Network structure diagram [3].

GoogLeNet (Inception): GoogLeNet, introduced by Szegedy et al. from Google Research in 2014, introduced the concept of the Inception module. It aims to address the trade-off between network depth and computational efficiency. The Inception module performs multiple convolutions with different filter sizes and concatenates their outputs, allowing the network to capture features at different scales. GoogLeNet was the winner of the ILSVRC 2014 competition and demonstrated state-of-the-art performance with a significantly reduced number of parameters. As shown in Figure 4.

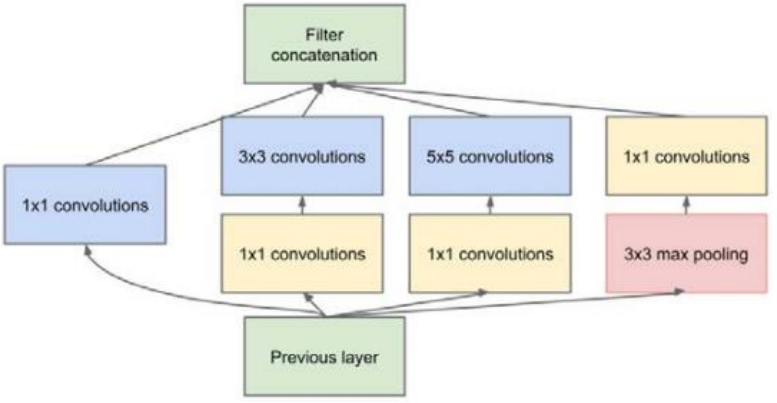


Figure 4. GoogLe network structure diagram [4].

ResNet: ResNet (Residual Network), proposed by Heetal. in 2015, introduced the concept of residual connections to address the degradation problem in deep networks. Residual connections allow information to bypass certain layers, enabling the network to learn residual mappings. This architecture significantly improved the training and performance of very deep networks. ResNet won the ILSVRC 2015 competition and has been widely adopted in various computer vision tasks. As shown in Figure 5.

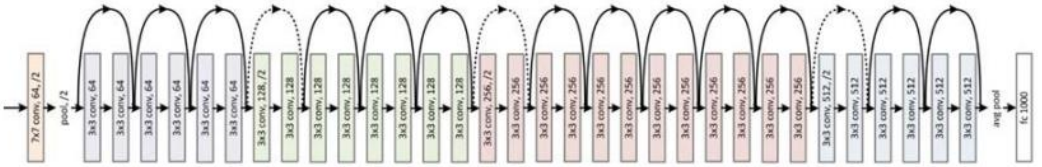


Figure 5. ResNet34 Network structure diagram [5].

DOI: 10.54254/2755-2721/32/20230212

These DCNN architectures have had a significant impact on the field of computer vision, demonstrating exceptional performance in image classification, object detection, and other related tasks. They serve as foundational models and have paved the way for subsequent advancements in deep learning and computer vision research.

3. System analysis and application research

3.1. Experimental framework

The experimental framework for image classification based on DCNN and visual cognition typically involves the following components: As shown in Figure 6.

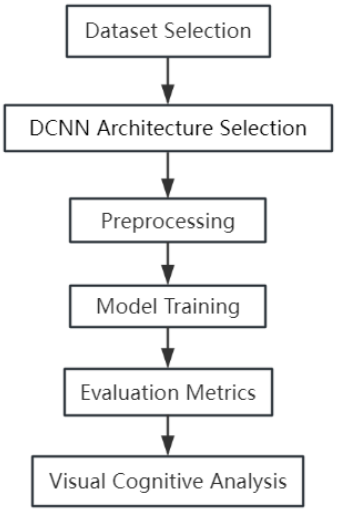


Figure 6. Experimental process diagram [6].

3.2. Data set

The ImageNet dataset is an open, publicly available dataset that allows researchers and developers to use it freely for image classification tasks.

The images in ImageNet are from real-world scenes and contain a variety of objects, backgrounds, and perspectives. Compared with the synthetic data set, ImageNet images are closer to the actual application scene, which makes the trained model have better generalization ability. As shown in Figure 7.

Figure 7. The ImageNet dataset [7].

DOI: 10.54254/2755-2721/32/20230212

3.3. Research methods/approach

DCNN Architecture Selection: AlexNet, VGGNet, GoogLeNet, or ResNet, based on the specific requirements of the experiment. Consider factors like model complexity, performance, and availability of pre-trained models. As shown in Figure 8.

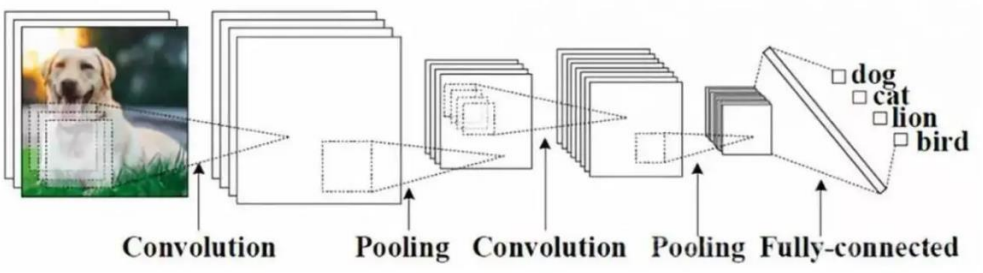


Figure 8. The convolutional neural network [8].

The formula for convolution is shown in Formula (1):

$$X_j^l = f(\sum_{i=m_j} k_{ij}^l * X_i^{l-1} + b_j^l) \quad (1)$$

The formulas for average and maximum pooling are shown in (2) and (3):

$$X_j^l = f(\frac{1}{m}\sum_{i=m_j} X_i^{l-1} + b_j^l) \quad (2)$$

$$X_j^l = f(max_{i=mj} X_i^{l-1} + b_j^l) \quad (3)$$

Preprocess the dataset and images to ensure uniformity and compatibility with the chosen DCNN architecture. Here are some common preprocessing techniques: Image Resizing: Resize the images to a consistent size to ensure uniformity in the input data. This step is necessary because images in a dataset may have different resolutions and aspect ratios. The input image size of all models was scaled to 224 × 224, and then the input feature extraction model was used, and the feature extraction model was pre-trained on ImageNet in PYTORCH model library to extract the full-connection layer, the output size of AlexNet is 256 × 6 × 6, the output size of VGG is 512 × 7 × 7, the output size of GoogLeNet is 1024 × 1 × 1, and the output size of ResNet is 512 × 1 × 1.

Data Normalization: Normalize the pixel values of the images to bring them within a specific range or distribution. Common normalization techniques involve scaling the pixel values between 0 and 1 or standardizing them using mean and standard deviation. Normalization helps to mitigate the impact of varying intensity levels and facilitates stable model training.

Data Augmentation: Augment the dataset by applying transformations to the images. Data augmentation techniques can include random rotations, translations, flips, and crops. By introducing variations in the training data, data augmentation helps improve the model's generalization capability, reduces overfitting, and increases the effective size of the dataset.

Noise Reduction: Apply noise reduction techniques, such as Gaussian blurring or median filtering, to remove or reduce image noise. This can enhance the clarity of the images and reduce the influence of noise during feature extraction and classification.

Model Training: Train the DCNN model using the labeled dataset. This typically involves feeding the preprocessed images through the network, adjusting the model's weights and parameters using optimization algorithms like stochastic gradient descent (SGD) or Adam, and iteratively updating the model to minimize a predefined loss function.

DOI: 10.54254/2755-2721/32/20230212

3.4. Experimental results and analysis

3.4.1. Evaluation metrics. Model Training: The training of the Deep Convolutional Neural Network model commences with the ImageNet dataset, comprising millions of images spanning a wide variety of categories. Each training cycle, also known as an epoch, involves processing a batch of images with the model. The model's weights get updated via an optimization algorithm known as stochastic gradient descent, and subsequently, the model's performance is assessed based on the training data. As the training process unfolds, the model's accuracy and the associated loss at the end of each epoch are recorded. This allows for tracking the model's learning progress over time. To better visualize the progression of the training, a graph is plotted. The horizontal axis represents the number of training epochs, while the vertical axis shows the corresponding training accuracy and loss values. This graphical representation helps in understanding the relationship between the number of iterations and the performance of the model. As shown in Figure 9.

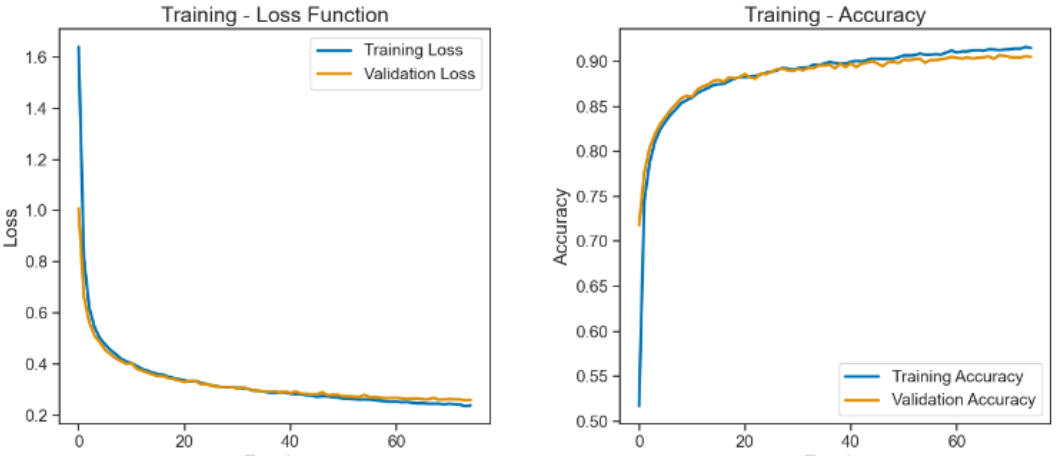


Figure 9. The relationship between training accuracy and loss [9].

3.4.2. Visual cognitive analysis. Incorporate visual cognitive analysis techniques to understand the model's behavior and relate it to human visual cognition. compare the model's predictions with human perception and decision-making.

To delve into the interpretability of DCNNs and gain insights into the learned feature representations and attention focus, several visualization techniques can be employed. Some commonly used techniques include:

Activation Maps: Activation maps, also known as feature maps or activation patterns, visualize the response of individual neurons or filters in the DCNN to specific input stimuli. By visualizing the activation maps of different layers, researchers can identify which regions of the input image trigger higher activations in the model, providing clues about the learned features and the model's perception of different visual patterns.

Class Activation Maps (CAM): Class activation maps highlight the regions of an input image that contribute most to the model's decision for a specific class. CAMs are derived from the gradients of the output with respect to the feature maps, providing insights into which image regions are relevant for the model's classification decision. As shown in Figure 10.

DOI: 10.54254/2755-2721/32/20230212

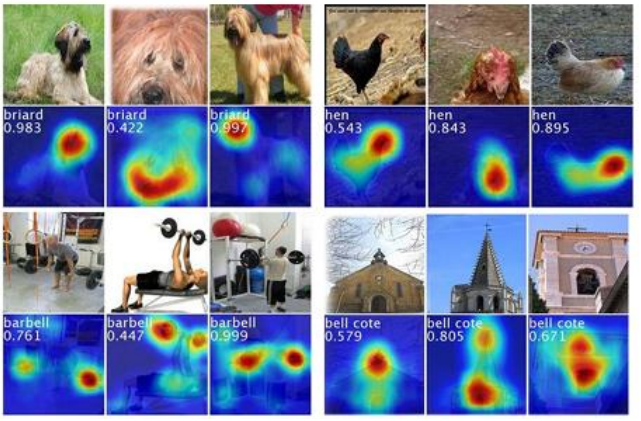


Figure 10. The CAMs of four classes from ILSVRC [10].

Saliency Maps: Saliency maps highlight the most salient or informative regions of an input image that contribute to the model's prediction. These maps are computed by measuring the sensitivity of the model's output to small changes in input pixels. Saliency maps can reveal which parts of an image receive the most attention from the model during classification. As shown in Figure 11.

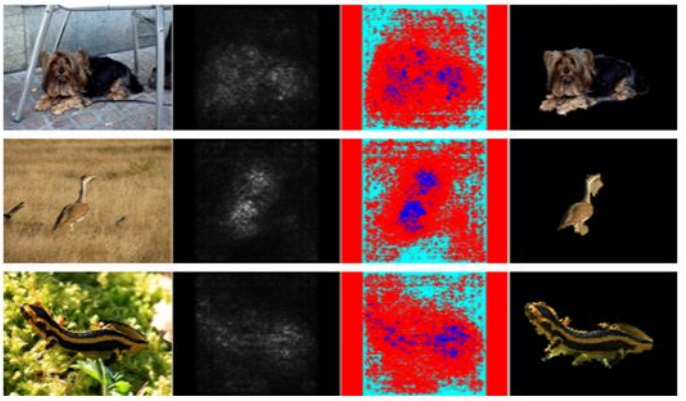


Figure 11. Object recognition and segmentation in image [11].

4. Challenges

Interpretability and explainability: DCNNs are black-box models, making it difficult to interpret and understand their internal workings and decision processes. This poses a challenge for visual cognition research in image classification. From the perspective of understanding human visual cognition and decision-making, mapping and explaining the outputs of DCNNs to human cognitive processes remain an open question.

Data bias and imbalance: DCNNs often require large amounts of labeled data for training in image classification. However, real-world datasets often exhibit class distribution imbalance and labeling errors. This can lead to DCNNs learning biases toward common classes or poor recognition performance for rare classes. Addressing data bias and imbalance is a challenge in image classification research.

Adversarial attacks and robustness: DCNN models in image classification can be susceptible to adversarial attacks, where slight perturbations to input images cause the model to produce incorrect classification results. This suggests differences between the processing of visual information by DCNNs and human visual cognition. Improving the robustness of DCNNs and their ability to resist adversarial attacks to better simulate human visual cognition is an important research direction.

Transfer learning and generalization: DCNN models may achieve good performance on the training set but still face challenges in generalizing to unseen data. How to transfer the learning capabilities of

DOI: 10.54254/2755-2721/32/20230212

DCNNs from one task to another and how to maintain performance stability in different domains or environments are issues that need to be addressed.

These challenges require further research and exploration, combining DCNNs with visual cognition research to improve the performance of image classification and understand the mechanisms of human visual cognition. Addressing these challenges will drive advancements in computer vision and cognitive science, facilitating the development of more effective and interpretable methods for image classification.

5. Conclusion

This study presents an exhaustive exploration of image classification leveraging Deep Convolutional Neural Networks (DCNNs), and how it aligns with human visual cognition. DCNNs are renowned for their ability to learn hierarchical features autonomously from raw pixel data. The study evaluates their performance using the challenging ImageNet dataset, employing well-known DCNN architectures such as AlexNet, VGGNet, GoogLeNet, and ResNet. The results underscore the potential of these networks in complex image classification tasks and transfer learning scenarios. Interpretability forms a crucial part of this study. Various visualization techniques are utilized to interpret the learned feature representations. Through visualizing activation maps and class-specific saliency maps, invaluable insights into the regions of interest that influence the model decisions are gathered. These visualizations illuminate the features learned by the model, revealing the model's perception of visual patterns.

The study additionally tackles challenges associated with adversarial attacks, data bias, and generalization capabilities in DCNNs. Methods to enhance model robustness and adaptability across multiple domains are investigated, striving to align the models' behavior with the intricacies of human cognitive processing.

In summary, the research provides deeper insight into image classification based on DCNNs and its relationship with human visual cognition. It exemplifies the capabilities of DCNNs in large-scale image classification tasks and their potential in transfer learning scenarios. The use of visualization techniques leads to valuable insights into the learned features and the model's attention focus. The visual cognitive analysis offers evidence of the models' strengths and limitations compared to human perception.

The conclusions drawn from this study serve as a stepping stone for future advancements in the field of computer vision research. By bridging the gap between machine-based image understanding and the complexities of human visual cognition, this research sets the groundwork for the development of more interpretable and human-aligned image classification models. As the landscape of deep learning continues to evolve, this research contributes significantly to the quest for more robust, interpretable, and human-like Artificial Intelligence systems in the field of image classification.

References

[1] Yan Jianpu, Y. (2022). Research on brain-computer hybrid intelligent computing method for image classification task [Dissertation]. Xidian University.

[2] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Proceedings of the 26th Conference on Neural Information Processing Systems (NeurIPS).

[3] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In Proceedings of International Conference on Learning Representations (ICLR), Boston.

[4] Szegedy, C., Liu, W., Jia, Y. Q., & others. (2015). Going deeper with convolutions. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston.

[5] He, K. M., Zhang, X. Y., Ren, S. Q., & others. (2016). Deep residual learning for image recognition. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas.

DOI: 10.54254/2755-2721/32/20230212

[6] Guangzhu, X., Zequn, Z., Silu, Y., & others. (2021). Flower Image Classification system based on lightweight deep convolutional neural network. Data Acquisition and Processing, 36(4), 756-768. https://doi.org/10.16337/J. 1004-9037.2021.04.014.

[7] Huiyong, W., Chunjie, X., Xiaoming, Z., & others. (2020). Image correlation measure based on DCNN classification. Computer Applications Research, 37(2), 625-629. https://doi.org/10.19734/J. ISSN. 1001-3695.2018.04.0487.

[8] Chen, Y., Argentinis, J., & Weber, G. (2016). IBM Watson: How cognitive computing can be applied to big data challenges in life sciences research. Clinical Therapeutics, 38(4), 688−701.

[9] Zhou, A., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2015). Learning Deep Features for Discriminative Localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). arXiv:1512.04150.

[10] Russakovsky, J., Deng, H., Su, J., Krause, S., Satheesh, S., Ma, Z., Huang, A., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & FeiFei, L. (2015). Imagenet large scale visual recognition challenge. International Journal of Computer Vision.

[11] Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. arXiv preprint arXiv:1312.6034.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230213

Deploying human body detection technologies in security systems: An in-depth study of the FASTER-GCNN algorithm

Chao Jiang

UWEC computer science college, University of Wisconsin Eau-Claire, Eau-Claire, 54701-6698, United States

jiangc0802@uwec.edu

Abstract. The field of human body detection, a pivotal area in computer vision, merits comprehensive discussion. Remarkable advancements have been achieved in the techniques for human body detection over the past few decades, with significant applications spanning various sectors. This discussion delves into the potential of human detection technology within the realm of security - a field that necessitates efficient and accurate human detection technology to promptly identify potential threats, suspicious behaviors, or unusual activities. Deep learning-based human detection algorithms have substantially improved capabilities in this domain, facilitating real-time tracking and identification of the human form, thereby enabling security personnel to respond swiftly. This paper employs the Faster-RCNN algorithm for model training, utilizing the Information and Automation Research (INRIA) database. The deep learning-trained model proves highly accurate in human body detection, effectively recognizing human movements and behaviors. Such capabilities hold immense potential for implementation within the security sphere, including video surveillance systems and other similar applications where effectiveness is crucial.

Keywords: Faster-RCNN, INRIA, deep-learning, human-detection.

1. Introduction

The function of the security sector primarily lies in safeguarding and preserving public safety, and mitigating security risks. In the contemporary era, security technologies and measures exert significant influences and implications on society, corporations, and individuals alike. The key roles in the realm of security encompass crime prevention through mechanisms such as video surveillance and intrusion detection. These measures help deter criminal activities, decrease the crime rate, and uphold societal security. They also protect property, guarantee public safety, facilitate monitoring and management, and enable timely response to emergencies. Incorporating human detection can sufficiently meet the demands of the security sector. This technology can identify potential threats, suspicious conduct, or anomalous activities, subsequently issuing a preliminary warning. In doing so, it aids in preventing criminal occurrences. Human detection technology can facilitate real-time monitoring and automatic alarms, swiftly initiating responsive and remedial actions. This significantly shortens the response duration to security incidents, enhances handling efficiency, and minimizes potential losses. Moreover, it amplifies the safety levels of public areas, transportation hubs, and crucial facilities, thereby ensuring

DOI: 10.54254/2755-2721/32/20230213

the security of citizens' and tourists' lives and properties. Additionally, this technology aids in maintaining societal order, thwarting terrorist assaults, and responding to emergencies.

Surveillance videos, images, and data obtained can serve as vital clues and evidence, assisting investigators in solving cases. This technology aids law enforcement in tracking suspects, identifying vehicles implicated in crimes, or observing other significant details. The application of human body detection technology in the security sector elevates public awareness and concern about security matters. By raising consciousness about security risks, this can stimulate individuals and organizations to take anticipatory security actions and contribute to collective efforts to maintain societal safety.

2. Background

Through research and surveys, the field of security is often overlooked for the vast majority of people, with technology and applications investing relatively little in security, and the deployment and maintenance of security systems usually requiring considerable investment [1]. For organizations or individuals with limited budgets, other urgent needs may take precedence and security may be put on the back burner [2]. Some regions or organizations may have a low perception of security risks and perceive the likelihood of a security incident as low, so that security measures are not seen as an urgent need. Some specific scenarios where the security field may not be able to provide a solution to meet a specific need due to technological limitations or immaturity, resulting in it not being a preferred consideration [3]. However, the importance of the security field is unquestionable and plays a major role in preventing accidents and disasters, maintaining social order, and protecting lives and property, and with the development of society, security is increasingly emphasized. Various security threats and criminal activities pose a threat to public safety and personal property, so the study of security technology and measures is an important way to meet the needs of society, so the field of security is worth being studied, the human body detection technology in the field of security will have a very good progress and the future of this paper, through the training of the model for the model and the deep learning, the use of Faster-RCNN algorithm is applied to detectron2 carrier, so that the human body detection is more accurate and has good results in the field of security [4].

3. The FASTER-GCNN algorithm

Detectron2 is renowned for its exceptional performance in the field of object detection and segmentation. It is a library built on top of the PyTorch deep learning framework, taking full advantage of the computational power of Graphics Processing Units (GPUs). Due to its high level of optimization and parallelization, Detectron2 enables efficient training and inference on large-scale datasets, offering fast and accurate results [5]. This library also provides numerous pre-trained models and model libraries, including traditional object detection models like Faster Region-Convolutional Neural Networks and RetinaNet, as well as instance segmentation models such as Mask Region-Convolutional Neural Networks. These pre-trained models can be employed as starting points, enabling users to rapidly construct and train high-performance models of their own.

The Faster Region-Convolutional Neural Networks is an advanced deep learning algorithm for object detection. Proposed in 2015 by Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun et al., this model improves upon its predecessors, Region-based Convolutional Neural Networks and Fast Region-Convolutional Neural Networks, by introducing a Region Proposal Network. This innovation facilitates end-to-end object detection, significantly enhancing detection speed [6]. The functioning of the Faster Region-Convolutional Neural Networks algorithm involves several stages, As shown in Figure 1:

- An input image is processed by a convolutional neural network to produce a feature map.
- A Region Proposal Network generates potential object regions using a sliding window approach. For each region, it predicts the probability of the presence of an object and the offset from the anchor box to the object's bounding box.
- These proposed object regions are then filtered using a technique called non-maximum suppression. This process eliminates overlapping regions, retaining only those with the highest confidence.

DOI: 10.54254/2755-2721/32/20230213

• The remaining candidate object regions are passed to a detection network that utilizes Region of Interest pooling to extract fixed-size feature maps.

• The detection network has two components: the classification branch predicts the likelihood of each object class within the candidate regions, while the regression branch estimates the offsets to adjust the bounding box coordinates.

• Lastly, non-maximum suppression is applied to the candidate regions after the classification and regression phases, yielding the final set of object detections.

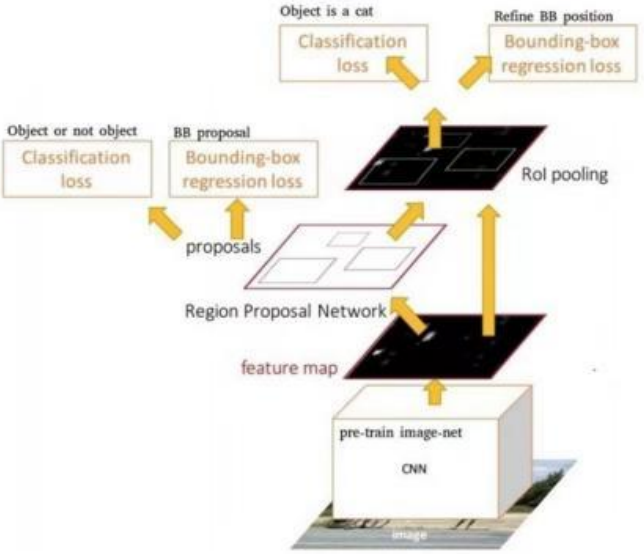


Figure 1. The architecture of Faster-RCNN (Photo/Picture credit: Original).

Fast R-CNN. RPN generates candidate boxes for objects, and Fast R-CNN extracts features from these boxes and classifies them [7]. The RPN uses a sliding window approach to generate these regions, and for each region, it predicts the probability of containing an object and the offset from the anchor box to the object's bounding box. After non-maximum suppression, the remaining candidate object regions are passed to the detection network. The detection network uses RoI (Region of Interest) pooling to extract fixed-size feature maps from the candidate regions. Detection Network: The classification branch predicts the probability of each candidate region containing each object class, and the regression branch predicts the offsets to adjust the bounding box coordinates of each candidate region [8]. The red box A represents the box before unregression, the blue box G' represents the box after regression. Each frame is represented as (x,y,h,w) and h,w denote the height and width of the frame [9].

(1) Do the panning first:

$$Gx' = Aw * dx(A) + Ax \tag{1}$$

$$Gy' = Ah * dx(A) + Ay \tag{2}$$

(2) Do the scaling again:

$$Gh' = Aw * dh(A) + Ah \tag{3}$$

$$Gw' = Ah * dw(A) + Aw \tag{4}$$

The parameters to be learned at this point are dh(A), dw(A), dx(A), dy(A). The corresponding learning objectives can be defined as follows.

$$t_x = \frac{(G_x - A_x)}{A_x} \tag{5}$$

DOI: 10.54254/2755-2721/32/20230213

$$t_y = \frac{(G_y - A_y)}{A_y} \tag{6}$$

$$t_h = \log \frac{G_h}{A_h} \tag{7}$$

$$t_w = \log \frac{G_w}{A_w} \tag{8}$$

A linear regression model can be used to define the learning process for d*(A): d*(A)=W*T*ϕ(A). The design loss function is: loss=i=1 ∑ 100 (t* i−W*T*ϕ(Ai)) The corresponding function optimization objective is: W* ‘ =argminw *(i = 1 ∑ 100(t* i−W*T*ϕ(Ai)) + ｜ ｜λ W*T ｜ ｜) W* can be solved by a stochastic gradient descent algorithm.

4. The database

The INRIA database is a computer vision-related database resource maintained by the computer vision team at the French National Institute for Information and Automation Research for target detection and pedestrian detection tasks. The INRIA Person Dataset contains a number of positive and negative sample images. Positive sample images contain labeled pedestrian targets, while negative sample images have no pedestrian targets. The dataset has a total of about 2,500 images containing about 1,200 positive samples (pedestrian targets) and about 1,300 negative samples (non-pedestrian targets) [10]. The images in the dataset are of high resolution and good quality, covering a wide range of scenes and complexities. Each positive sample image is labeled with the location and bounding box of the pedestrian target and is used to train and evaluate the performance of the pedestrian detection algorithm.

5. Experimental analyses

Following an in-depth comparative analysis of the Faster R-CNN and HOG+SVM algorithms, it is discerned that the former proves more precise in human body detection. Faster R-CNN is an advanced deep learning model for target detection, which leverages a Region Proposal Network (RPN) to facilitate an end-to-end detection process.

Historically, HOG and SVM have been employed as a classic algorithm pairing for target detection. In this combination, Histogram of Oriented Gradients extracts the image's features, while Support Vector Machine categorizes the targets from non-targets. The procedure includes several stages: image pre-processing to grayscale to eliminate illumination effects; gradient calculation at each pixel point; division into cells for gradient direction assignment; normalization of each cell's gradient direction histogram for shadow and lighting robustness; and, if color is considered, the inclusion of a color histogram in the HOG feature. The culmination of these steps is a feature vector formed by concatenating all cell gradient histograms.

Support Vector Machine is a supervised learning algorithm for binary classification. In the HOG+SVM target detection model, SVM segregates the target from non-target regions. The process initiates with preparing labeled image samples and extracting their HOG features, followed by labeling the HOG feature vectors for SVM training. The detection phase utilizes the trained SVM classifier to detect the image's target region by extracting HOG features via a sliding window and classifying them through SVM. The sliding window's size and step can be adjusted to the target size and detection accuracy. Ultimately, while HOG-SVM relies on machine learning, Faster R-CNN harnesses deep learning. By classifying the database images with the trained model, it is apparent that Faster R-CNN's recognition is superior. HOG+SVM, on the other hand, does not efficiently recognize potential targets, with some pedestrians escaping detection. This inefficiency is untenable in a security system demanding high safety factors. The superiority of Faster R-CNN in human body detection is further illustrated in a subsequent comparison of pedestrian detection using both models on the same set of images. As shown in Figure 2.

DOI: 10.54254/2755-2721/32/20230213

Figure 2. A comparison of pedestrian detection on images using these two models (Photo/Picture credit: Original).

6. Conclusion

In essence, the Faster-RCNN screening and detection functionality proves highly efficient in security systems. Comparative studies with other algorithms reveal the precision and effectiveness of Faster-RCNN, making it an ideal choice for high-stake security applications. However, it is essential to recognize some inherent limitations in the use of Faster-RCNN within the security paradigm. Faster-RCNN is a sophisticated deep learning model, composed of a Region Proposal Network and a Target Detection Network, both of which demand substantial computational resources and considerable storage space. This proves difficult for embedded devices with limited resource capabilities. In security situations, the potential subjects may range in size from small (like pedestrians) to significantly large (like vehicles). Faster-RCNN employs predefined anchor frames for target detection, but these frames may not cover all target scales, resulting in under-detection of small objects and misdetection of larger ones. Frequently, multiple targets may overlap significantly, as seen with crowded pedestrians. In such scenarios, Faster-RCNN could yield redundant detection results, causing inaccuracies in target counts or incomplete detection frames. Speed is crucial in security scenarios, but Faster-RCNN's detection rate on large-scale images can be slow, making it unsuitable for high real-time performance requirements. Complex security situations may present targets from varied viewpoints or obscured by other objects. These conditions may lead to the Faster-RCNN failing to detect targets accurately, resulting in false detections. Additionally, Faster-RCNN requires extensive, well-labeled training data for optimal detection results, a task that can be costly and time-consuming within the security sector. Tailoring the

DOI: 10.54254/2755-2721/32/20230213

model to specific security situations is also necessary to ensure real-world application success, representing yet another limitation. Nevertheless, future improvements to complement and enhance the security system could include the deployment of lighter weight target detection algorithms, model optimization for specific scenarios, or the incorporation of other sensors and technologies to bolster detection accuracy and real-time performance. Emerging trends suggest that the integration of reinforcement learning into target detection and security will gradually become more commonplace. By learning through environmental interaction, security systems can adaptively optimize detection strategies, delivering superior performance in complex environments. As target detection technology expands, concerns around data privacy and security will also rise. Future advancements will prioritize creating intelligent, efficient, comprehensive, and secure target detection systems while preserving user privacy and data security. The evolution of target detection technologies signals a move towards more intelligent, efficient, comprehensive, and secure solutions. These advancements will profoundly influence the development of security systems and other fields, resulting in improved safety and convenience for society at large.

References

[1] Hung G L , Sahimi M S B , Samma H ,et al.Faster R-CNN Deep Learning Model for Pedestrian Detection from Drone Images[J].SN Computer Science, 2020, 1(2):116.

[2] Albinali H , Alzahrani F A .Faster R-CNN for Detecting Regions in Human-Annotated Micrograph Images[C]//2021 International Conference of Women in Data Science at Taif University (WiDSTaif).2021.

[3] Lukac Y N .Pedestrian Detection based on Faster R-CNN[J].International Journal of Performability Engineering, 2019, 15(7).

[4] Baussard A , D'Acremont A , Quin G ,et al.Faster-RCNN with a compact CNN backbone for target detection in infrared images[C]//Conference on Artificial Intelligence and Machine Learning in Defense Applications.2020.

[5] Panigrahi S , Raju U S N .Pedestrian Detection Based on Hand-crafted Features and Multi-layer Feature Fused-ResNet Model[J].International Journal on Artificial Intelligence Tools, 2021.

[6] Zhao R , Li C , Ye S ,et al.Butterfly Recognition Based on Faster R-CNN[C]//IOP Publishing.IOP Publishing, 2019:032048-.DOI:10.1088/1742-6596/1176/3/032048.

[7] Gao F , Fu L , Zhang X ,et al.Multi-class fruit-on-plant detection for apple in SNAP system using Faster R-CNN[J].Computers and Electronics in Agriculture, 2020(176-):176.

[8] Kihara H , Ikejiri K , Ishizaki S ,et al.A Study of the Suitable Electrode Position on the Head for High Accuracy R-peak Detection from Electrocardiogram[J].The Proceedings of the Symposium on sports and human dynamics, 2019:A-29.DOI:10.1299/jsmeshd.2019.A-29.

[9] Huang X , Li X , Hu Z .Cow tail detection method for body condition score using Faster R-CNN[C]//2019 International Conference on Unmanned Systems and Artificial Intelligence (ICUSAI).2019.

[10] Zhang W , Mi Z , Zou Y ,et al.Joint HFaster-RCNN and Bayesian Posterior Probabilities for Pedestrian Detection[C]//2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE).IEEE, 2019.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230214

Predictions of diabetes through machine learning models based on the health indicators dataset

Xinyi Ren

Lancaster University, Lancashire, LA1 4YW, The United Kingdom

1625201499@qq.com

Abstract. Diabetes is a chronic disease that is widespread in the United States. Patients with diabetes will lose the ability to effectively regulate blood glucose levels and the disease can lead to increased economic burden for patients and generate enormous public health impact. The main purpose of this paper is to find out the indicators that are highly associated with diabetes and build a model to predict diabetes. The original dataset is from BRFSS (the Behavioral Risk Factor Surveillance System). For this project, a cleaned dataset on Kaggle for the year 2015 was used, which has 253,680 survey responses to CDC (Centers for Disease Control and Prevention)'s BRFSS with the target variable diabetes and 21 feature variables. The Chi-square test was applied to investigate the association between feature indicators and diabetes and built several machine learning models for predicting the disease. The selected model is Cat Boost Classifier with 86.6% accuracy for the testing set. According to the Permutation Feature Importance based on the Cat Boost Classifier, the most important 5 features were General Health (GenHlth), BMI (Body Mass Index), Age, high blood pressure (HighBP), and high cholesterol (HighChol) variables.

Keywords: diabetes prediction, machine learning, health indicators, classification.

1. Introduction

Diabetes is one of the most common and widespread chronic diseases in the US [1]. A diabetes patient does not produce enough insulin to regulate sugar in the body and can have many complications. In 2018, among the US population, 34.3 million people of all ages had diabetes, taking up 10.5% of the population [2,3], which increases the risk of complications of diabetes such as heart disease, cardiovascular events, microvascular disease, and even premature death. It can be seen that diabetes has a great impact in the US and poses a threat to personal health. The Behavioral Risk Factor Surveillance System (BRFSS) is the largest health-related telephone survey system in the United States, completing more than 400,000 adult interviews each year [4]. The primary purpose of this system is to collect data on health-related risk behaviours, chronic conditions such as diabetes, and the use of preventive services among US residents.

In this project, the 2015 BRFSS dataset, with 21 health indicators that may be associated with diabetes, was analysed. The goal is to identify the indicators highly correlated with diabetes and develop predictive models for diabetes, which could help facilitate early diagnosis and intervention. There are 23580 records from the 2015 BRFSS dataset. Diabetes generally includes two main types: type 1 and type 2, accounting for 5% and 95% respectively [2]. The dataset in this project does not separate type 1

DOI: 10.54254/2755-2721/32/20230214

and type 2 diabetes. Due to the extremely high proportion of type 2, the characteristics analysed in this paper are more likely to be applicable to type 2 diabetes. The features that are highly correlated with diabetes were found by using correlation analysis and the Chi-square test. Through these selected feature variables, several models of supervised machine learning algorithms were built for predicting diabetes, including Gaussian Naive Bayes, Decision Tree, Random Forest, Logistic Regression, Gradient Boosting, Linear Discriminant, and Cat Boost classifiers. These classifiers were evaluated by training and testing accuracy and Mean Squared Error to find the best model, thus providing help for the early diagnosis of diabetes, which is important to the prevention of the onset of complications.

2. Diabetes health indicators dataset

The US Diabetes Health Indicators Dataset, originally collected by BRFSS, is cleaned and published by Alex Teboul on Kaggle [5]. The dataset is a collection of 23580 records of adult respondents in the United States. It has 22 variables in total, including 21 feature variables and 1 label variable indicating diabetes or not [5]. The binary target variable takes value "1", indicating a positive result for diabetic or prediabetic, while "0" indicates a negative result for non-diabetic. Figure 1 shows the unbalanced distribution of diabetes cases in the dataset (86.1% for non-diabetic, 13.9% for diabetic and prediabetic). The 21 indicators are indicated by numeric and categorical variables respectively. The only numeric variable is the BMI index which has an interpretation of BMI status. As seen in Figure 2, BMI below 18.5 means underweight, 18.5–24.9 means healthy weight, 25.0–29.9 means overweight, and 30.0 and above means obesity.

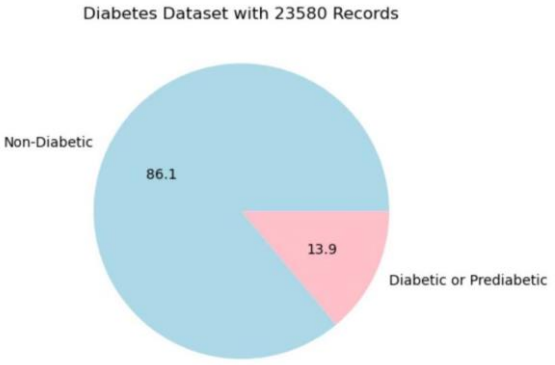


Figure 1. Distribution of Diabetes Dataset.

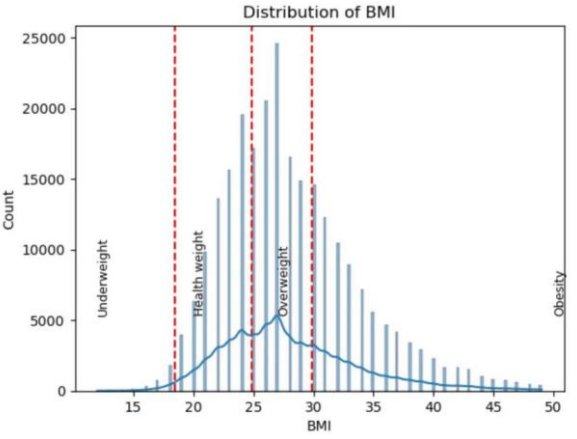


Figure 2. Distribution of BMI in Dataset.

The other 20 categorical features include 4 aspects. The features related to personal information are age, gender, education level, income, having health care service or not, and having financial difficulties to see a doctor or not. The binary features of physical disease denote high blood pressure, high cholesterol, cholesterol check, stroke, heart disease attack, and whether the patients have difficulty walking or not. The features that show the respondents' self-assessment of their health status are general health, mental health, and physical health. The binary features of personal habits are physical activity, smoking, fruits, veggies, and heavy alcohol consumption.

3. Method

3.1. Feature selection

Feature selection is a crucial step for removing irrelevant features and picking a subset of highly discriminant features for the target variable from the original dataset. The advantages of feature selection are a reduction in the execution time of the classifier and an improvement in model accuracy.

DOI: 10.54254/2755-2721/32/20230214

3.1.1. Feature correlation. Before selecting the feature, the correlation diagram (Figure 3) was plotted to check the correlation between the target variable "Diabetic" and 21 feature variables. The features of health issues and diseases, such as general health, high BP, walking difficulty, high BMI, high Cholesterol, and heart disease or attack, are highly and positively correlated with diabetes. The features denoted that personal habits have much less correlation with having diabetes. Among the features of personal information, education level and income are more correlated with diabetes in a negative way. Overall, the correlation diagram illustrates some indicators of influence in predicting diabetes, which is helpful feature selection.

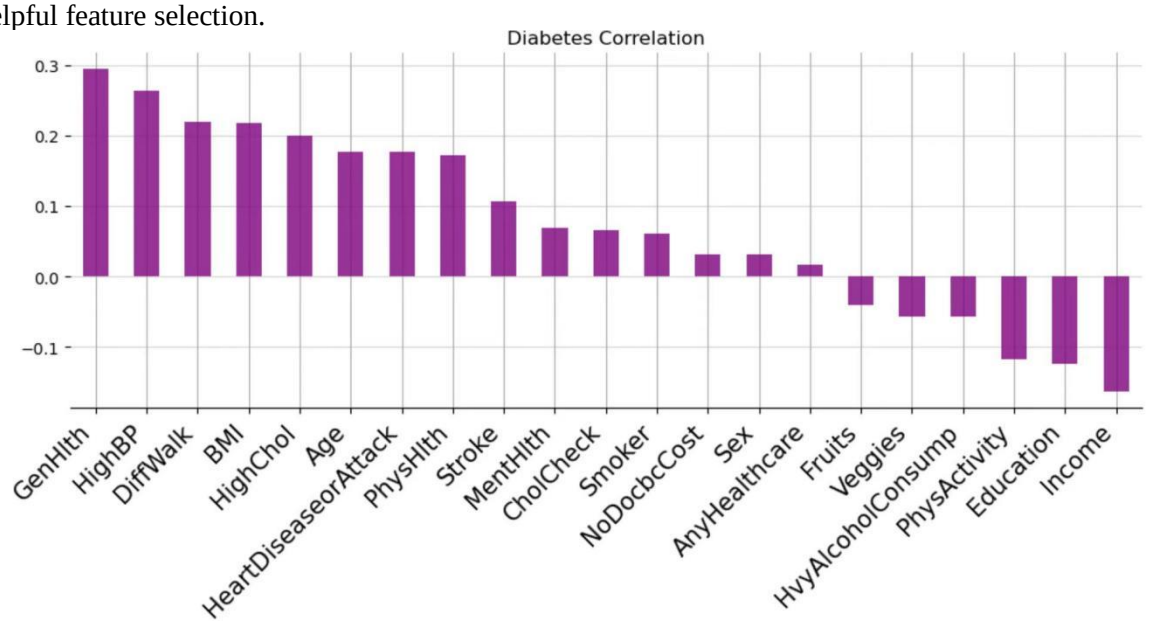


Figure 3. Correlation Diagram.

3.1.2. Chi-square test. It is not enough to select features only based on the correlation diagram. For further feature analysis, the Chi-square test was applied to this project, which helps to solve the problem in feature selection by testing the relationship between the feature variable and the target variable [6]. A Chi-square test is used in this project to test the independence of two features.

The formula for calculating a Chi-square statistic is [7]:

$$x^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \quad (1)$$

Where O denotes the observed values, and E denotes the expected values.

By calculating the Chi-square value, the relationship between the independent category feature and the dependent category feature can be determined. In feature selection, the goal is to select the features with a high Chi-square value, which are highly dependent on the target "Diabetic". The Chi-square score of 21 feature variables is calculated, and Table 1 displays the best 15 features to be selected and their Chi-square scores. The other six features will be removed from the dataset due to their low Chi-square scores; they will not participate in the training of the models. The dropped features also show a lower correlation in the previous correlation diagram, which is expected.

DOI: 10.54254/2755-2721/32/20230214

Table 1. Feature with Chi-square Score.

Feature	Chi-square Score
PhysHlth	133424.41
MentHlth	21029.63
BMI	18355.17
DiffWalk	10059.51
HighBP	10029.01
GenHlth	9938.51
Age	9276.14
HeartDiseaseAttack	7221.98
HighChol	5859.71
Income	4829.82
Stroke	2725.23
PhysActivity	861.89
HvyAlcoholConsump	779.42
Education	756.04
Smoker	521.98

3.2. Data preprocessing

3.2.1. Data splitting. Before starting to train the model, the data set needs to be preprocessed. The dataset is cleaned without any missing values. And randomly shuffling the dataset ensures the random distribution of the data. In machine learning, if the data set is not shuffled, the "bias" of the model might occur in the training process, which reduces its generalisation ability, thereby reducing the training accuracy. For example, if a classification model was built in which the first 80% of the initial data is the first class and the last 20% is the second class, the accuracy of the model will be extremely low. The low correlation feature variables in Section 2 dropped, and the best 15 features remained as the final feature variables in the model training and testing. The data was divided into target variable data as 'X' and 15 feature variable data as "y". The portion of the dataset to allocate to the test set is 20%, and the counterpart of the train set is 80%.

3.2.2. SMOTE algorithm. Recall Section 2, the number of non-diabetic records is much greater than that of diabetic records. The imbalance of the training dataset might lead to bias and the poor performance of model training. Therefore, before building the model, the imbalance of the dataset needs to be resolved. The imbalance can be handled by undersampling and oversampling strategies. The undersampling is to select data randomly from the majority class until two classes have the same number of records [8]. However, the reduction of the majority class might lose useful information. Thus the SMOTE (Synthetic Minority Oversampling Technique) preprocessing algorithm was used to balance the training data. The basis of SMOTE was to generate synthetic samples for the minority class until the data is balanced [8]. Moreover, it can assist the classifiers to improve their generalisation capacity. Figure 4 displays the rebalance of diabetes distribution before and after oversampling.

DOI: 10.54254/2755-2721/32/20230214

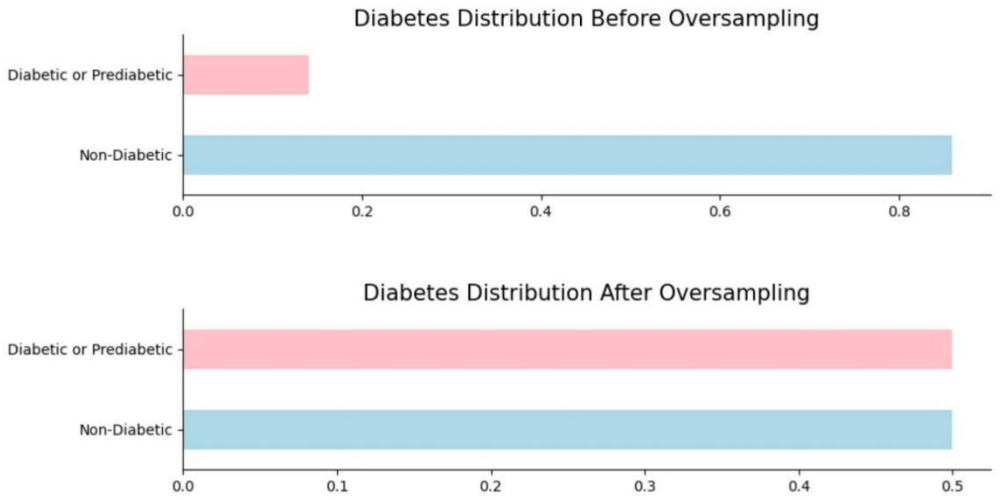


Figure 4. Diabetes Distribution Before and After Oversampling.

3.2.3. Feature scaling. The final step for data preprocessing is feature scaling which can transform features of a dataset to improve the performance of machine learning models and reduce the time for training models. If the original index value is directly used to train models, the features with a large value will be emphasised, and the features with a small value will be weakened. The feature scaling ensures that all features contribute equally to the model and prevents the domination of features with larger values [9]. The feature scaling method used in this project is standardisation scaling.

3.3. Models and evaluation

3.3.1. Model training and results. Several supervised machine learning classifiers have been applied to predict diabetes with the processed training set and testing set, including Gaussian Naive Bayes, Decision Tree, Random Forest, Logistic Regression, Gradient Boosting, Linear Discriminant, and Cat Boost classifiers. The accuracy, precision, and mean square error measures are applied for analysing the performance of the models above. In order to calculate accuracy and precision for training and testing data, there are four cases that need to be considered [10].

True positive (TP): record is classified as positive and is actually positive.
False positive (FP): record is classified as positive and is actually negative.
True negative (TN): record is classified as negative and is actually negative.
False negative (FN): record is classified as negative and is actually positive.
The accuracy is the proportion of the total number of predictions that were correct:

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{FP} + \text{TN} + \text{FN}) \quad (2)$$

The precision is the proportion of positive records that were correctly predicted as positive:

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \quad (3)$$

The four classifiers with an accuracy of more than 80% are the Cat Boost Classifier, Random Forest Classifier, Gradient Boosting Classifier, and Decision Tree Classifier. As seen in Table 2, the training accuracy of the Decision Tree and Random Forest Classifier is close to 100%, but the accuracy of the two classifiers decreases for the test set, and the precision is very low. Therefore, these two models may be overfitting. Overfitting refers to matching the data of the training set too closely and precisely so that it cannot fit the data of the testing set well. The Cat Boost Classifier has the highest accuracy (86.6%) and highest precision (55.6%). Thus the selected model is the Cat Boost Classifier.

DOI: 10.54254/2755-2721/32/20230214

Table 2. Summary of Models.

Model	Train Accuracy	Train Precision	Test Accuracy	Test Precision
Cat Boost	0.925	0.978	0.866	0.556
Random Forest	0.990	0.994	0.849	0.420
Gradient Boosting	0.886	0.901	0.838	0.421
Decision Tree	0.990	0.997	0.802	0.298
Gaussian Naïve Bayes	0.717	0.731	0.741	0.308
Logistic Regression	0.752	0.740	0.729	0.307
Linear Discriminant	0.752	0.734	0.721	0.302

3.3.2. Permutation feature importance. Permutation Feature Importance (PFI) is a method applied to calculate the importance of features independent of the model type. As a result of the Cat Boost classification model, feature importance measures how much each feature affects the target. The PFI was helpful to know which features are more important in the selected model and which features affect the prediction results, thus interpreting the performance of the model. The following steps are applied to calculate the feature importance score [11]:

1. Use the selected model and record the original score of the model.
2. Shuffle the value of a feature, use the model to predict again, and calculate the score on the test set. The reduction of model performance represents the importance of this feature.
3. Restore the value of the disrupted feature, and repeat step 2 on the next feature until the importance of each feature is obtained.

Accuracy was selected as a representation of model performance, which is the "score" for permutation. In that case, the features were sorted from high to low based on the decrease in accuracy, and the top feature has the highest importance. For each feature variable, the reduction in accuracy was to calculate and record for 30 random shuffles [11]. To visualize these records, boxplots (Figure 5 and Figure 6) are created for the train set and test set respectively.

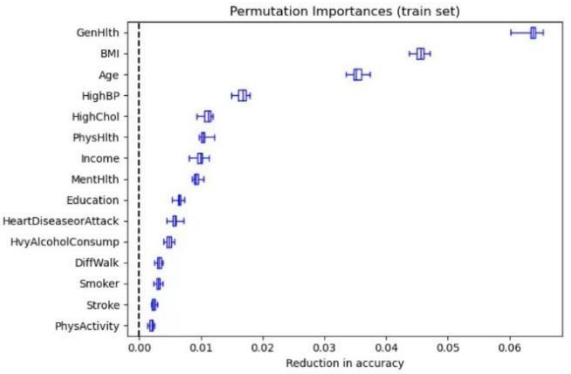


Figure 5. PFI for Train Set.

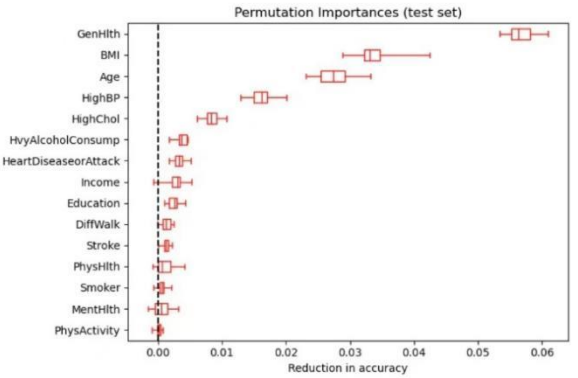


Figure 6. PFI for Test Set.

According to the boxplot of the permutation importance feature for the Cat Boost model, the most important feature was General Health (GenHlth), followed by BMI, Age, high blood pressure (HighBP), and high cholesterol (HighChol) variables. The permutation feature importance in the train set and test set is very similar.

DOI: 10.54254/2755-2721/32/20230214

4. Conclusion

This paper applies a dataset that represents the distribution of diabetes disease from BRFSS in 2015. The goal of this paper is to find the features that are correlated with diabetes and train a predictive model with a reasonable level of accuracy. The correlation diagram and the Chi-square test have been applied to select the best 15 features from 21 features in total. After the data preprocessing and model training, the results suggest that the Cat Boost classifier is suitable for predicting diabetes with the selected features. However, there are still some limitations existing in the model. The precision of the model is not as expected. One of the possible reasons is that the data set is unbalanced. The SMOTE algorithm was implemented in the training set to rebalance the data, but the testing set is still unbalanced (it is incorrect to use SMOTE for the full set, which affects the purity of the data set and leads to overfitting). The imbalance of the test set may cause the model to be less precise on the test set. In the future, the performance of the Cat Boost Classifier can be improved by collecting more data on diabetic patients.

References

[1] Kumari, V. A. and Chitra, R. (2013). Classification of Diabetes Disease Using Support Vector Machine. International Journal of Engineering Research and Applications (IJERA), 3, 1797-1801.

[2] U.S. Department of Health and Human Services Centers for Disease Control and Prevention. (2020). National Diabetes Statistics Report Estimates of Diabetes and Its Burden in the United States. https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf.

[3] Fang, M., Wang, D., Coresh, J. and Selvin, E. (2021). Trends in Diabetes Treatment and Control in U.S. Adults, 1999-2018. The New England journal of medicine, 384(23), 2219–2228.

[4] National Center for Chronic Disease Prevention and Health Promotion. Division of Population Health. https://www.cdc.gov/brfss/index.html.

[5] Teboul, A. (2021). Diabetes Health Indicators Dataset. https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-datase t/code.

[6] Bahassine, S., Madani, A., Al-Sarem, M. and Kissi, M. (2020). Feature selection using an improved Chi-square for Arabic text classification. Journal of King Saud University - Computer and Information Sciences Volume, 32(2), 225-231.

[7] Kumar Gajawada, S. (2019). Chi-Square Test for Feature Selection in Machine learning. Published in Towards Data Science. https://j-pcs.org/temp/JPractCardiovascSci1169-9537648_023857.pdf.

[8] Satwik, M. (2017). Handling Imbalanced Data: SMOTE vs. Random Undersampling. International Research Journal of Engineering and Technology (IRJET), 4(8), 317-320.

[9] Hanan, A., Yuan, X. H., Esterline, A., Khorsandroo, S. and Lu, X. C. (2021). Studying the Effects of Feature Scaling in Machine Learning. Ph.D. Dissertation. North Carolina Agricultural and Technical State University. Advisor(s) Xu, Jinsheng. Order Number: AAI28772109.

[10] Gürsoy, M. İ. and Alkan, A. (2022). Investigation Of Diabetes Data with Permutation Feature Importance Based Deep Learning Methods. Karadeniz Fen Bilimleri Dergisi. The Black Sea Journal of Sciences. ISSN (Online): 2564-7377.

[11] Li, S. (2022). Best Practice to Calculate and Interpret Model Feature Importance: With an example of Random Forest model. Published in Towards Data Science. https://towardsdatascience.com/best-practice-to-calculate-and-interpret-model-feature-importance-14f0e11ee660.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230215

Study on Fatigue Driving Detection based on Physiological Characteristics of Drivers

Mingjun Jiang[1,3], Xinran Yang[2]

[1]School of Computer and Information Engineering, Ren'ai College of Tianjin University, Kunming, China
[2]Jiangxi Huis High School, Nanchang, China

[3]1911432137@mail.sit.edu.cn

Abstract. The 21st century is an information age. With the rapidly developing economy, transportation has also been equally affected. The number of people's travel has also increased. With the popularity of transportation, people's travel has become more convenient. According to the data, the number of deaths from traffic accidents still exceeds 60,000 and injuries exceed 250,000. At present, the three main causes of traffic accidents are fatigue driving, speeding driving and drunk driving. Driver fatigue status is crucial for safe driving. Fatigue may reduce the alertness, affect the reaction time and increase the risk of accidents. The earliest research on fatigue detection in the world was carried out using high-precision medical equipment. Researchers analyzed the changes of EEG signals and ECG signals to determine whether people were in a fatigue state. The acquisition of physiological signals needs to paste the sensor in the human skin, which then interferes with the driver's operation. With the advent of the Internet era, the promotion of artificial intelligence technology, and the rise of image processing technology and machine learning has further promoted the development of fatigue driving detection. Therefore, the research on the detection and evaluation of fatigue driving is particularly important, and the research on fatigue driving detection technology provides a guarantee for road safety. This paper will analyze, and study based on the four aspects of vehicle driving mode, driver physiological characteristics, driver behavior characteristics and driver facial characteristics detection, with the advantages of convenient operation and low cost.

Keywords: Traffic Safety, Fatigue Driving, Physiological, Behavior, Facial Characteristics.

1. Introduction

While providing convenience for people's life, work and travel, the automobile also causes serious personal injury and huge economic and property losses. There are many causes of traffic accidents, among which the traffic accidents caused by fatigue driving is 12%, which is a major factor affecting driving safety. Most of the accidents are largely related to the driver's lack of sleep the previous dad [1]. If the driver in the day before the sleep time to seven hours, is the initiative of the probability of traffic accident will increase greatly, and the day before the driving sleep time not to four hours of the driver, the probability of traffic accident is the driver meet the normal sleep time 15 times, therefore, fatigue driving is likely to lead to car accident. The main cause of fatigue driving is the reduction in the perceptual sensitivity and the ability to identify the distance when the driver is in a state of

DOI: 10.54254/2755-2721/32/20230215

exhaustion. At the same time, their reactions are becoming dulled, thinking and judgment is slow, the reaction time will be extended by about 0.2s, when stimulated by complex situations, the reaction time will double [1]. The characteristics of tired driving can be summarized as: frequent blinking, yawning and nodding, body tilt, slow reaction, etc. These fatigue characteristics may all be important factors in causing major or major traffic accidents. Therefore, fatigue driving detection has become an important research topic in the field of traffic safety, which has a very long-term significance to People's Daily life and the economic development of the country. In this paper, fatigue detection based on driver behavior characteristics is an indirect detection method, which mainly uses head sensors and machine vision technology to detect the changes of the driver's head and face, including eye closure ratio, yawning frequency, nodding frequency and line of sight direction deviation angle.Another is that the fatigue detection based on the physiological characteristics of the driver, which uses the sensor in direct contact with the driver's body to obtain physiological parameters to distinguish the fatigue or not. The methods include EEG signal detection [2], ECG signal detection, eye telecom signal detection and heart rate variability detection [3]. By testing whether the driver is tired in the driving process, so as to reduce the fatigue driving behavior in the driving process, so as to improve the driving safety and ensure safe driving. Reduce a variety of car accidents caused by fatigue driving in China. Our detection based on driver physiological characteristics could use the frequency change of EEG images to visually reflect the driver's sleepy state, use ECG to determine whether physiological burden is abnormal, and collect biological signals of muscle at rest or contraction through EGM technology [4]. Driver face features are detected by face detection by YOLOX [1] and face keypoint detection and head posture by PFLD [5]. Finally, the relevant indicators are used to judge whether the driver is tired.

2. Fatigue driving detection study

2.1. Detection based on the driver's physiological characteristics

Through the physiological data monitoring of the driver to judge the driving state mainly through the analysis of the physiological state of the driver is awake, has a good driving state. The detection of physiological data when the driver fatigue state will deviate from the normal data track.

2.1.1. Electroencephalography. The data monitoring of EEG is closely related to the sleep structure. The frequency change of the EEG image can directly reflect the sleepy state of the driver, so as to judge the fatigue degree of the driver. According to the basis of EEG waveform test fatigue: when the cerebral cortex is in different states, the EEG performance is different: δ wave, frequency 1 ~ 3.5 H z; θ wave, frequency 3.5 ~ 7.5 H z; α wave, frequency 7.5 ~ 1 2.5 H z; β wave, frequency 1 2.5 ~ 3 0 H z [2]. The is found that among the four ripple attributes, α wave is the most able to reflect the driver's desire to sleep. When the brain needs to sleep, the α wave frequency increases significantly and the β wave frequency decreases significantly. However, because the EEG method requires professional instruments and the brain signals are susceptible to external information, the test conditions are harsh, and the price of individual differentiation is too high, the practicability of the EEG method is limited and cannot be put into the market.

2.1.2. Electrocardiogram. ECG is mainly used to detect the physiological burden of the driver. When the physiological burden of the driver is too large, the ECG will send out a striking and regular downward trend signal, indicating that the driver is in a state of fatigue. The most important physiological indicators to determine whether the physiological burden is abnormal are heart rate and heart rate variability (HRV), respectively. Through some simulated driving experiment, the fatigue degree can be quantified by four ECG time and frequency domain indexes [3].

The advantages and disadvantages of ECG method are obvious, but the equipment of this method is easy and the system is easy to cut. This technology is expected to carry out dynamic detection and real-time processing, but the disadvantages are that the sensitivity and analysis and judgment ability of

DOI: 10.54254/2755-2721/32/20230215

the hormone are poor, and it needs to cooperate with other technologies to make up for the technical loopholes that cannot be broken in this technology.

2.1.3. Electromyography. EGM technology collects biological signals from muscles at rest or during contraction to determine the functional status of peripheral nerves, neurons, neuromuscular joints and the muscle itself. With the emergence of physiological fatigue with the deepening of fatigue EMG data will gradually decline. Wang Feng et al designed a front-end surface myographic signal (SEMG), which has the advantages of high input resistance [4], low noise and strong interference resistance. This technique uses wavelet analysis to determine the characteristics of SEMG during muscle fatigue, which is more flexible than Borleaf transformation.

This technique is very suitable for dynamic signal processing and can accurately and efficiently extract the nonstationarity of signals. It is the most complete detection method to detect the driver's driving state at present.

2.1.4. Test method. EEG, ECG, and EMG all use the same test sequence. The sequence is divided into four parts: 1) Collect information: collect information about heartbeat, brain signals and muscle expansion. This step is the only one to do in the cockpit, and the other three steps are done on a cloud processor. 2) paramete rcalculation:electrocardiogram, electroencephalogram and EMG cannot be used directly as calculation data, because generally the machine is a data comparison (graphics is too complex for data, not convenient for accurate comparison), the step usually use specific formula to collect the graphic information into digital data. 3) Data preprocessing: Use specific formulas to calculate the data of graphic information. 4) Parameter comparison: submit the data from the first three steps to the cloud processor for comparison with the expected threshold data. If the expected threshold is exceeded, it will be judged as abnormal or otherwise, and the results will be sent to the cloud server [6].

After sending to the cloud server, the server will not directly convey the results to the driver, because the cloud service will simultaneously receive three information about the comparison of the EEG data, ECG data, and EMG data with the expected threshold. When two or more items of data are greater than the expected threshold, when the result given by the cloud server in the fourth step of the detection method sequence is abnormal, the cloud server will determine the driver status as fatigue driving, and the cloud server will send the results to the driver in the first time.

2.2. Detection based on driver facial features

2.2.1. Face detection based on YOLOX. Face detection technology is mainly divided into two categories. One is the traditional computer vision algorithm, which primarily extracts features from the input image manually, and then trains the detector with the extracted features, so as to use the detector to complete face detection. The other type is deep learning technology. On the one hand, target detection network suitable for various tasks can detect faces, such as Faster-RCNN and YOLO series algorithm; on the other hand, face detection, such as CascadeCNN and MTCNN. The detection accuracy of such method is slightly higher than that of YOLO series, but the detection speed cannot meet the requirements of real-time system. Therefore, considering the demand of fatigue driving detection for face detection in this paper, the target detection network YOLOX is selected to conduct the driver's face detection. The prediction network module YOLOX Head can output the prediction box of the detected target, and its structure is shown in Figure 1.

DOI: 10.54254/2755-2721/32/20230215

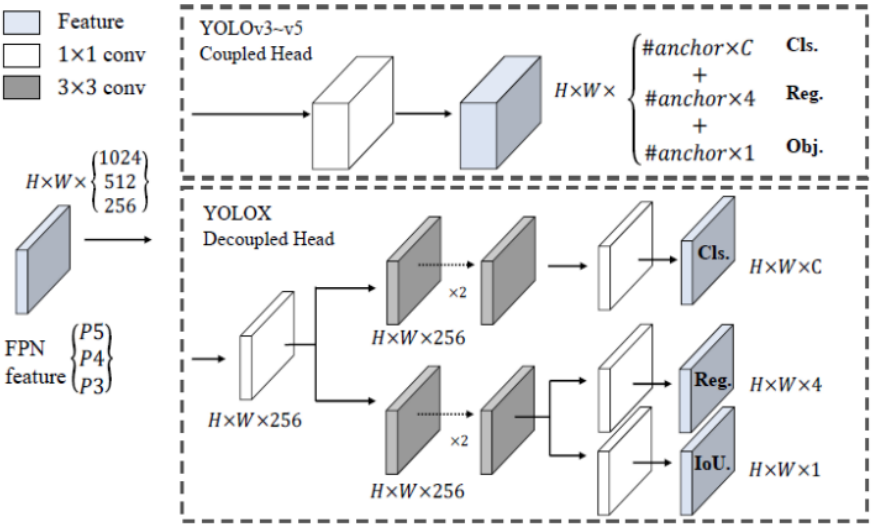


Figure 1. YOLOX Head, Network structure[1]

YOLOX use of Decoupled Head (predicted branch decoupling) [4] not only greatly improves the convergence rate, but also brings 4.2%AP improvement over the end-to-end decoupling of YOLOv3 to v 5. Decoupled Head, 1 * 1 convolution was used to first unify the features of originally different channel to 256 for dimension reduction, and then two parallel branches were used for classification and regression. The FPN feature pyramid is able of output three enhanced features of different shape, which are then input into the YOLOX Head prediction network to obtain the prediction results. Finally, three predictions can be obtained for each feature layer, which are: Reg (h, w, 4): the regression parameters used to judge each feature point, which can be obtained after adjustment Forecast box; (2) Obj (h, w, 1): used to judge whether each feature point contains an object; (3) Cls (h, w, num_classes): used to judge the type of objects included in each feature point. Finally, the three predictions are stacked to obtain the prediction results of the detected target.

2.2.2. Face keypoint detection and head pose estimation based on PFLD: 1) Face keypoint detection and head posture estimation: In the recognition of faces, we need to detect the key points of the face through the key points in the figure below to judge whether the person is tired in the process of driving. In this paper, 98 key points are used to identify the face diagram, as shown in Figure 2, to outline the facial features such as eyebrows, eyes, nose, mouth and face in a certain order.

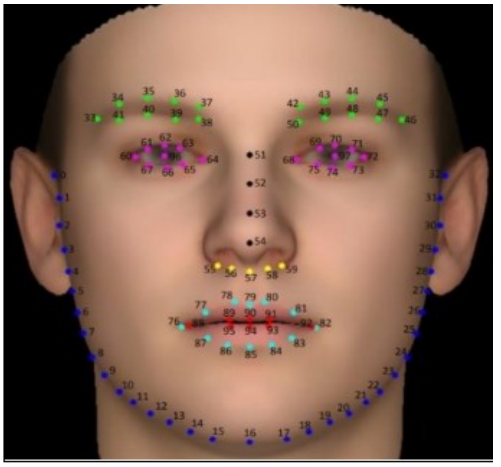


Figure 2. A Schematic diagram of the 98 key points identifying the human face[7]

DOI: 10.54254/2755-2721/32/20230215

Head posture estimation is to judge the driver in the process of driving, if fatigue, will cause a certain head Angle deviation, or the opening of the eyelid reduced. The above figure is the 2D model constructed. We can build a 3D model to better judge the degree of increase and decrease of various indicators of drivers. In 3D space, the head posture can be represented by three Euler angles: pitch (pitch), yaw Angle (yaw), and roll Angle (roll). The elevation angle represents the de-rotation using the x-axis, the yaw angle indicates the rotation of the y-axis, and the rolling angle represents the rotation of the z-axis as shown in Figure 3. With the central point of the head as the origin of the spatial coordinate system, pitch refers to the rotation around the x axis, indicating the amplitude of the nod up and down; yaw refers to the rotation around the y axis, indicating the amplitude of the left and right head, and roll refers to the rotation around the Z axis, indicating the amplitude of the left and right head.

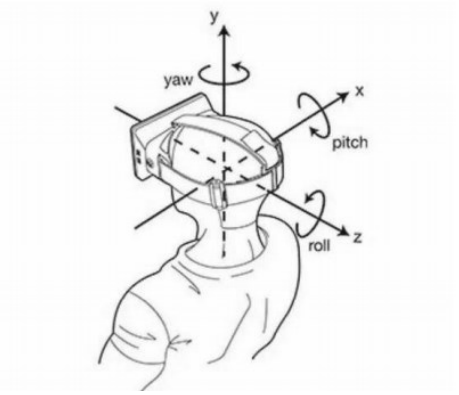


Figure 3. Schematic diagram of Euler angles[8]

2) PFLD, and the network structure:In the real driving environment, the key point detection faces complex lighting conditions in the driving process, and the face may be blocked by the driver wearing a mask. During the driving process, the driver may change positions in real time, and may be affected by the shooting equipment and lighting environment, resulting in different image quality.

The above situations will cause certain interference to the accuracy and speed of key point detection. In view of the difficulties of these face key point detection, Guo Xiaojie et al. proposed PFLD algorithm [9]. The algorithm is a practical algorithm with fast running speed and high accuracy, which breaks through the speed and accuracy limit of the previous face key point detection algorithm. The algorithm is not only fast, but also achieves the best evaluation index on the same data set, that is, maintains the high accuracy of the face key point positioning, and greatly reduces the complexity of the model. Therefore, the PFLD model was used to detect face keypoints. The network structure of the PFLD model consists of two parts: predicting landmark backbone network and the head pose auxiliary network. The Figure 4 showed the schematic diagram of Euler angles.

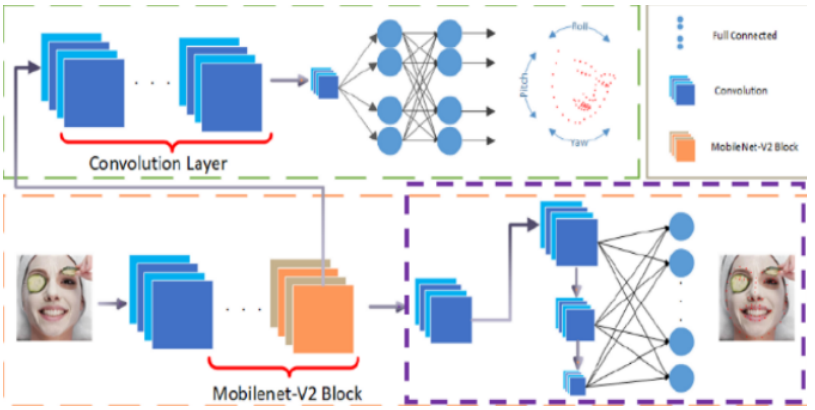


Figure 4. Schematic diagram of Euler angles[5]

DOI: 10.54254/2755-2721/32/20230215

The structure-optimized MobileNet-V2 lightweight network is adopted as the backbone network of the PFLD model [9]. The network is used to locate the position coordinates of facial key points. Due to its unique network structure, the quantity of parameters and calculations of the model are greatly reduced, and the execution speed of the model is improved.head pose The auxiliary network is a branch of the backbone network for head pose prediction during training to improve the positioning accuracy of key points. By default, the secondary network is not used during testing. The purpose of this is to adjust the loss parameters according to the acquired head attitude angle during the training process, so that the model pays more attention to rare samples and samples with too large attitude angle, and the predicted key point position coordinates are more stable and more robustness.

The PFLD model optimizes the network structure of Mobilenet-V2, using multiscale fusion to enhance the model expression capacity. The multi-scale full connection layer can increase the receptive field and can also better obtain the overall structure of the face, and then improve the positioning accuracy of the network, so as to obtain more accurate key point coordinates of the face.

The Loss function can calculate a value obtained after each training, which is the value between the real value and the predicted value, so as to ensure that the next training time can get a smaller error. In addition to network model structure and the quality of the selected samples that can determine the model training, the loss function also plays a crucial role.

The general loss function is difficult to deal with data imbalance, such as L2-loss loss function:

$$\mathrm{L}:=\frac{1}{\mathrm{M}}\sum_{\mathrm{m}=1}^{\mathrm{M}}\sum_{\mathrm{n}=1}^{\mathrm{N}}\gamma_{\mathrm{n}}||\mathrm{d}_{\mathrm{n}}^{\mathrm{m}}|| \tag{1}$$

Where M is the number of samples, and N is the number of feature points, representing the weight values of different key points. The loss function of the PFLD algorithm takes into account that the number of samples of different categories in the training sample may differ greatly, and the head pose angle obtained in the branch is applied to the loss penalty, so that the rare sample is given higher weight for further refinement. The optimized loss function;

$$\mathrm{L}:=\frac{1}{\mathrm{M}}\sum_{\mathrm{m}=1}^{\mathrm{M}}\sum_{\mathrm{n}=1}^{\mathrm{N}}(\sum_{c=1}^{C}\omega_n^c\sum_{k=1}^{K}(1-\cos\theta_n^k))||d_n^m||_2^2 \tag{2}$$

among, $\sum_{c=1}^{C}\omega_n^c\sum_{k=1}^{K}(1-\cos\theta_n^k$for the final sample weight, C represents different categories of faces. In the PFLD paper, faces are divided into multiple categories, including positive face, side face, tilted head and exaggerated expression. K represents the deviation of the three Euler angle values of the head pose estimation, and w is the given weight corresponding to the category. The loss function can solve the problem of unbalanced training samples, adjust the training weight of different categories of samples, and give a small weight to the samples with large data volume (such as positive face, samples with small θ value), thus reducing the contribution of samples with large sample size to the model training; for rare samples (such as side face, lower face, head, local occlusion, etc.), so as to improve the contribution in the backpropagation of the gradient, so as to solve the imbalance of training samples.

2.2.3. Extraction of fatigue features and status determination. a. Judgment of eye fatigue: a. PERCLOS is the physical quantity of fatigue/sleepiness proposed by the Carnegie Mellon Institute after repeated experiments and arguments [10]. PERCLOS It is recognized as the most effective and widely used indicator of fatigue. It has three classification criteria:

1) P70: When the upper eyelid covers more than 70% of the pupil area, it is regarded as closed eyes, and then the proportion of closed eyes time per unit time is calculated;

2) P80: When the upper eyelid covers more than 80% of the pupil area, it is regarded as closed eyes, and then the proportion of closed eyes time per unit time is calculated;

3) EM: When the upper eyelid covers more than half of the pupil area, it is considered as a closed eye, and then the unit is calculated Proportion of time.

The ratio of eye closure time to unit time calculated according to the P80 standard is the PERCLOS value per unit time, and the principle is shown in Figure 5.

DOI: 10.54254/2755-2721/32/20230215

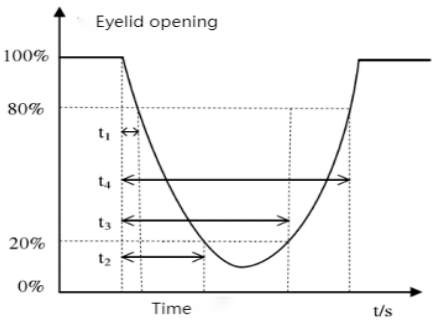


Figure 5. The P80 standard principle in PERCLOS[4]

The formula for calculating the PERCLOS value is shown in formula 3:

$$f = \frac{t_3 - t_2}{t_4 - t_1} \times 100\% \tag{3}$$

b. Blink frequency

Blink frequency refers to the number of blinks per unit time, and the driver has a certain blink frequency in the awake state Small amplitude fluctuations in the range of. If the driver feels tired and conscious, his eyes will feel dry, and alleviate dryness by frequent blinking, the corresponding blinking frequency parameter becomes larger; when the fatigue gradually increases, the blink time will become longer, and the driver will suddenly open his eyes at a moment of fatigue, and then continue to blink after a short period of time, changing the blink frequency between brief awareness and blur until closed. Therefore, the ratio of the number of blinks to the statistical time. Blink frequency is like:

$$f = \frac{N_B}{N_p} \tag{4}$$

Characteristics of mouth fatigue: During driving, the driver will have his mouth close, talk or yawn. The driver is tired During tired and frequent yawning, the opening and duration of the mouth will be significantly different from that of awake, so the aspect ratio (Mouth Aspect Ratio, MAR) can be calculated to determine the opening of the mouth, so as to distinguish the three states according to the threshold value. The key point distribution of the extracted mouth area of the face is shown in Figure 6.

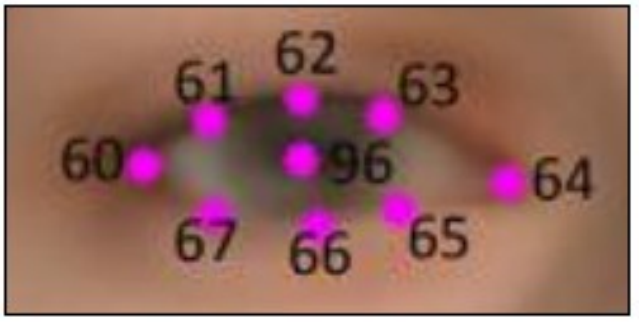


Figure 6. The distribution of key points in the face and mouth[4]

The MAR is calculated as in Eq:

$$EAR = \frac{|y_{61} - y_{67}| + |y_{63} - y_{65}|}{2|x_{60} - x_{64}|} \tag{5}$$

DOI: 10.54254/2755-2721/32/20230215

2.2.4. Characteristics of head fatigue: a. Nodding frequency: The nodding frequency refers to the frequency of nodding per unit time, which is one of the head characteristics of the driver. The nodding frequency was performed as indicated:

$$f = \frac{N_N}{N_T} \tag{6}$$

This is the number of nodding movements in time T and the total number of frames collected in time T.

b. Head posture and Euler angle: Head posture euler Angle is another can be used to identify the driver fatigue head characteristics, mainly by patch, yaw and roll three head posture Angle to reflect the change of the driver head movement, the parameter value can be used to calculate the nod frequency and head abnormal posture, which can assist in determining the driver's state of fatigue. The calculation formula of the head abnormal posture index is shown in the formula:

$$f = \frac{N_P}{N_T} \tag{7}$$

This is the number of frames judged as abnormal attitude in the statistical time T, and the total number of frames collected during time period T.

3. Discussion

Assessing driver fatigue through physiological signals can yield relatively accurate and objective results that more accurately reflect the driver's true state. These devices collect data continuously, allowing real-time monitoring of the driver's condition and thus early detection of signs of fatigue. However, physiological signals can be affected by many factors, including individual differences, environmental factors, and emotional states. The use of EEG, ECG and EMG devices may require drivers to wear special equipment or attach sensors to their bodies, which may cause discomfort and may even interfere with normal driving behaviour. Moreover, such equipment can be quite expensive and also requires regular maintenance and calibration to ensure accuracy and reliability.

PFLD-based detection of face keypoints can run quickly on a variety of devices to adapt to real-time needs, PFLD can detect many keypoints of the face to deal with the detection of keypoints of the face in multiple viewpoints, and the detection of the driver's eyes, mouth, and head fatigue features can also improve driving safety and help prevent traffic accidents. However, PFLD requires a large amount of data with accurate labelling for training, and data acquisition and labelling is a difficult and time-consuming task. Under changing lighting conditions, PFLD may suffer from a decrease in detection accuracy. This may affect the accurate detection of driver fatigue.

A feasible solution for detecting fatigue driving is derived based on driver physiological characteristics and PFLD-based methods for estimating face keypoints and head pose. Firstly, a large amount of physiological data and facial images of drivers during driving need to be collected. These data are also labelled to identify fatigue and non-fatigue states. Then use PFLD for face key point detection and head pose estimation method. PFLD can effectively identify the driver's key points such as eyes and mouth, as well as head pose. This information will be used as a key input for detecting the driver's fatigue state. These features are correlated with the labelled fatigue state and trained to obtain a model capable of recognising fatigued driving. In practical applications, real-time collected driver physiological data and facial images are input into the model to detect the driver's fatigue state in real time. If the driver is detected to be in a fatigue state, can take appropriate measures.

4. Conclusion

Road traffic safety and driver's life safety guarantee the possibility of fatigue driving detection technology, which has important social significance. Although the technology has emerged in recent years due to economic development and academic breakthroughs, this technology still needs to be improved. 1) The fatigue driving testing standards need to be further clarified. Fatigue driving perception includes cross-integration of biological, biochemical, behavioral science, ergonomics, and

DOI: 10.54254/2755-2721/32/20230215

other fields. Existing detection methods cannot accurately classify fatigue levels and judge the relationship between fatigue and detection indicators, so this level and relationship need to be quantified in future studies. Through further study of fatigue characteristics and characteristics, construct the perception quality of fatigue driving. 2) Information fusion is one of the most important directions of fatigue detection. Due to the individual differences and complexity of physiological factors is difficult to estimate the driver fatigue, so there is only one standard and evaluation system, namely based on the recruitment information convergence measurement method can complement different sensor methods under the premise of comprehensive analysis of fatigue, artificial intelligence, digital image processing and mobile communication technology more effectively improve the detection accuracy and reliability. 3) Fatigue detection technology based on computer vision will continue to be further studied. Although fatigue driving monitoring devices are close to commercial, the high cost will hinder the wide application of inspection equipment, and computer desktop technology is expected to provide simple, cheap, comfortable, reliable wide range of market inspection products.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Wang S, Research on fatigue driving detection technology based on facial features. 2022 Shenyang University of Technology 64.
[2] Xu L. Design of a vehicle-mounted intelligent cockpit controller system. 2023 Times Automobile 114-116.
[3] Zhang R, Zhu T, Zhou Z. Review on the methods of driver fatigue driving detection. 2022 Computer Engineering and Applications. 58(21) 53-66.
[4] Lv G, Software architecture design and development of intelligent cockpit controller. 2022 Shandong University 98.
[5] Jing F, Research and implementation of driving fatigue detection based on deep learning. 2021 University of Electronic Science and Technology of China 83.
[6] Bo C, Xu G,Song P. A. Review of the methods of fatigue driving detection. 2013 J. Dalian Nationalities University 15 266-271.
[7] Jing B, Research on design and control strategy of vehicle intelligent cockpit controller. 2022 Henan University of Science and Technology 85.
[8] Wang T, Research on the status quo and development trend of automotive intelligent cockpit design. 2021 Times Automobile 23 158-159.
[9] Zhang M, Research and implementation of driving fatigue detection based on deep learning. 2020 Southwest University 72.
[10] Wang D, Research on the status quo of domestic driver fatigue driving detection methods. 2018 Think Tank Era 41 118-119.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230216

Research on performance comparison of patrol path planning techniques for mobile robots in nuclear power plants

Shuying Liu

Maynooth International Engineering College, Fuzhou University, Fuzhou City, Fujian Province, 350000, China

shuying.liu.2021@mumail.ie

Abstract. Path planning, as a basic problem of mobile robots, is important in the application of industrial patrol robots. This paper takes the nuclear power plant as an example and solves the problem of multi-target patrol path of patrol robot. Firstly, this paper processes multi-target points applying Euler's formula to obtain a reasonable order of patrol target points. Then three path planning algorithms, A* algorithm based on graph search, RRT* algorithm based on sampling and Q-learning algorithm based on reinforcement learning, are applied on path planning, and combined with Minimal-Jerk algorithm for optimizing trajectories. The performance of the results is finally compared using two evaluation metrics. The acceleration variance and route analysis in planar images are used as evaluation metrics in this paper. It is considered that the smaller the acceleration variance, the more smoothly the robot can move. The gentler the route is, the more effective the robot movement is.

Keywords: plant patrol, path planning, trajectory algorithms, performance comparison.

1. Introduction

Industrial patrol is closely related to factory production safety [1]. Mobile robots equipped with various technologies are a good solution for problems such as poor instrumentation patrols, difficult manual inspection operations and lack of effective solutions for high-risk sites. For mobile robots, good patrol route planning is fundamental [2,3]. This paper takes a mobile robot in a nuclear power plant as an example and applies many kinds of path planning algorithms to get the patrol path. The path is then made more suitable for robot movement by a trajectory optimization algorithm. Finally, acceleration variance and path curve analysis are used as evaluation metrics to estimate path stability through acceleration variance, and path smoothness is analyzed through path curve.

For path planning, graph search-based algorithms are the most common path planning algorithms. This kind of algorithm is based on iterative or improved iterative logic, but is prone to exponential explosion problems for high dimensional cases. In practical scenarios, it is certainly best to find the optimal path, but more often than not, it is sufficient to find a sub-optimal or feasible better path. Hence the emergence of sample-based planning algorithms, which are centred on random sampling. Alternatively, reinforcement learning-based path planning algorithms refer to the learning of the system from the environment to the behavioral mapping. It aims to maximize the value of the reward signal function [4]. Reinforcement learning methods may yield results that are better adapted to the real environment than path planning based on a fixed model of graph searching and sampling.

DOI: 10.54254/2755-2721/32/20230216

However, the resulting paths produced by the three path planning algorithms described above have more inflection points and are not suitable for the robot to follow the path directly. This can lead to more problems with decelerating and turning, making the acceleration of the motion non-linear and demanding on the robot's motor. With a known series of trajectory points, taking into account dynamics constraints and environmental constraints, the use of curves from computer graphics techniques to generate a feasible smooth path that allows for continuous acceleration can result in an optimized route that is robot friendly [5].

Therefore, this paper chooses to select the typical algorithm that works well among the three path planning algorithms, with the A* algorithm representing graph search-based path planning, the RRT* algorithm representing sampling-based path planning and the Q-Learning algorithm representing reinforcement learning-based path planning. After obtaining results by the above three algorithms, this paper applies the Minimal Jerk algorithm to complete trajectory optimization using derivative constraints, continuity constraints, and obstacle constraints [6].

2. Methodology

2.1 Environmental characterization of nuclear power plants

This paper takes the nuclear power plant as an application scenario to design the patrol path of a mobile robot. A section of the 3D design map from the nuclear power plant is intercepted and simulated, as shown in Figure 1 (a). Based on the specific design data, it is transformed into a plane raster map. At the same time, the patrol path has to pass through multiple objective points based on the analysis of the fire-prone points of the specific nuclear power plant equipment. The final plane map is used as the basic environment, as in Figure 1 (b).

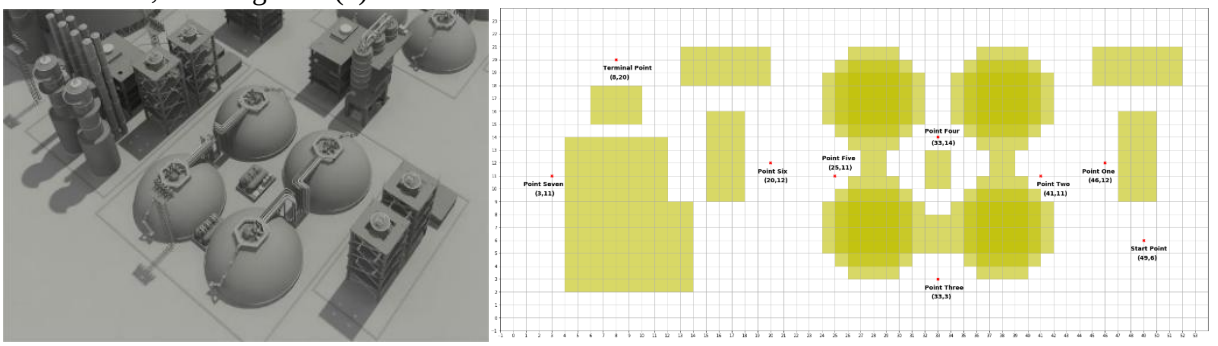


Figure 1. Nuclear power plant scene map.

2.2. Multi-objective path planning for mobile robot patrols

2.2.1. Processing of multi-objective points for path planning. This paper derives relatively better patrol paths by calculating the Euclidean distance between multiple objective points. The equation for calculating the Euclidean distance is shown in Equation 1.

$$\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{1}$$

The procedure for obtaining the patrol order of multiple objectives is as follows. First, add 9 patrol points to an array. Start with the "start point" first, and at the same time delete the ‘start point’ that has been reached in the array. Calculate the Euclidean distance between the ‘start point’ and the rest of the points in the array [7]. Move to the nearest point and use it as the new start point. Then, delete the current point in the array that has been reached. Finally, the search for the nearest point is repeated until the patrol point ‘end point’ is reached. After experimentation, the final patrol order is: start point - first point - second point - fourth point - fifth point - sixth point - third point - seventh point - end point.

DOI: 10.54254/2755-2721/32/20230216

2.2.2. Search-based planning. The A* algorithm, based on graph traversal search, is very commonly used in path planning tasks. With the research of scholars, it has good performance and accuracy. Therefore, this paper uses the A* algorithm as a representative search-based method to complete the initial planning of patrol paths.

The A* algorithm which is a heuristic algorithm uses a priority queue [8]. Elements in the priority queue are given priority, with the highest priority element being first deleted. The heuristic search is planned through the valuation function of the location. The valuation function is represented as shown in Equation 2.

$$f(n) = g(n) + h(n) \tag{2}$$

where n represents node n, f(n) is the valuation function of node n, g(n)is the actual cost from the initial node to node n in the real state space, and h(n) is the estimated cost of the best path from n to the target node, which can guide the search algorithm towards the end point, mainly using the Euclidean distance or Manhattan Distance.

2.2.3 Sampling-based planning. The RRT algorithm, as a typical sampling-based algorithm, is often compared and discussed with the A* algorithm [9]. However, this paper argues that the A* algorithm has undergone many evolutions, just using the original RRT algorithm for comparison is not the right standard. Therefore, this paper uses an optimized RRT algorithm, the RRT* algorithm, as one of the comparison algorithms.

The path obtained by RRT* algorithm is asymptotically optimal, i.e., the path planned becomes more optimal and less costly as time increases. RRT* is based on the original RRT algorithm, with optimizations for reselecting the parent node and reconnecting.

2.2.4. Reinforcement Learning-based Planning. Q-Learning in Reinforcement Learning is the most basic algorithm, but applied to path planning, it usually can only be based on the grid to achieve four directions movement to obtain a path with the highest Q value. In order to better compare with A* and RRT* algorithms, this paper preliminarily optimizes the Q-learning algorithm so that it can move along eight directions and obtain a reasonable Q value. This simple optimized Q-Learning algorithm is used in this paper as a representative of reinforcement learning-based planning.

Q-learning algorithm aim to learn expectation of the benefit of the state action value function [10]. Q-learning is solved using a value iteration approach, the core iterative formula of which is shown in Equation 3.

$$Q_{k+1}(s,a) = r(s,a) + \gamma \cdot \max_{a\hat{} \in A}\{Q_k(s\hat{},a\hat{})\} \tag{3}$$

where $Q_{k+1}(s,a)$is the $(k+1)$th iteration function, s and a denote the current state and the action performed and they belong to state space S and action space A, respectively, $r(s,a)$ denotes the immediate reward after the execution of action a in state s, $s\hat{}$and $a\hat{}$ denote the next state and action, and γ denotes the discount factor.

DOI: 10.54254/2755-2721/32/20230216

2.3. Trajectory optimization and performance comparison metrics

2.3.1. Trajectory optimization. Minimum-jerk algorithm directly constrains the position, velocity, and acceleration of the head and tail by a total of 6 equations, so the optimization parameters must provide more than 6 degrees of freedom. 5th order polynomials have 6 coefficients, so the minimum order of a polynomial that meets the requirements is 5, so a 5th order polynomial can be chosen to represent each segment of the trajectory. By constructing an objective function that satisfies the derivative constraint, continuity constraint and obstacle constraint simultaneously, continuity optimization is ultimately accomplished in position, velocity, and acceleration. This paper uses the minimal-jerk algorithm for trajectory optimization, which minimizes the total jerk by solving for the coefficients of each segment of the trajectory while satisfying the constraints.

The trajectory is constrained using the path points obtained from path planning. Continuity constraint enables smooth transitions between adjacent trajectories. Obstacle constraint allows trajectories to be smooth without colliding into obstacles and boundaries. For paths with collisions, the midpoints will be taken for the points at the two ends of the collision and added to the initial path points, and the process will be repeated until no collisions occur.

2.3.2. Performance comparison metrics. The magnitude of acceleration is an important parameter for measuring the degree of change in the motion of an object. The variance of acceleration can be used to indicate whether the change in velocity is smooth or not, i.e., the smaller the variance of acceleration, the smaller the degree of change in velocity. Combined with the mobile robot for analysis, the mobile robot adjusts the acceleration of the movement by adjusting the rotational speed of the motor, and frequent changes in acceleration during the design process will increase the load of the motor, and at the same time led to a large difference between the simulation results and the actual performance. Therefore, this paper considers that the smaller the acceleration change, the better the trajectory optimization effect.

From the obtained map, the path curve can be clearly seen. A smooth path curve is the goal of trajectory optimization. The Minimal-jerk algorithm will increase the path points on the original path when satisfying the obstacle constraints, which leads to oscillations of the path at the corners. Therefore, this paper argues that the smoother the path in the resulting graph, the better the trajectory optimization.

Ultimately, this paper determines to use acceleration variance and path curve analysis as evaluation metrics, estimating path stability through acceleration variance and analyzing path smoothness through path curves.

3. Experiment results and analysis

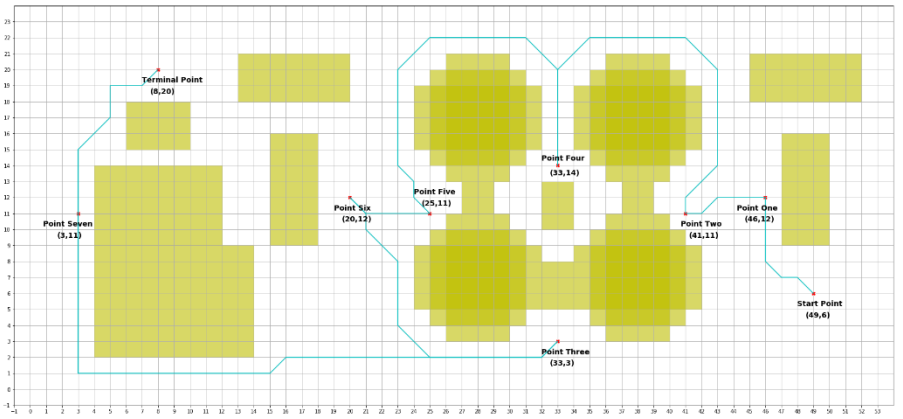
3.1. Multi-objective path planning for mobile robot patrols

3.1.1. Experimental procedure for three types of path planning. The result of the determined patrol sequence applying the A* algorithm to is shown in Figure 2 (a). As the A* algorithm is optimized based on a raster map, each step has 8 directions, so it is clear from the figure that the path is strictly along the raster line.

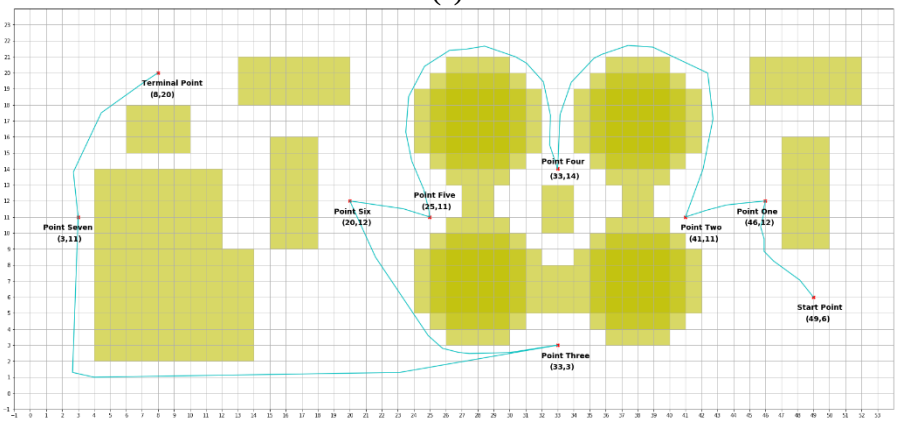
Applying the RRT* algorithm and setting the number of iterations to 20,000, the good initial path is found in the determined patrol sequence, and the obtained path results are shown in Figure 2 (b). Because the RRT* algorithm is based on a random tree that iterates through the map after random sampling to find the best path, the figure shows the original path is smoother than the raster-based path planning algorithm.

Setting the number of iterations to 20,000, the good initial path in the determined patrol sequence applying the Q-Learning algorithm is shown in Figure 2 (c). In this paper, the Q-value is calculated by deducting 1 point for "up, down, left and right" actions and 2 points for "up right, down right, up left and down left" actions according to the length of the movement path. This method optimizes movement in only four directions.

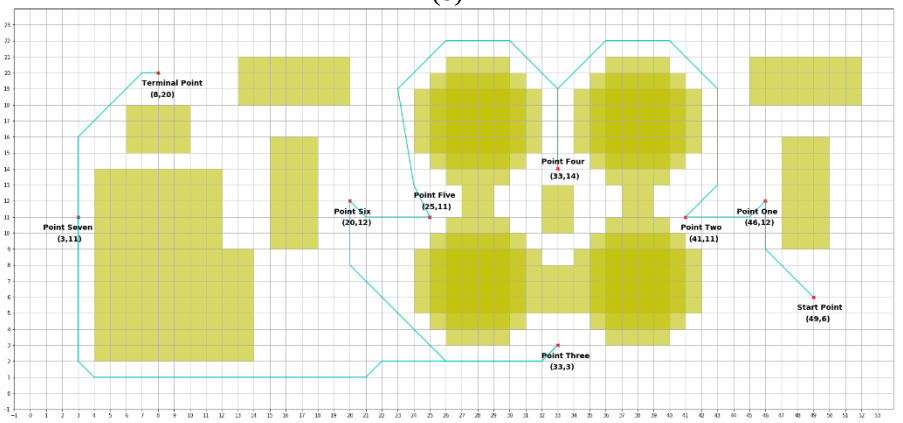
DOI: 10.54254/2755-2721/32/20230216



(a)



(b)



(c)

Figure 2. Original path maps applying three representative algorithms.

DOI: 10.54254/2755-2721/32/20230216

3.2. Trajectory optimization and performance comparison

3.2.1. Trajectory optimization. Applying the Minimal-jerk algorithm, trajectory optimization is performed for each patrol segment of the results obtained from the three path planning algorithms separately, and the results obtained after integration of the routes are implicated in Figure 3.

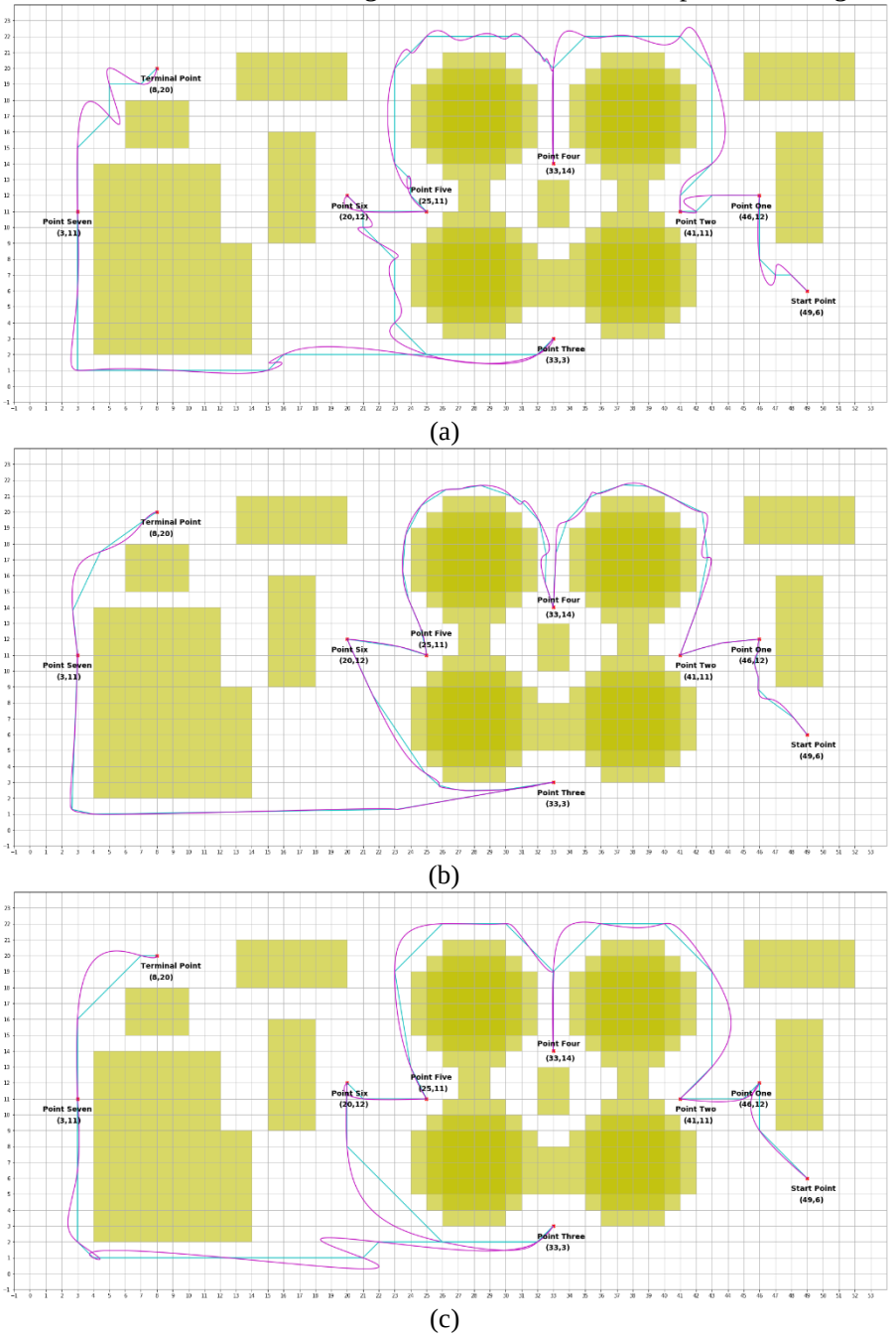


Figure 3. Path maps after trajectory optimization with the application of three representative algorithms.

Most of the trajectory optimization makes the turning points of the path smooth, ensuring that the speed at the start and end points is reduced to zero for operations such as flame recognition and turning. However, in the vicinity of obstacles with circular edges, there is a part of the optimization that allows the curve to have a back-and-forth route in order to pass through all the path points and avoid the

DOI: 10.54254/2755-2721/32/20230216

obstacles. For this problem, Dubin's curve can be added to the application of the Minimal jerk algorithm. Dubin's curve finds the shortest smooth path connecting the points in a given curvature range using arcs and line segments, given two points in the plane and the direction of motion.

Therefore, combining the metrics of the route in the image, the Q-learning algorithm is optimal, followed by the A* algorithm. the RRT* algorithm is less effective after trajectory optimization as there is already some optimization of the route itself.

3.2.2. Performance comparison. Data visualization of the time and acceleration obtained after trajectory optimization for each section of the path was also carried out, as shown in Table 1.

Table 1. Data visualization after trajectory optimization.

	Segment	Time	Acc_max	Acc_min	Acc_var
A star	**1**	8	1.195646795	-1.099370938	
	2	6	3.473192473	-3.950846006	
	3	16	1.132717089	-1.944708846	
	4	40	1.0414029	-0.82221811	
	5	4	6.444610778	-4.441829464	
	6	18	2.804554038	-2.971793668	
	7	18	2.902686338	-3.122272144	
	8	10	2.114110073	-1.928609996	
	Total	120	6.444610778	-4.441829464	1.335669913
RRT star	**1**	16	0.322229195	-0.941182813	
	2	8	1.026808454	-1.344704875	
	3	24	1.026808454	-1.344704875	
	4	30	0.730899365	-0.904737421	
	5	4	2.57232128	-2.103306515	
	6	16	1.776005243	-0.928693255	
	7	14	4.982078066	-5.970267201	
	8	8	1.402541805	-0.734994606	
	Total	120	4.982078066	-5.970267201	1.432004361
Q-learning	**1**	4	5.937767322	-3.36799225	
	2	4	3.322899681	-3.910393499	
	3	12	2.207471676	-2.759411464	
	4	16	1.846009635	-1.603679727	
	5	4	6.444610778	-4.441829464	
	6	8	3.344967509	-2.853384159	
	7	16	7.844555797	-9.923111617	
	8	6	2.202703002	-3.505841076	
	Total	70	7.844555797	-9.923111617	5.980346245

In Table 1, Time indicates the time at which it will move after optimisation. Acc_max indicates the maximum value of acceleration in the patrol path. Acc_max indicates the minimum value of acceleration in the patrol path. In the Total row, the first is the calculated global time, and the second and third are the maximum and minimum acceleration values along the entire patrol path. Acc_var indicates the variance of the acceleration over the full patrol path.

DOI: 10.54254/2755-2721/32/20230216

The data implicate after trajectory optimization the variance of acceleration for the A* algorithm is approximately 1.34, the variance of acceleration for the RRT* algorithm is approximately 1.43 and the variance of acceleration for the Q-learning algorithm is approximately 5.98. Therefore, based on the metrics, this paper concludes that the optimized A* algorithm and RRT* algorithm have lower motor requirements for the robot.

Combining the optimized trajectories shows that the A* algorithm and the RRT* algorithm add more path points into the optimization in order to avoid obstacles near circular obstacles. This resulted in an increase in the robot's movement time, accompanied by an overall decline in acceleration. Therefore, this could be the reason for the small difference in their acceleration.

4. Conclusions

In summary, for the A* algorithm, the algorithm has a short running time and also results in a better path. After trajectory optimization there are some oscillating routes around circular obstacles, but the overall acceleration variance is small and less demanding on the robot motor.

For the RRT* algorithm, the algorithm has a long run time and may be more demanding on the robot drive board. Because the algorithm itself is not limited by the raster map, the path obtained is more closely matched to the path after trajectory optimization. However, the optimization time is limited in order to go through all the path points, so the robot has a slower travel time and is less demanding on the robot motors.

For the Q-learning algorithm, the algorithm runs in few time, while the result is a better path. Because of the algorithm's reward system, the robot's movement path is more realistic in order to get a higher score, so the algorithm obtains fewer path points. This means more freedom in the path, making the optimization of the trajectory simple. However, in the middle of each path, the robot will be faster. As a result, the acceleration of the robot can fluctuate considerably while ensuring that the starting and ending points are reduced to zero. This is reflected in the large variance of the acceleration obtained, which is more demanding on the robot's motor.

The addition of Dubin's curve to the Minimal-jerk algorithm will be considered to improve the oscillation of the path. In order to get good results, many path points are generated during initial planning. These path points will be compulsory in the subsequent optimisation. Delete some of the path points involved in the optimisation may result in a good optimised path and faster running speed.

References

[1] Alhassan A B, Zhang X, et al. Power transmission line inspection robots: A review, trends and challenges for future research [J]. International Journal of Electrical Power & Energy Systems, 2020, 118: 105862.

[2] Sánchez-Ibáñez JR, Pérez-del-Pulgar CJ, García-Cerezo A. Path Planning for Autonomous Mobile Robots: A Review. Sensors. 2021; 21(23):7898.

[3] Mohanty, P.K., Singh, A.K., Kumar, A., Mahto, M.K., Kundu, S. (2022). Path Planning Techniques for Mobile Robots: A Review. In: Abraham, A., et al. Proceedings of the 13th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2021). SoCPaR 2021. Lecture Notes in Networks and Systems, vol 417. Springer, Cham.

[4] Chen, H., Ji, Y. & Niu, L. Reinforcement learning path planning algorithm based on obstacle area expansion strategy. Intel Serv Robotics 13, 289–297 (2020).

[5] F. Bullo and W. T. Cerven, "On trajectory optimization for polynomial systems via series expansions," Proceedings of the 39th IEEE Conference on Decision and Control (Cat. No.00CH37187), Sydney, NSW, Australia, 2000, pp. 772-777 vol.1, doi: 10.1109/CDC.2000.912862.

[6] Richter, C., Bry, A., Roy, N. (2016). Polynomial Trajectory Planning for Aggressive Quadrotor Flight in Dense Indoor Environments. In: Inaba, M., Corke, P. (eds) Robotics Research. Springer Tracts in Advanced Robotics, vol 114. Springer, Cham.

DOI: 10.54254/2755-2721/32/20230216

[7] Y. Yuan, J. Liu, W. Chi, G. Chen and L. Sun, "A Gaussian Mixture Model Based Fast Motion Planning Method Through Online Environmental Feature Learning," in IEEE Transactions on Industrial Electronics, vol. 70, no. 4, pp. 3955-3965, April 2023, doi: 10.1109/TIE.2022.3177758.

[8] Karaman S, Frazzoli E, Sampling-based algorithms for optimal motion planning. The International Journal of Robotics Research. 2011;30(7):846-894. Doi:10.1177/0278364911406761.

[9] J. Chen, M. Li, Y. Su, W. Li, Y. Lin. "Direction constraints adaptive extended bidirectional A* algorithm based on random two-dimensional map environments", Robotics and Autonomous Systems, 2023Karaman S, Frazzoli E. Sampling-based algorithms for optimal motion planning. The International Journal of Robotics Research. 2011;30(7):846-894. doi:10.1177/0278364911406761.

[10] C. Jin, Y. Lu, R. Liu and J. Sun, "Robot Path Planning Using Q- Learning Algorithm," 2021 3rd International Symposium on Robotics & Intelligent Manufacturing Technology (ISRIMT), Changzhou, China, 2021, pp. 202-206, doi: 10.1109/ISRIMT53730.2021.9596694.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230218

Employing the BERT model for sentiment analysis of online commentary

Bowen Li[1,4], Xiaolu Liu[2] and Ruijia Zhang[3]

[1]International Business School, Henan University, Zhengzhou, 451460, China
[2]College of Computer Science and Technology, National University of Defense Technology, Changsha, 410073, China
[3]College of Artificial Intelligence, Tianjin University of Science and Technology, Tianjin, 300457, China

[4]224240823@henu.edu.cn

Abstract. The objective of this research is to carry out a tone and semantic sentiment analysis of network comments on new media platforms by leveraging the BERT model. With the burgeoning popularity of social media, network comments, rich in emotional and tonal features, have emerged as a significant part of the online culture. Accurate interpretation and analysis of these comments' sentiment and semantic meanings are paramount to grasping online public opinion and user psychology. In this study, the BERT model, lauded for its bidirectional encoding and contextual understanding capabilities, is selected to scrutinize the sentiment and tone of network comments on new media platforms. Through a process of pre-training and fine-tuning, the sentiment attitudes and polarity of comments are accurately identified along with their conveyed tonal features, such as joy, anger, and sarcasm. Conducting an accurate tone and semantic sentiment analysis of network comments on new media platforms facilitates a profound understanding of user preferences and trends in public opinion. This can assist in optimizing content recommendations, enhancing user experiences, and increasing the operational effectiveness of new media platforms. The outcomes of this research will bear significant implications for studies and applications in online culture, offering invaluable references and guidance in related domains.

Keywords: sentiment analysis, BERT model, new media platforms.

1. Introduction

As social media gains traction and the internet continues to rapidly evolve, online comments have become a crucial element of daily communication. They are typified by concise language, varied expressions, and often harbor an abundance of emotional information. Accordingly, sentiment analysis of online comments has grown into a challenging but practically significant task. The ability to discern users' sentiment towards specific topics, products, or events can yield valuable insights for areas such as social media marketing and public opinion monitoring.

An effective approach to enhancing various natural language processing tasks involves language model pre-training [1]. The BERT (Bidirectional Encoder Representations from Transformers) model, a pre-trained deep bidirectional representation model, captures contextual information from both left

DOI: 10.54254/2755-2721/32/20230218

and right contexts, enabling rich language representations to be learned from unlabeled text. BERT has achieved remarkable success across a multitude of natural language processing tasks, including question answering and language inference [2]. This research, therefore, seeks to harness the BERT model for sentiment analysis of online comments, aiming to refine its performance through fine-tuning. The proposed experimental methodology to conduct sentiment analysis on online comments involves first curating a representative dataset comprising notable online comment data from various recent social media platforms. Subsequently, the dataset undergoes preprocessing, involving the extraction of pertinent features and the incorporation of specialized tokens. The processed dataset is then employed to train the BERT model, followed by fine-tuning to boost the accuracy of sentiment classification. The primary aim of this experimental phase is to optimize the performance of the BERT model on sentiment analysis tasks, particularly pertaining to online colloquialisms. By improving the accuracy of sentiment classification, it concurrently seeks to enhance the model's capability to comprehend and capture nuanced emotions, such as surprise or disgust. This goal is addressed by implementing several techniques, including dataset preprocessing and filtering, adjustments to the BERT model architecture parameters, refinement of the loss function algorithm, optimization of training batch configurations, and the implementation of measures to counter overfitting. The model demonstrated exemplary performance on the test dataset, meeting the predetermined accuracy and F1 score benchmarks. This affirms the effectiveness of the approach for sentiment classification tasks. Additionally, a confusion matrix is utilized to analyze the model's predictive outcomes across various sentiment categories, offering deeper insights into its performance. Sentiment analysis was conducted on online comments using the BERT model, introducing a novel mixed loss function, CombinedLoss. This function melds weighted cross-entropy loss with standard cross-entropy loss and includes L2 regularization to combat overfitting. By leveraging different weights during model training to counter data imbalances across sentiment categories, performance improvement was observed across all categories. After extensive training and testing, the model yielded satisfactory results in the context of sentiment analysis tasks. This research illustrates that a combined loss function, incorporating weights and L2 regularization, can effectively enhance sentiment analysis performance when fine-tuning the BERT model. Importantly, this methodology proves valuable for tackling class imbalance scenarios and improves the model's recognition of minority sentiment categories. This work presents a feasible approach to the sentiment analysis of online comments and lays the groundwork for further exploration of sentiment analysis applications in social media marketing and public opinion monitoring. The findings of this research can act as a beneficial reference for researchers and practitioners in related fields, spurring the development and application of sentiment analysis techniques in practical environments.

2. Related works

Sentiment analysis encompasses the examination of sentiments, opinions, attitudes, and emotions expressed toward specific entities such as topics, products, individuals, and organizations. The goal is discerning the author's viewpoint [3]. A vast body of research has been conducted in this domain, exploring diverse approaches from rule-based methods and bag-of-words techniques to machine learning algorithms [4]. For instance, Peter D. Turney introduced the concept of semantic orientation for unsupervised classification, analyzing sentiments in comments to determine positive or negative orientation [5]. A machine learning-based method combined with semantic sentiment analysis for extracting predictions of suicidal ideation using Twitter data was proposed by Marouane Birjali et al. Furthermore, Penalver-Martinez et al [6] implemented a semantic ontology approach to boost feature extraction effectiveness and applied vector analysis techniques for movie review sentiment analysis.

The crux of most machine learning-based sentiment analysis research is the enhancement of feature extraction algorithms [7]. Notably, Zichao Yang et al. proposed a hierarchical attention network for boosting document-level sentiment analysis via optimized feature extraction algorithms [8]. Moreover, Soujanya Poria et al. refined the feature extraction algorithm by computing the mutual information between features and sentiment categories, thus selecting the most informative feature set to boost sentiment analysis efficacy. In addressing these challenges, this research applies techniques such as

DOI: 10.54254/2755-2721/32/20230218

dataset preprocessing and filtering, fine-tuning of BERT model parameters, enhancement of the loss function algorithm, training batch configuration, and mitigation of data overfitting risks among other approaches.

3. Proposed methods

The objective was to conduct sentiment analysis of internet slang utilizing the BERT model. A selection of representative internet slang data was gathered from contemporary social platforms. Following data integration, the dataset underwent pre-processing, involving feature extraction from the data and the inclusion of special tokens. The processed dataset was then fed into the BERT model for training, with subsequent fine-tuning of the model to enhance its accuracy in sentiment determination. The comprehensive workflow of the model is illustrated in Figure 1.

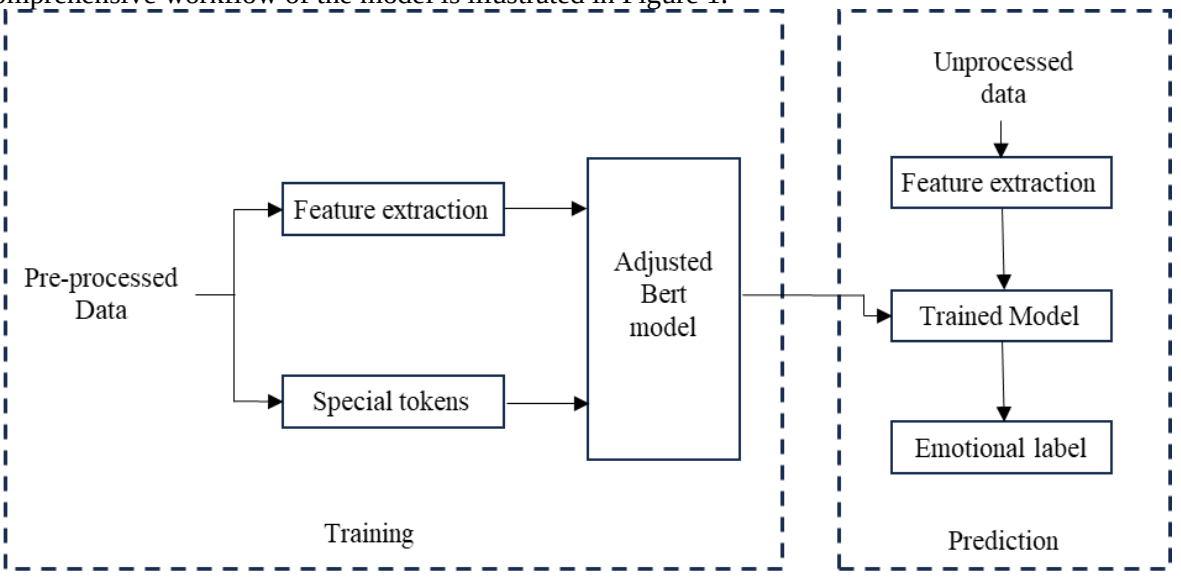


Figure 1. The overall run of the model (Photo/Picture credit: Original).

Table 1. The corpus statistics and label distribution [9].

Emotion	Number	Percentage (%)
Like	4540	11.45
Happiness	9959	25.11
Sadness	14052	35.43
Anger	4562	11.50
Disgust	4876	12.29
Fear	661	1.67
Surprise	1011	2.55
Sum	39661	100.0

3.1. Pre-processing

PyTorch and the Transformers library were utilized for the implementation of the BERT model [1]. A dataset class was designed specifically to transform the text and labels into a format compatible with the BERT model. With the aid of BERT's tokenizer, the text was tokenized and the tokenized results were converted into IDs within the vocabulary. Furthermore, special tokens such as '[CLS]' and '[SEP]' were incorporated, and padding or truncation was applied to ensure a fixed-length input sequence. During the experiment phase, extensive data preprocessing steps were undertaken on the collected internet slang dataset, which ranged from cleaning the data and handling missing values to tokenizing the text. These measures were aimed at enhancing the cleanliness and consistency of the data, with the ultimate goal of improving model performance during both training and prediction phases.

DOI: 10.54254/2755-2721/32/20230218

3.2. Dataset

A significant quantity of internet slang data is sourced from an open-source database on GitHub. Stored within the data directory, this dataset serves as an emotion analysis corpus, with each sample meticulously annotated with one sentiment label [9]. The sentiment labels, manually assigned, span seven distinct emotions: 'happiness', 'sadness', 'anger', 'disgust', 'fear', 'surprise', and 'like'. This broad range of data provides a comprehensive portrayal of varying sentiment expressions in internet slang. The dataset has been divided into training, validation, and testing sets in an 8:1:1 ratio and is encoded in UTF-8. The specific quantity and percentage of slang for each emotion are detailed in Table 1 [10].

3.3. Sentiment analysis

The BERT model, an acronym for Bidirectional Encoder Representations from Transformers, is utilized for sentiment analysis. As a pre-trained deep learning model, BERT has been proven to achieve outstanding results in various natural language processing tasks. The 'BERT-base-uncased' variant is selected for this study, incorporating 12 layers of Transformer architecture, 110M parameters, and a lowercase English vocabulary. Such a choice leverages the robust capabilities of BERT in grasping contextual information and discerning sentiment within internet slang.

3.4. Classification

In this classification task, the BERTForSequenceClassification class is utilized, merging the BERT model with an overlaying classification layer. This model configuration encompasses seven categories, aligning with the seven sentiment labels present in the dataset. To optimize model performance during training, the Adam optimizer and the cross-entropy loss function are employed.

3.5. Training and evaluation

After preprocessing the data, we train the BERT model using the preprocessed dataset. The training process involves iterating over the dataset, adjusting the model's weights through backpropagation, and fine-tuning the model's parameters to optimize its performance. We monitor the training progress, including the loss values and accuracy, to ensure the model's convergence.

Once the model is trained, we evaluate its performance on the validation set. This evaluation involves computing various metrics, such as accuracy and F1 score, which provide insights into the model's ability to correctly predict sentiment labels. Additionally, we visualize the confusion matrix to gain a deeper understanding of the model's performance across different sentiment categories. To evaluate the performance of the trained model, we split the dataset into a training set and a validation set. The training set is used to train the model, while the validation set serves as an independent benchmark for evaluating its performance. We employ accuracy as the evaluation metric, which measures the proportion of correctly predicted sentiment labels.

4. Experiments

4.1. Experimental procedure

The goal of these experiments is to enhance the BERT model, boosting its performance in sentiment analysis tasks, especially when dealing with internet slangs. Simultaneously, these enhancements aim to improve the accuracy of sentiment classification and the capability to comprehend complex sentiments such as surprise and disgust. The primary research question posed is: "How can the accuracy of BERT models be improved for sentiment analysis of the internet slang corpus?" This question is approached through several methods including preprocessing and filtering of the dataset, adjusting the model parameters of the BERT model, enhancing the loss function algorithm, setting up training batch configurations, and implementing measures to prevent data overfitting.

To prepare for model training, the text dataset is preprocessed, which includes using BERT's word splitter to separate words and packaging the processed data into a PyTorch dataset. A loss function, entitled CombinedLoss, is employed. It blends a weighted cross-entropy loss with a standard cross-

DOI: 10.54254/2755-2721/32/20230218

entropy loss and is specifically designed to manage category imbalance by assigning higher weights to a few categories. The original dataset is split in an 80/20 ratio to form a training set and a validation set. This split facilitates the evaluation of the model's generalization ability during the training process and helps prevent overfitting. Training parameters are defined, including the learning rate, weight decay, etc., followed by the use of a Trainer for model training. To further guard against overfitting, EarlyStoppingCallback is employed, which sets a condition to cease training early based on the accuracy of the validation set. The model is ultimately evaluated based on the test set, and test data prediction is performed to obtain predictive labels for the model. The trained model is saved for future use. The experimental setup can be observed in Figure 2.

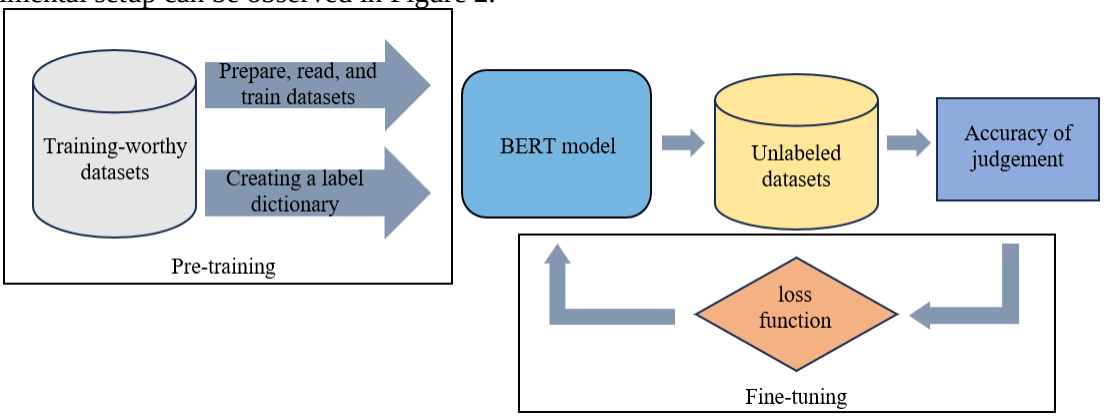


Figure 2. Experimental setup (Photo/Picture credit: Original).

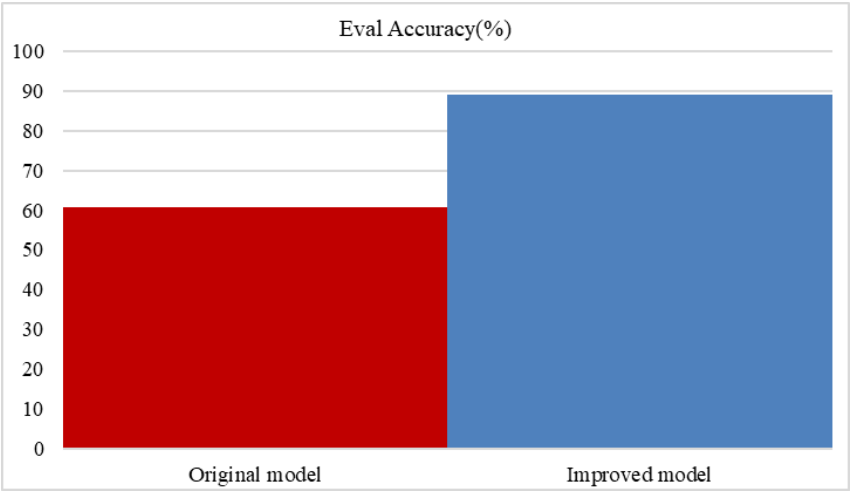


Figure 3. The comparison of accuracy between the improved model and the original model (Photo/Picture credit: Original).

4.2. Results

Ultimately, the model demonstrates satisfactory performance on the test set. Specifically, it meets the benchmarks set for accuracy and F1 scores, thereby validating the approach's effectiveness in the sentiment classification task. A confusion matrix analysis of the model's predictions further elucidates its performance across different sentiment categories. While the model provides reasonable accuracy for most sentiment categories, some misclassifications in certain categories do occur. The approach taken in this research incorporates sentiment analysis based on the BERT model, with the introduction of a novel hybrid loss function, CombinedLoss. This function merges weighted cross entropy with standard cross entropy loss, and includes an L2 regularization term to thwart overfitting. Differing weights in the model training help balance the data imbalance between various sentiment categories, thereby enhancing

DOI: 10.54254/2755-2721/32/20230218

the model's performance across all categories. Post training and testing, the model exhibits commendable results on the sentiment analysis task.

This study reveals that using a hybrid loss function with weights and L2 regularization can effectively fine-tune the BERT model to improve sentiment analysis performance. In situations of category imbalance, this method notably enhances the model's ability to recognize minority classes. The following segment presents the experimental results. Figure 3 illustrates a comparison of accuracy between the enhanced model and the original model. Figure 4 showcases the alterations in precision, loss function, and F1 value during the model's improvement process. As shown in Table 2.

Table 2. Performance Metrics Across Iterations for the Enhanced BERT Model in Sentiment Analysis.

iteration	1	2	3	4	5	6	7
Accuracy	41%	43%	53%	59%	61%	79%	89%
Loss Decrease	1.66	1.49	1.35	1.30	1.17	1.07	0.57
F1 Score	0.15	0.18	0.24	0.42	0.60	0.77	0.82

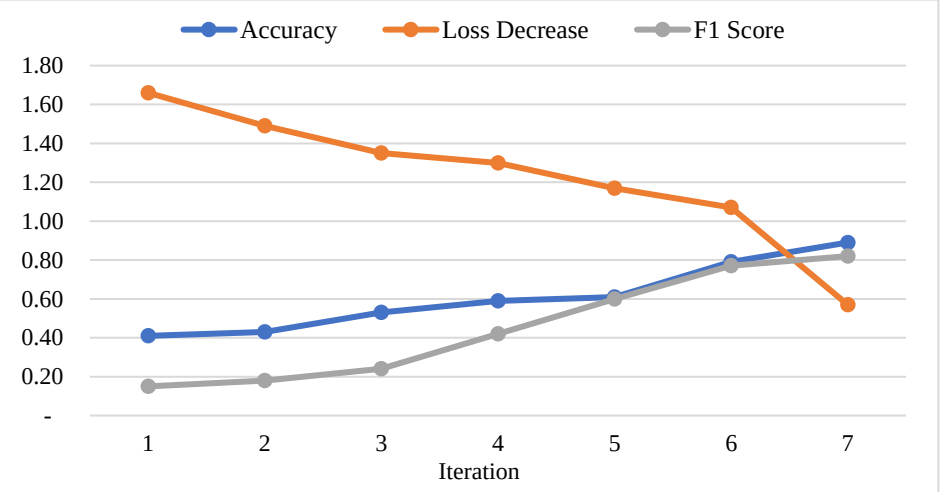


Figure 4. The changes in precision, loss function, and f1 value during the improvement of the model (Photo/Picture credit: Original).

5. Conclusion

This research reveals that the application of a weighted combined loss function and L2 regularization significantly enhances the performance of the BERT model in sentiment analysis tasks. Remarkably, this methodology shows a marked improvement in the recognition of minority classes in instances of class imbalance.

These findings hold substantial practical value for sentiment analysis and natural language processing, offering an effective solution to the common class imbalance issues frequently found in real-world datasets. Moreover, they illustrate how the integration of multiple loss functions and regularization methods can help avoid overfitting and bolster the model's generalization capability. From a theoretical standpoint, this research introduces an innovative model training strategy that optimizes the model through the amalgamation of various loss functions and regularization methods. This novel approach presents a fresh viewpoint on how to address class imbalance problems in natural language processing tasks, and provides valuable insights for the training of deep learning models.

While this study concentrates on text data, future research could extend sentiment analysis to multimodal data, including images, audio, and videos. By integrating sentiment information from diverse data modalities, a more holistic understanding and analysis of emotional expressions can be achieved, thereby broadening the application spectrum of sentiment analysis. In conclusion, this research underscores the practical value of utilizing weighted combined loss functions and L2

DOI: 10.54254/2755-2721/32/20230218

regularization to improve sentiment analysis performance with the BERT model. It also suggests potential future research directions, such as the extension of sentiment analysis to multimodal data, and provides novel insights for tackling class imbalance issues in natural language processing tasks.

Authors contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Devlin, J., Chang, M.W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 4171-4186.

[2] Cao, Y., Sun, Z., Li, L., & Mo, W. (2022). A Study of Sentiment Analysis Algorithms for Agricultural Product Reviews Based on Improved BERT Model. Symmetry, 14(1), 1604.

[3] Birjali, M., Beni-Hssane, A., Erritali, M. (2017). Machine Learning and Semantic Sentiment Analysis based Algorithms for Suicide Sentiment Prediction in Social Networks. Procedia Computer Science, 113, 65-72.

[4] Turney, P.D. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), 417-424.

[5] Peñalver-Martinez, I., Garcia-Sanchez, F., Valencia-Garcia, R., Rodríguez-García, M.Á., Moreno, V., Fraga, A., & Sánchez-Cervantes, J.L. (2014). Feature-based opinion mining through ontologies. Expert Systems with Applications, 41, 5995-6008.

[6] Yang, Z., et al. (2016). Hierarchical attention networks for document classification. Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies.

[7] Sun C, Huang L, Qiu X. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence[J]. arXiv preprint arXiv:1903.09588, 2019.

[8] Ray B, Garain A, Sarkar R. An ensemble-based hotel recommender system using sentiment analysis and aspect categorization of hotel reviews[J]. Applied Soft Computing, 2021, 98: 106935.

[9] AlQahtani A S M. Product sentiment analysis for amazon reviews[J]. International Journal of Computer Science & Information Technology (IJCSIT) Vol, 2021, 13.

[10] Sudhir P, Suresh V D. Comparative study of various approaches, applications and classifiers for sentiment analysis[J]. Global Transitions Proceedings, 2021, 2(2): 205-211.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230219

YOLO model-based target detection algorithm for UAV images

Anqi Wei

School of Communication & Information Engineering, Shanghai University, Shanghai, 200444, China

waq99@shu.edu.cn

Abstract. The increasing popularity of drones has paved the way for their utilization in various sectors, including civil, commercial, and government agencies. These unmanned aerial vehicles have proven to be invaluable in capturing images and videos from vantage points that were once difficult to access, leading to a wide range of applications. Images captured by drones often have target objects that are small in the frame and a large number of photos or videos captured, so that it is difficult for people to find the target objects in the photos. Nowadays, target detection of images captured by drones through deep learning methods, such as the YOLO algorithm, can greatly help people's work. In this paper, the authors of this paper have investigated for the last three years, for target detection of UAV images, optimization based on the original YOLO algorithm to achieve improved detection results. The research in this paper summarizes the existing research results and is of great significance to the subsequent research and application of UAV image processing.

Keywords: yolo, drone image, UAV image, target detection.

1. Introduction

Commercial small aerial vehicles, also known as drones, have the advantage of being portable and more flexible during flight. Drones on the market today are often equipped with high-definition cameras and have real-time sharing capabilities that allow users to capture images and analyze them at any time. Nowadays, using drones for real-time monitoring is becoming increasingly common, and the technology can be applied to human flow monitoring, road traffic flow assessment, forest fire inspection, and large motor equipment inspection. The technology reduces the need for personnel for such tasks or reduces the pressure on personnel, and personnel are less likely to need to travel to dangerous areas, providing more security for staff. The use of drones for inspection and finding, in addition to the hardware equipment requirements, also needs to be able to picture quickly to complete the target detection and the need for relatively high recognition accuracy.

There are the following problems to be solved in target recognition of images using UAVs. The first one is the dataset. In early image target detection algorithms, such as face recognition, license plate recognition, etc., most images are portraits or the front of the object. Much research is also based on the training of such datasets obtained. However, due to the flight characteristics of UAVs, the images captured are all top-down views, and the algorithms also need to be retrained on the data under this characteristic. Second, the size of the target detection object captured by the UAV, even at high

DOI: 10.54254/2755-2721/32/20230219

resolution, will be smaller than the target object captured by the flat camera. The algorithm needs to enhance the detection performance based on this property, such as improving the characterization of target features, which can improve the correct rate of target object detection. In addition, images captured by UAVs often have the problem of rotational angle because it is impossible to restrict the UAV's camera during flight [1]. Therefore, the algorithm's detection cannot be limited to horizontal or vertical situations but must have good detection for any rotation angle.

As a popular mainstream target detection algorithm, the YOLO algorithm is fast, small, and guarantees a good detection rate [2]. Many researchers have based on this algorithm framework and further optimized the framework for the characteristics of UAV images. In this paper, the authors investigate the research results in this direction in the last three years, so that the subsequent research can be further deepened.

2. Research results in the past three years

The YOLO algorithm is one of the most popular deep learning-based target detection algorithms and has now been released in YOLOv7. YOLOv4 builds on YOLOv3, a major breakthrough in real-time target detection and improved accuracy. YOLOv5 builds on the success of YOLOv4, further improving accuracy and enhancing the model's generalization capabilities. YOLOv7 is the latest version, optimized for speed and accuracy.

Luís Augusto Silva et al. used YOLOv4, YOLOv5, and YOLOv7 for target detection in UAV images, and achieved 59.9% mAP in YOLOv5, 65.70% mAP in YOLOv5 with Transformer Prediction, and 73.2% mAP in YOLOv7 [3]. For UAV images, Transformer Prediction Head (TPH) is added to YOLOv5, improving the large-scale change of target objects in UAV images. And effective methods for target detection in UAV images were screened, and self-trained classifiers were used to improve the classification accuracy under too vague criteria. The project also improved the classification of the dataset of pavement damage, which made the algorithm more effective in detection through more detailed classification.

Jinsu An et al. proposed a method to improve the YOLOv5 network with CBAM to implement an algorithm for target detection for UAV vision and achieved a mAP of 22.56% [4]. The addition of CBAM to the network of the original YOLOv5 realized the introduction of combinatorial fast structures. Because CBAM combines channel attention and spatial attention, the improved YOLOv5 has a convolutional block attention module. CBAM strengthens the performance of the attention module based on the BAM module while remaining lightweight and generalized, and can be flexibly added within any CNN architecture and train the model end-to-end. During the optimization of the three parts of YOLOv5, CSPDarknet53 was used for the backbone, PANet for the neck, and B×(5+C) for the output layer of the head. With this optimization, the algorithm has a better performance for the extraction of feature information.

Zhengwei Li et al. proposed R-YOLOv5, a lightweight rotating target detection algorithm, which achieves an mAP of more than 80% on different datasets and shows good generalizability [5]. The algorithm is based on YOLOv5 and optimized for the backbone feature extraction network, the neck feature fusion network, and the prediction head. The angle prediction branch is combined in the part of the prediction head, and the Circular Smooth Labeling (CSL) angle classification method is introduced to be able to measure the distance between the angle labels, which makes YOLOv5 able to detect the scenes with unknown rotation angles. The problem of tiny objects in UAV images not being able to retain feature information to higher level feature information after multiple convolution operations is also partially solved by feature fusion embedded in the Swin Transformer block (STrB). In UAV images, there is also the problem that similar objects cannot be detected correctly due to more noise in the feature information. Using the Feature Enhanced Attention Module (FEAM), which incorporates the Multihead Self Attention (MHSA) module, enhances the ability of the network to capture the information and ensures detection accuracy. There is a problem of object scale distortion in the images captured by the UAV. By adding the Adaptive Spatial Feature Fusion Structure (ASFF) to the head of YOLOv5, the algorithm can adapt to objects of different scales and no longer loses object information.

DOI: 10.54254/2755-2721/32/20230219

The algorithm reduces the excessive computational complexity during feature fusion in the backbone network, increases the utilization of detailed information, and improves multi-scale feature fusion in the head.

Oyku Sahin et al. In response to the problems of high viewing angles, large target scale variations and unstable image quality of UAV images, and an improved network structure was used [6]. The network structure utilizes a combination of Convolutional Neural Networks (CNNs) and Feature Pyramid Networks (FPNs) to effectively capture target information at different scales in the image, thus improving target detection accuracy. The team also proposed a new loss function for optimizing the training process of the network. This loss function combines the position loss and confidence loss of the target frame, as well as the categorization loss, which integrates multiple aspects of target detection and enables the network to better learn the position and category information of the target. In addition, to cope with the problem of changing target scales in UAV images, a target scaling technique is introduced in the paper for appropriate processing of targets at different scales in the image, improving the robustness and performance of target detection. The experimental results show that YOLODrone can achieve higher target detection accuracy and faster detection speed in UAV images, proving the superiority and practicality of the method.

Sushil Kumar et al. utilized deep learning techniques, based on the YOLO V5 algorithm, to improve the accuracy and efficiency of target detection in Unmanned Aerial Vehicle (UAV) surveillance images [7]. The algorithm adopts a one-stage detection strategy, transforms the target detection problem into a regression problem, and utilizes a feature pyramid network (FPN) for multi-scale feature fusion, which results in an excellent performance in dealing with target scale variations. Optimizing the YOLO V5 algorithm for the characteristics of UAV surveillance images, the researchers introduced a series of improvements. The method can better adapt to the complex scenes of UAV surveillance images by introducing special convolutional layers, attention mechanisms and data enhancement techniques. Meanwhile, applying pre-training and migration learning enables the model to be trained on small sample data with better generalization ability. To verify the method's effectiveness, the researchers constructed a UAV surveillance image dataset containing various types of targets and conducted extensive experiments on the dataset. The experimental results show that the target detection and recognition method based on the YOLO V5 algorithm achieves significant performance improvement in UAV surveillance images. Compared with traditional methods and baseline models, the method has significant advantages in target detection accuracy and efficiency, and is able to identify and localize various types of targets in surveillance scenes more rapidly.

Weibiao Chen et al. proposed the DSM-YOLO v5 algorithm, which aims to improve the accuracy and efficiency of target detection in UAV aerial images [8]. The paper chose the YOLO v5 algorithm as the basic framework, and the algorithm adopts a one-stage (one-stage) detection strategy, which performs well in dealing with target scale variations by transforming the target detection problem into a regression problem, as well as utilizing a feature pyramid network (FPN) to achieve multi-scale feature fusion. Optimized for the characteristics of UAV aerial images, the paper proposes the DSM (Digital Surface Model) mechanism. The DSM technique further improves the accuracy and robustness of target detection by acquiring the surface elevation information and fusing it into the YOLO v5 algorithm. The introduction of the DSM helps to better localize and recognize the target. The experimental results show that the DSM-YOLO v5 algorithm is able to achieve significant performance improvement in UAV aerial images. Compared with the traditional methods and benchmark models, the algorithm has obvious advantages in target detection accuracy and detection speed.

Songyun Zhang proposed a fast target detection method for UAV imagery based on MobileNet-YOLO V4 model [9]. The authors used MobileNet as the basic network structure, which is a lightweight convolutional neural network with fewer parameters and computational complexity, suitable for target detection on devices with limited resources. Combining MobileNet with YOLO V4, which is an advanced model in the field of target detection, a series of technological improvements, such as the CIOU loss function, the SAM module, and PANet, are used to improve the accuracy and

DOI: 10.54254/2755-2721/32/20230219

robustness of target detection. The authors used a series of optimisation measures to further improve the speed of target detection. For example, the number of parameters of the model is reduced and the network structure is optimized by network pruning and quantization techniques, which enables the model to perform target detection in UAV imagery quickly and efficiently. In addition, this paper also carries out data enhancement and preprocessing for the characteristics of UAV images, which increases the diversity of samples and improves the generalization ability of the model. Experiments are conducted on UAV image datasets containing various types of targets and different complex scenes, and the results show that the method achieves significant performance improvement in the UAV image target detection task. Compared with traditional methods and other target detection models, the method based on the MobileNet-YOLO V4 model has obvious advantages in terms of speed and accuracy.

Xianghong Cheng et al. improved the YOLO V5 algorithm to detect small targets efficiently and accurately in UAV aerial images [10]. The research team adopted the YOLO V5 algorithm as the basic framework. Aiming at the low-resolution characteristics of small targets, the team introduced a higher-level feature pyramid network to enhance the representation of small targets. To suppress the background interference, this paper adds an attention mechanism, which enables the algorithm to focus more on the important features of small targets, thus improving the accuracy of detection. To verify the effectiveness of the improved YOLO V5 algorithm, the researchers constructed a UAV aerial image dataset containing many small target samples and conducted a series of experiments. The experimental results show that the improved YOLO V5 algorithm significantly improves UAV aerial images' performance. Compared with the traditional method and the benchmark model, the algorithm has obvious advantages in small target detection accuracy and detection speed.

3. Conclusion

In recent years, with the rapid development of UAV technology, UAV image target recognition plays an increasingly important role in military, civil and industrial fields. Among them, the target recognition technology based on YOLO (You Only Look Once) algorithm has attracted much attention. In this paper, we conduct in-depth research on the improvement of the YOLO algorithm for UAV image target recognition in the past three years, in order to understand the performance of the algorithm in solving the problems of rotational angle, small target pixels, etc., and to explore the more detailed enhancement achieved by different versions of YOLO on these improvements.

Our research identified eight important studies that have done a great deal of exploratory work on the particular challenges of UAV imagery. First, researchers have proposed a series of solutions to the problem of target rotation, which is prevalent in UAV imagery. Some of these methods are based on YOLO and introduce a rotation invariance module, which enables the algorithm to better handle targets with inconsistent rotation angles. These improvements effectively improve the accuracy of target recognition and enhance the application of UAVs in dynamic environments.

Second, another group of researchers proposed a series of innovative solutions for the problem of too small target pixels in UAV images. These methods mainly focus on the feature extraction part of the YOLO algorithm, which effectively enhances the perception of small targets by introducing the attention mechanism and image pyramid structure. The results show that these improved algorithms have outstanding performance in recognizing small targets, which greatly improves the detection rate of UAVs on small targets and provides strong support for dealing with complex and changing practical application scenarios.

It is worth noting that although all of these studies improved on the YOLO algorithm, they did not retain the advantages of its inherent light weight and fast detection. This advantage is especially important for UAV image target recognition in today's demands for efficient computation. These researches improve the performance and meet the demand for real-time and practicality in practical applications, enabling UAV technology to work even better in target searching, monitoring and tracking.

DOI: 10.54254/2755-2721/32/20230219

In addition, some of the research projects have constructed their own datasets to better validate the performance of the algorithms. By using targeted datasets, these studies can more fully demonstrate the superiority of their improved algorithms and optimize them for specific scenarios. This trend in dataset construction has provided UAV image target recognition research with more reliable evaluation criteria, allowing algorithms to be trained to produce better results and be better adapted to specific mission requirements.

In summary, research on improving the YOLO algorithm for UAV image target recognition has made great progress in the past three years. By improving the algorithm for problems such as rotation angle, too small target pixels, and constructing a customized dataset while retaining the algorithm's advantages, researchers have made positive contributions to the development of UAV technology. However, it is also important to realize that the challenges faced by target recognition in the real world are complex and diverse, and continuous efforts are still needed to further improve the robustness and accuracy of the algorithms in the future to promote the application of UAV technology in a wider range of fields.

References

[1] Li Z, Liu X, Zhao Y, Liu B, Huang Z and Hong R 2021 *Journal of Visual Communication and Image Representation* **77** 103058
[2] Jiang P, Ergu D, Liu F, Cai Y and Ma B 2022 *Procedia Computer Science* **199** 1066–73
[3] Silva L A, Leithardt V R Q, Batista V F L, Villarrubia González G and De Paz Santana J F 2023 *IEEE Access* **11** 62918–31
[4] An J, Putro M D, Priadana A and Jo K-H 2023 *2023 IEEE International Conference on Industrial Technology (ICIT) 2023 IEEE International Conference on Industrial Technology (ICIT) (Orlando, FL, USA: IEEE)* pp 1–6
[5] Li Z, Pang C, Dong C and Zeng X 2023 *IEEE Access* **11** 61546–59
[6] Sahin O and Ozer S 2021 *2021 44th International Conference on Telecommunications and Signal Processing (TSP) 2021 44th International Conference on Telecommunications and Signal Processing (TSP) (Brno, Czech Republic: IEEE)* pp 361–5
[7] Kumar S and Kumar C 2023 *2023 International Conference for Advancement in Technology (ICONAT) 2023 International Conference for Advancement in Technology (ICONAT) (Goa, India: IEEE)* pp 1–5
[8] Chen W, Jia X, Zhu Zh et al. *Computer Engineering and Applications* **1-11**[2023-07-27].http://kns.cnki.net/kcms/detail/11.2127.TP.20230705.2129.004.html
[9] Zhang Song Yun. *Jiangxi Science* 2023 **41(02)** 339-342+355.DOI:10.13990/j.issn1001-3679.2023.02.020.
[10] Cheng X, Cao Y, Hu Y et al. *Flight Control and Detection* 2023 **6(01)** 80-85.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230221

A study on search techniques in the game-tree

Shaojia Zhang

Hengshui High School of Hebei, Hengshui, 053000, China

jlovei85072@student.napavalley.edu

Abstract. Artificial intelligence has developed a lot in the game field and search techniques on game-trees are essential for AI-game-playing. As many techniques for searching the game-trees have been published, the time consumption of search has decreased and the accuracy of it has increased. This paper would provide a comprehensive review of existing algorithms and improvements, including their mechanisms and performances. These techniques are first divided into two sections, techniques based on Alpha-Beta pruning and techniques based on Monte-Carlo tree search. Furthermore, techniques based on Alpha-Beta pruning are further subdivided into narrow window properties and information from previous searches based on the specific aspects they focus on. Finally, this paper concludes by summarizing the performance of these techniques and identifying their suitable application scenarios, as well as suggesting potential directions for future research.

Keywords: game-tree search, Minimax, Alpha-Beta, improvement, windows, Monte-Carlo tree search.

1. Introduction

For a long time, people are keen to build AI programs to compete against humans in games. While humans were always winners in the early time, with the invention and development of many game-tree algorithms, computers are now in some sense able to defeat humans in many games, such as Tic-Tac-Toe, Checkers, Chess, and even Go. Studying search techniques in game-tree and further improving or enhancing them can make AI in games smarter and reduce the cost of such machine thinking. All techniques can be divided into two sections, techniques based on Alpha-Beta pruning and techniques based on Monte-Carlo tree search.

Most game-tree search algorithms are defined in the context of a two-person and zero-sum game-tree of perfect information. In such a game-tree, the root represents the initial game state; each node represents a position in the game; a node's ply, introduced by Arthur Samuel in [1], represents the depth of that node or the number of moves to reach that position; all branches of one node represent all legal moves from that position; leaves, which have no successors, are terminal positions, from which the result of the game can be determined—win, lose, or draw. In the most ideal case expected to happen, the entire game-tree has been generated and searched before the computer knows every strategy for every position of the game. However, this case only happens in simple games, such as Tic-Tac-Toe, but it is unachievable for more complex games, such as Chess, because the time complexity and space complexity increase exponentially in a game-tree. Therefore, evaluation systems and pruning algorithms are needed to predict results without reaching leaves and reduce useless nodes that do not affect final

DOI: 10.54254/2755-2721/32/20230221

results, respectively. This paper reviews several sorts of evaluation mechanisms and search algorithms to make the development and category of search techniques on game-trees clear.

This paper would first introduce some basic algorithms to help readers understand other advanced algorithms. Furthermore, it focuses mainly on improvements and algorithms based on the Alpha-Beta algorithm, the classic algorithm used for searching game-trees, and also discusses Monte-Carlo tree search, which is the most popular search technique recently. It also summarizes characteristics, advantages and disadvantages, and suitable circumstances for these algorithms.

2. Background

This section introduces the basic principles and basic algorithms of game-tree search.

2.1. Minimax and Negamax algorithm

In the Minimax algorithm, two players are commonly called Max and Min. Max moves at odd ply and tries to maximize the score while Min moves at even ply and tries to minimize the score [2]. All leaves have scores evaluated according to the positions they represent. For any other node, if it is at an odd ply, its score is the maximum of scores of all its successors; similarly, if it is at an even ply, its score is the minimum of scores of all its successors. All nodes get their values through one of these two processes recursively. Max would try to move to the node that has the maximal value among the successors while Min would try to move to the node that has the minimal value among the successors.

The Negamax algorithm is a variant of the Minimax algorithm [2]. It is simpler and more convenient than the Minimax algorithm because it does not need to be checked who the next mover is. It is defined that the score of a position for one player is the negation of that for the other one. Except for leaves, every node's value is the maximum of the negation of scores of all its successors, shown as $V = \max(-V_1, -V_2 \ldots - V_b)$. Both two players would try to move to the node that has the maximal value among the successors.

The time complexities of both the Minimax algorithm and Negamax algorithm are $O(b^{ply})$, in which b is the branching factor (the average number of successors each node has) and ply is the depth of the entire game-tree.

2.2. Alpha-beta algorithm

Either the Minimax algorithm or the Negamax algorithm is considered a brute-force algorithm, so pruning or optimization is needed to reduce the time consumption. The Alpha-Beta algorithm defines alpha and beta as the lower bound and upper bound of the score, which means that any value less than alpha or greater than beta is unimportant for final results [3]. For the example shown in Figure 1, no matter what value X is, the score of node-1 is certainly 7. In this case, the subtree of node-5 is not necessary to be searched because it would not have any influence on the final result, so it can be cut off.

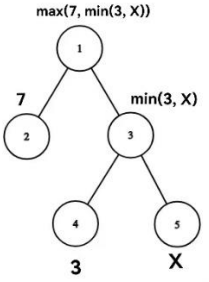


Figure 1. A sample tree that can cause a cutoff.

This algorithm can be unstable, for the time complexity is $O\left(b^{\frac{ply}{2}}\right)$ in average cases but $O(b^{ply})$ in worst cases. Although the time consumption is much less than Minimax and Negamax algorithms, it is still very large for complex games.

DOI: 10.54254/2755-2721/32/20230221

3. Improvements based on the Alpha-Beta algorithm

Since the pure Alpha-Beta algorithm still has problems with time complexity and stability, lots of improvements based on the Alpha-Beta algorithm have been proposed. In this paper, they are divided into two sections according to which aspect they optimize.

3.1. Algorithms based on the property of narrow windows

The efficiency of the Alpha-Beta algorithm significantly depends on the values of alpha and beta. The greater alpha is and the less beta is, the more cutoffs can be caused. Since alpha should be less than or equal to beta, it can also be written as the narrower the window is, the better the performance of the Alpha-Beta algorithm is. There are two techniques based on this property, aspiration window search and minimal window search.

3.1.1. Aspiration window search. Aspiration window search stipulates the initial window is $(\mathrm{V}-\mathrm{e}, \mathrm{V}+\mathrm{e})$, where V is the estimated value of the position and e is the expected error limit, rather than $(-\infty, +\infty)$ used in the original Alpha-Beta algorithm. If the actual score of a position lies within the window of $(V-e, V+e)$, which is expected, it will not only return the correct result but also prune lots of subtrees since the initial window is much narrower than that of Alpha-Beta algorithm so that there is a good chance to find some subtrees useless. In contrast, if the actual score does not lie within the window of $(V-e, V+e)$, this subtree needs to be re-searched. In this case, the initial alpha is reset to $-\infty$ when the actual score is less than $\mathrm{V}-\mathrm{e}$ (failing low) while the initial beta is reset to ∞ when the actual score is greater than $\mathrm{V}+\mathrm{e}$ (failing high).

Such re-searching increases time consumption a lot. The ordering of moves is quite important to aspiration window search and it performs nearly perfectly in the case of searching strongly ordered trees. It is better to use the aspiration window search together with the iterative deepening and transposition tables. Alpha-Beta with aspiration windows is more effective than Alpha-Beta in most cases [4-5].

3.1.2. Minimal window search. The core property of minimal window search [6] is that to prove a subtree inferior is faster than to confirm its exact score. It is based on the supposition that the move that is going to be searched is the best move and other moves are inferior. The case that all other moves prove to be inferior is highly expected, but if it does not happen, a new supposition continues until the real best move is found. This approach performs significantly well, especially when used together with reordering mechanisms. There are two typical algorithms using minimal windows: principal variation search and memory-enhanced test driver.

Principal variation search (PVS), equivalent to Negascout search, is an algorithm based on minimal windows. It searches the first subtree with a full window (alpha, beta) and gets the exact score V of the first subtree. Then it searches other subtrees with null windows (V, V+1) to find whether there are moves that fail high, which means they are better than the first move. If a move is found to be better, re-searching for the subtree of this move is needed to get the exact score of this move and continue similar verifications. According to reference [7], PVS runs faster than aspiration window search. One of the properties of PVS is that the tree searched by it is asymmetrical because PVS always tries the best moves first.

MTD(f) is short for Memory-enhanced Test Drive [8]. Unlike PVS performs a full window search, MTD(f) completely uses null windows for searching, so it in many cases can outperform PVS. It is based on the idea to guess the Minimax value iteratively. The score V for a position starts with a value guessed to be the best or closest to the actual score. Then the algorithm uses a null window (V, V+1) to search the tree and get the information about how to adjust V due to whether the result fails high or low. Furthermore, since it does lots of re-searches, the algorithm has to use a transposition table to retrieve information of the same position. Although the algorithm re-searches the same nodes quite many times, it is still pretty much cheaper to do a null window search than do a full window search.

3.2. Enhancements based on results from previous searching

In a complex and huge game-tree, there is a high probability that there are several nodes having the same position, so it is not uncommon to encounter a position that is identical or similar to a position searched previously. Therefore, for better performance, it is attractive to use previous information that has already

DOI: 10.54254/2755-2721/32/20230221

been got without searching again. Several most popular techniques based on this idea and discussed in this paper are iterative deepening, transposition table and refutation table, and killer heuristic and history heuristic [6].

3.2.1. Iterative deepening. This method searches game-trees by iteration with a depth limit D which starts with 1 ply and at the beginning of each iteration, the limit is extended by 1 ply. The main idea of iterative deepening is to do better moves in D-ply search based on (D-1)-ply search. Since there is a high probability that the principal variation of (D-1)-ply search is a prefix of the principal variation of D-ply search and even further searches, first trying or examining the best moves of the last iteration is a good strategy. Moreover, the results of the last iteration can be used to set up alpha and beta in this iteration. Even though, the most superior point of iterative deepening is that it can give final results with high accuracy at any time. This is very useful in time-limited tests. It is usually used with transposition tables or refutation tables.

3.2.2. Transposition table and refutation table. The transposition table uses the idea of hash tables to store information as a large direct access table. When a node is reached, if the previous search of the position of this node reached the desired depth or height, the previously stored score of it can be directly used and further search of this subtree can be stopped. Otherwise, not the score but the best move of it can be used for subsequent faster searching. This optimization is quite significant for searching, especially for iterative deepening, in terms of time.

However, the memory consumption of the transposition table is huge since it records every position searched. An alternative to a transposition table is a refutation table, whose space complexity is merely $O(d \cdot b)$ or so. Often used with iterative deepening, it records the principal variation or continuation of each iteration and direct moves of the next iteration. According to [7], using refutation tables improve iterative deepening by roughly 30% in terms of time.

3.2.3. Killer heuristic and history heuristic. Killer heuristic is a method used for dynamically re-ordering moves in a search. At each ply, the best moves, which can cause a cutoff or get an excellent score, are recorded. Based on the thought that a move that is the best move for a position may also be the best move for another position that is similar to the previous one at the same ply, this approach would first search the best moves recorded previously in the hope of making cutoffs. Although it seems to be pretty good, it does not perform very well in practical implementation due to its critical uncertainty.

History heuristic is a generalization of killer heuristic. The judgements of best moves for both techniques are similar. Killer heuristic records only one or two best moves of each ply independently while history heuristic records all best moves of the entire search tree, no matter which ply the move is at. A killer move for killer heuristic has a high probability to be one of the best moves for history heuristic. Also, the history heuristic proved to have a better performance than the killer heuristic.

4. Monte-Carlo tree search

Since 2006 when Monte-Carlo tree search (MCTS) got unprecedented attention for searching game-trees, it has become a powerful technique used for complex AI games, such as Go and chess, and made a big breakthrough in the field of game-trees due to the huge difference between it and traditional Alpha-Beta algorithm [9]. With the use of MCTS and other AI implements, AlphaGo beat professional human Go players, which was a milestone for the whole field. Today MCTS is still under research for further improvements and other inspirations.

4.1. Mechanisms and characteristics

MCTS uses best-first search with a random sampling of the search space. It looks for the best move iteratively. There are four phases that are performed repeatedly in an iteration (Figure 2):

DOI: 10.54254/2755-2721/32/20230221

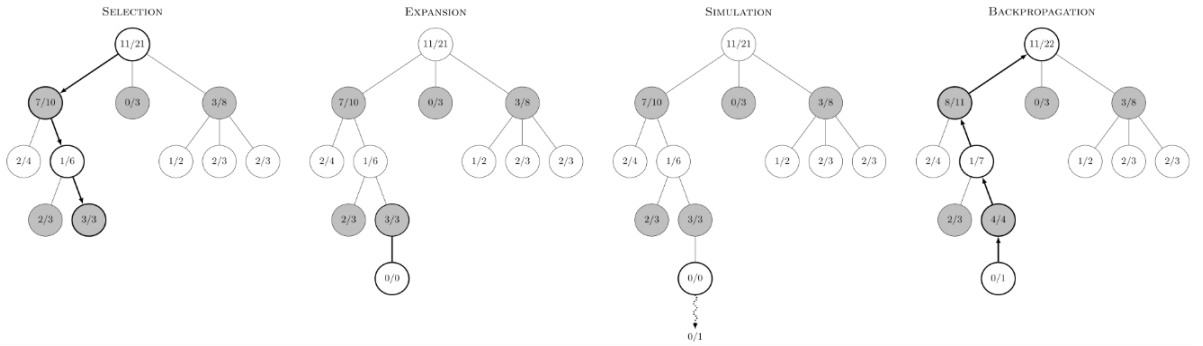


Figure 2. Four steps of monte-carlo tree search [10].

A) Selection: Start from the current root node and search the tree with a bias of visiting more promising moves until it reaches a leaf node, which is the terminal node at the current state but still has potential successors that have not been visited.

B) Expansion: Add one or more successors of the leaf node found in the selection phase to the search tree.

C) Simulation: Perform a playout (also called roll-out), which refers to a process of constantly choosing random moves to reach the very end of the game, from the new node to produce a sample outcome.

D) Backpropagation: Update back father nodes recursively with the simulation result.

Such repetition terminates when the computational limitation, maybe time or memory, is reached and an action would be applied before the next iteration.

In order to achieve the best performance, it is essential to balance exploitation (focusing on superior moves) and exploration (focusing on inferior moves), because there is a big chance that some moves are inferior now but superior when the search goes deeper, which are called “trap states”. Many techniques can be used for this purpose, such as bandit-based methods including upper confidence bounds (UCB), which is used to develop the upper confidence bounds for trees (UCT) algorithm, the most popular variant of MCTS.

The first characteristic of MCTS is that MCTS can return results at any time, just like iterative deepening, since MCTS also uses iteration and all information keeps updated. Moreover, the other one is that the tree generated by MCTS is asymmetric because it prefers better moves like principal variation.

4.2. Comparison with Alpha-Beta algorithm and advantages and disadvantages

Unlike the Alpha-Beta algorithm, which needs to determine exact position scores for all nodes, MCTS uses randomization to predict which is better. Alpha-Beta algorithm is based on depth-first search while MCTS is based on best-first search. According to several experiments, MCTS reaches better performance than the Alpha-Beta algorithm and other similar algorithms.

The biggest advantage of MCTS is that it can be applied to search trees without any domain-specific knowledge (human knowledge). That is also the reason why it can make a big success in Go, because the evaluation function in the Alpha-Beta algorithm needs domain-specific knowledge and a position can be evaluated correctly in chess but is very difficult to be evaluated in Go. Furthermore, it turns out that MCTS can also perform well in chess games. And this advantage has been magnified due to the use of convolutional neural networks (CNN) [11].

One obvious disadvantage is that even though there are methods to deal with exploitation-exploration dilemmas, it is still possible that MCTS misses better moves due to “trap states” and this might be the reason for its failures in competing against human players. Another disadvantage is that it is possible that sometimes playouts consume pretty much time because random moves may lead to deadlocks.

DOI: 10.54254/2755-2721/32/20230221

5. Conclusion

Different kinds of popular techniques for searching game-trees have been discussed in terms of their mechanisms, advantages, shortcomings, and practical performances which are influenced by time complexities, space complexities, and whether the trees are strongly ordered or random. And they are divided into different categories based on their essences. Brute-force algorithms, pure Minimax and Negamax algorithms, are only used in very simple games. For other complex games, the Alpha-Beta algorithm must be used, usually together with the iterative deepening and transposition tables or refutation tables, sometimes also with the history heuristic. MTD(f) outperforms PVS and PVS outperforms aspiration search. MCTS with AI technology is now widely used and is more used in Go than in any other game.

Currently, some algorithms of machine learning, such as regression and the neural network, have been combined with game-tree search algorithms. Despite they are used with improvement methods, some algorithms are still sometimes unstable and can be enticed by human players to go into losing moves. Future research about game-trees can try to make these algorithms with as much certainty as possible and can define some special human-strategies and transform them into a form that computers can understand.

References

[1] Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development, 3(3), 210–229.

[2] Heineman G T, Pollice G, Selkow S. Algorithms in a nutshell: A practical guide. " O'Reilly Media, Inc.", 2016.

[3] Russell, Stuart J.; Norvig, Peter. (2021). Artificial Intelligence: A Modern Approach (4th ed.). Hoboken: Pearson. pp. 149–150.

[4] Kaindl H, Shams R, Horacek H. Algorithms with and without Aspiration Windows. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13(12).

[5] Shams R, Kaindl H, Horacek H. Using Aspiration Windows for Minimax Algorithms, (International Joint Conference On Artificial Intelligence. 1991: 192-197.

[6] Elnaggar A A, Gadallah M, Aziem M A, et al. A comparative study of game tree searching methods. International Journal of Advanced Computer Science and Applications, 2014, 5(5): 68-77.

[7] Marsland T A, Campbell M. Parallel search of strongly ordered game trees. ACM Computing Surveys, 1982, 14(4): 533-551.

[8] Tommy L, Hardjianto M, Agani N. The analysis of alpha beta pruning and MTD (f) algorithm to determine the best algorithm to be implemented at connect four prototype. IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2017, 190(1): 012044.

[9] Świechowski M, Godlewski K, Sawicki B, et al. Monte Carlo tree search: A review of recent modifications and applications. Artificial Intelligence Review, 2023, 56(3): 2497-2562.

[10] https://upload.wikimedia.org/wikipedia/commons/thumb/2/21/MCTS-steps.svg/1200px-MCTS-steps.svg.png.

[11] Mastering the game of Go with deep neural networks and tree search. Nature, 2016, 529(7587): 484-489.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230222

House price prediction using machine learning: A case study in Seattle, U.S.

Jiapei Liao

Faculty of Science and Technology, University of Macau, Macau, 999078, China

Corresponding author: dc12784@umac.mo

Abstract. In recent years, predicting housing prices has become a prominent research topic. The motivation behind this research is the lack of precise analysis and comprehensive comparison of different machine learning models for Seattle housing prices. To address this issue, this paper utilizes a dataset obtained from Kaggle, consisting of Seattle housing prices. This study focuses on analyzing and comparing the performance of different machine learning models in predicting housing prices in Seattle. After preprocessing the data, eight different machine learning models are applied to predict housing prices in Seattle. A comprehensive comparison of these models is conducted to analyze their differences. The experimental results show that the differences in performance among the models are not substantial. However, StackingAveragedModels emerges as the top-performing model, with an RMSLE of 0.2328 and an R2 of 0.7771. These findings contribute to a better understanding of the predictive capabilities of different machine learning models for Seattle housing prices.

Keywords: Machine Learning, Seattle, Housing Price Forecast, StackingAveragedModels.

1. Introduction

House price prediction has become a popular research field, particularly with the advancement of artificial intelligence and the widespread use of machine learning algorithms [1]. Accurate house price prediction holds significant importance for real estate professionals, potential homebuyers, as well as government policies and urban planning, among other stakeholders [2]. It can greatly assist individuals in making informed decisions. Consequently, there is a need for a high-precision real estate price prediction model. House prices are influenced by various factors, such as geographical location, house size, and land area. By leveraging machine learning algorithms, reliable predictions can be obtained.

Homeownership is highly valued by Americans, as owning a home is seen as both a tool for economic mobility and a symbol of social status. Having a home can deepen one's sense of security and provide the flexibility to move to more desirable neighborhoods. Therefore, for American families, considering housing options is of great significance.

This paper utilizes Seattle, a prominent seaport city on the west coast of the United States, as a prime example to forecast its housing prices. Seattle serves as the vibrant seat of King County, Washington, and stands as the most populous city in the state of Washington and the entire Pacific Northwest region of North America.

However, Seattle's notorious traffic congestion ranks among the worst in the United States [3]. In recent years, the city's real estate market has been booming, with both property prices and rents

DOI: 10.54254/2755-2721/32/20230222

experiencing significant increases. The high cost of housing has resulted in a growing homelessness issue, making Seattle the second-largest city in terms of homelessness, just behind Los Angeles. The study aims to employ machine learning techniques in constructing a promising house price prediction model for Seattle.

A related research report on Kaggle trained and predicted the same dataset, yielding favorable results. However, it did not conduct a comparative analysis of multiple algorithm models. Despite the maturity of machine learning algorithms, there has not been a sufficiently in-depth analysis of the variations among different models. Therefore, this paper selects eight commonly used machine learning models to conduct an in-depth study of the Seattle housing dataset. A comprehensive comparison and analysis of these eight models were conducted to determine the best-performing model.

2. Method

In this study, the training and test sets were sourced from the Kaggle platform. The data indicators were subjected to an analytical examination, followed by preprocessing of the input data. Subsequently, an ensemble of machine learning models, including linear regression (Lasso), ElasticNet, KernelRidge (KRR), Random Forest (RF), GBoost, and XGBoost regression (XGB), were constructed and trained. Additionally, two ensemble approaches, namely AveragingModels and StackingAveragedModels, were employed. The resulting models were subsequently utilized to generate outcomes for further analysis.

2.1. Exploratory data analysis

According to previous research, the dataset's quality significantly impacts research outcomes. In this study, we utilized the Seattle Housing dataset from Kaggle [4], which consists of real housing price data for homes sold in Seattle, Washington. The variable targeted for prediction is the house price, while the independent variables consist of beds, baths, size, total land area, and zip code. The Seattle Housing Dataset comprises data about 2515 single family houses sold in Seattle, Washing, USA between August and December 2022. Each sample in the dataset consists of 8 features (shown in Table 1).

Table 1. Description of features in the Dataset

Features	Description
beds	The number of bedrooms within the house
baths	The count of bathrooms available in the property. Please note that a value of 0.5 indicates the presence of a half-bath, which includes a sink and toilet but lacks a bathtub or shower
size	The overall floor area of the property
size_units	The units in which the previous measurement is expressed
lot_size	The complete land area on which the property is situated. The ownership of the lot resides with the homeowner
lot_size_units	Units of the previous measurement
zip_code	A postal code utilized in the United States
price	The selling price of the property in US dollars

First, the normal distribution plot of the "price" variable and the quantile-quantile plot are drawn. (Figure 1). By examining the quantile-quantile plot, we can determine whether the "price" variable follows a normal distribution.

As illustrated in Figure 1, it is evident that the "price" variable deviates from a normal distribution. Therefore, in order to address this non-normality, a logarithmic transformation has been applied to the data. The transformation involves taking the natural logarithm of the values, using the formula log(1+x). This will generate a new set of transformed data. Subsequently, we can draw the normal distribution plot and the quantile-quantile plot (Figure 2) for the transformed data. In the quantile-quantile plot, we would expect the red line and the blue points to closely coincide, indicating conformity with the normal distribution. By comparing the quantile-quantile plot in Figure 1 with that in Figure 2, it can be observed

DOI: 10.54254/2755-2721/32/20230222

that after the logarithmic transformation, the red line aligns more closely with the blue points. This indicates that the data is more in line with a normal distribution.

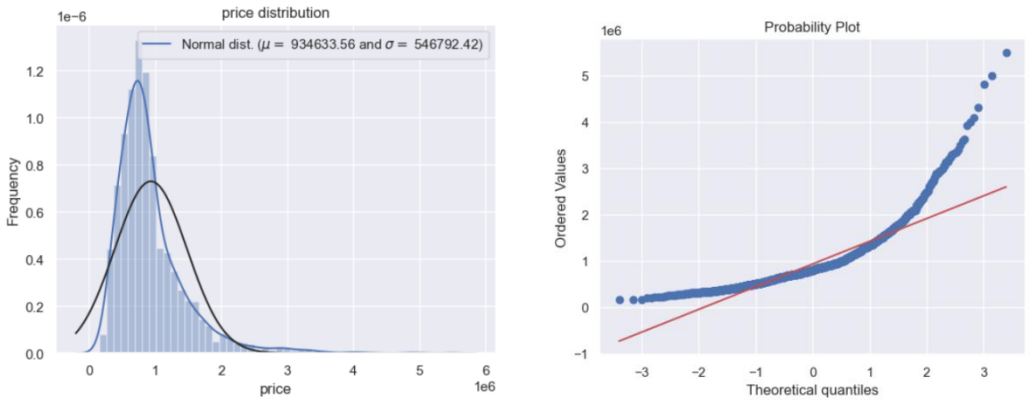


Figure 1. Normal Distribution Plot and Quantile-Quantile Plot of the 'Price' Variable

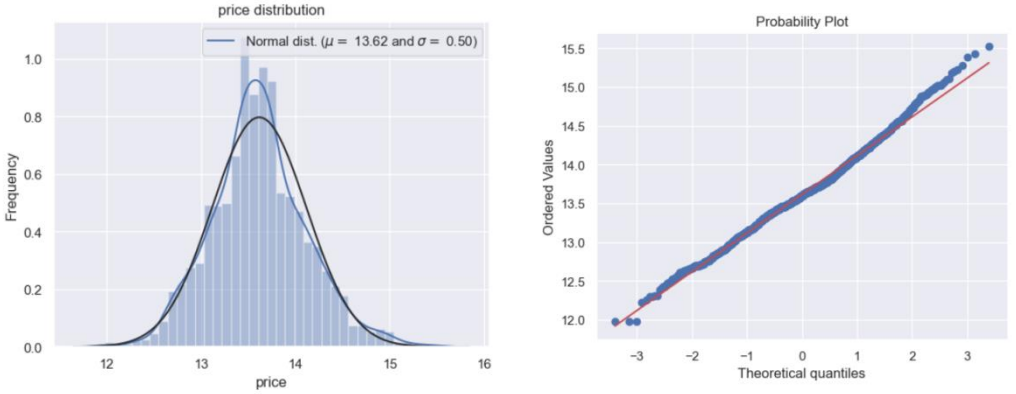
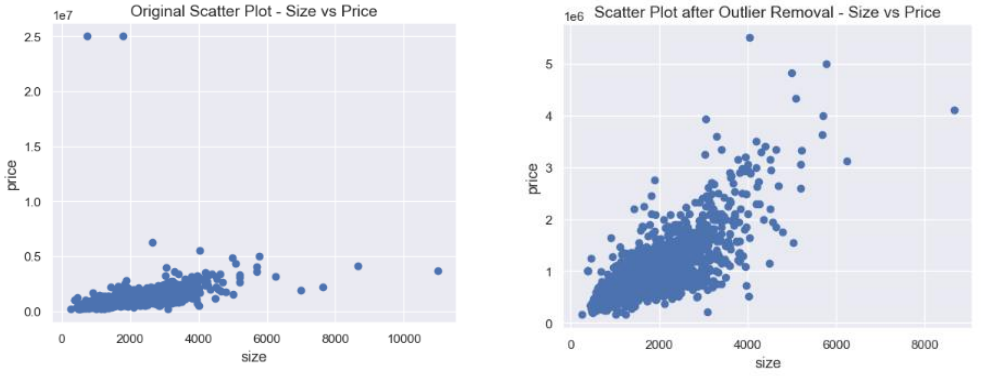


Figure 2. Normal Distribution Plot and Quantile-Quantile Plot of Logarithmic Transformed Data"

To analyze the relationship between each variable and the "price" variable, scatter plots were generated. Firstly, in Figure 3, a scatter plot of "size" and "price" was created using the scatter function. Upon examination, outliers were identified and subsequently removed, resulting in a revised scatter plot.

Next, a scatter plot of "lot_size" and "price" was generated using the scatter function. Similarly, outliers were detected and eliminated, leading to an updated scatter plot.



DOI: 10.54254/2755-2721/32/20230222

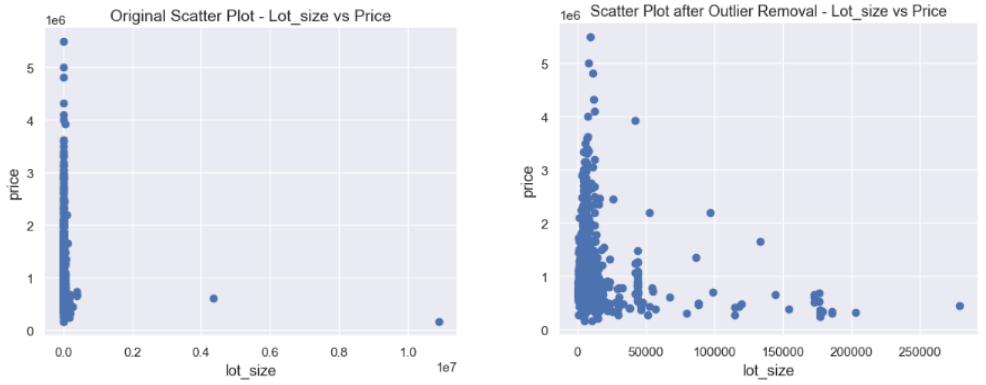


Figure 3. Scatter Plots Analyzing Relationships with the 'Price' Variable

After removing outliers, a heat map of the data was generated (Figure 4). The heat map illustrates the correlation between different variables in the dataset at this stage. It indicates that "beds", "baths" and "size" exhibit a moderate correlation with the "price" variable. This finding implies that if a multiple linear regression model is employed, the issue of multicollinearity needs to be taken into account. Multicollinearity refers to a high correlation between predictor variables, which can cause problems in interpreting the coefficients and stability of the model. To address this, appropriate techniques such as feature selection, dimensionality reduction, or regularization methods may be employed.

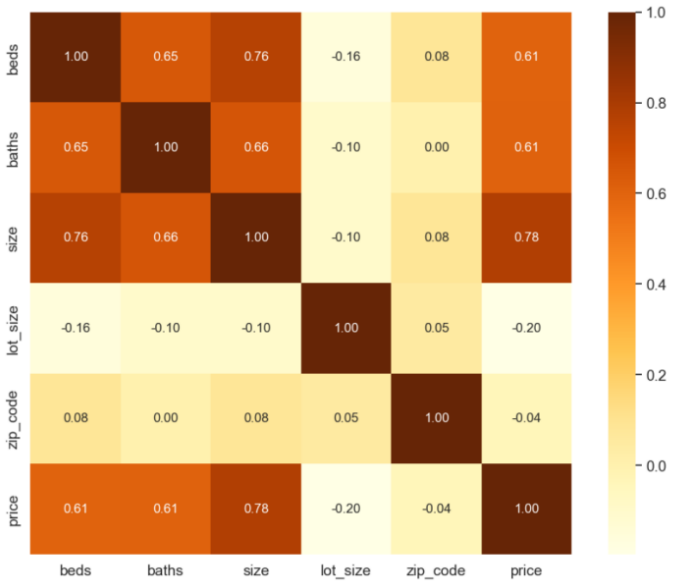


Figure 4. Heatmap illustrating the correlation between variables after logarithmic transformation

2.2. Data preprocessing

To obtain the final dataset, the training set and test set were combined. Upon inspecting the data, it was observed that there are missing values present in the dataset. Specifically, both "lot_size_unit" and "lot_size" contain missing values. In this case, the missing values in "lot_size" were filled with the median value of the "lot_size" column. Additionally, the missing values in "lot_size_unit" were represented as "sqft" units.

For variables that are not continuous, they were converted into categorical values. On the other hand, for variables that are not categorical, their skewness was examined. It was noted that four data features still

DOI: 10.54254/2755-2721/32/20230222

exhibited significant skewness. To address this, a Box-Cox transformation was applied with a lambda (λ) value of 0.15 to transform the data with skewness greater than 0.75.

Finally, all columns in the dataset were individually one-hot encoded, and the results were stored back in the "all_data" dataset. By performing one-hot encoding, the categorical variables such as "zip_code," "size_units," and "lot_size_units" were converted into multiple binary feature columns.

2.3. Model selection and construction

In this study, several classical machine learning algorithms and ensemble learning models were employed to predict housing prices. Firstly, a linear regression model was chosen. To address the issue of multicollinearity, two commonly used linear regression methods, Lasso regression and ElasticNet regression, were applied. These methods incorporate regularization terms during the model training process, thereby reducing the number of features or driving their coefficients towards zero through penalty terms on the model parameters. For nonlinear regression, the KernelRidge algorithm was utilized. As for ensemble learning models, three typical approaches were selected: Random Forest, Gradient Boosting, and XGBoost. To facilitate performance comparison and further analysis, two combined model algorithms were developed. AveragingModels involved averaging the prediction results of multiple models, while StackingAveragedModels utilized the predictions from multiple base models as inputs for metamodels. These ensemble methods aim to enhance prediction accuracy and robustness.

These models are introduced respectively:

- **Lasso:**

Lasso regression, also known as L1 regularization, is a technique that introduces a penalty term into the traditional linear regression model. It aims to provide an interpretable model while effectively managing the risk of overfitting. By adding the penalty term, Lasso encourages sparsity in the model by forcing some coefficients to be exactly zero. This promotes feature selection and helps in obtaining a simpler and more understandable model.

The optimization objective for Lasso can be expressed as:

$$\text{Lasso} = (1/(2 \times n\_samples)) \times \|y - Xw\|^2\_2 + alpha \times \|w\|\_1 \quad (1)$$

The first item measures the squared difference between the true target. values (y) and the predicted values based on the feature matrix (X) and the weight vector (w). alpha is the regularization parameter that controls the strength of the penalty term. In this experiment,the value of alpha is set to 0.0005 for the Lasso regularization.

- **ElasticNet:**

ElasticNet is a regularization technique that combines the L1 regularization of Lasso and the L2 regularization of Ridge regression [5].

Minimizes the objective function:

The optimization objective for ElasticNet can be expressed as:

$$\text{ElasticNet} = (1/(2 \times n\_samples)) \times \|y - Xw\|^2\_2 + alpha \times \left((1 - l1\_ratio)\right) \times \|w\|\_2^2/2 + l1\_ratio \times \|w\|\_1 \quad (2)$$

In this experiment, the value of alpha is set to 0.0005. l1-ratio is set to 0.9, indicating a higher emphasis on L1 (Lasso) regularization. The first term measures the squared difference between the true target values (y) and the predicted values based on the feature matrix (X) and the weight vector (w). The second term represents the regularization term, which consists of two parts. The first part, $\left((1 - l1\_ratio)\right) \times \|w\|\_2^2/2$ _1 corresponds to the L2 regularization (Ridge), and the second part, $l1\_ratio \times \|w\|\_1$, corresponds to the L1 regularization (Lasso). The alpha parameter controls the overall strength of the regularization. By adjusting the values of alpha and $l1\_ratio$, ElasticNet allows for flexible regularization and can effectively handle situations where there are both correlated and uncorrelated features.

- **KernelRidge:**

DOI: 10.54254/2755-2721/32/20230222

Kernel ridge regression (KRR) combines ridge regression, which applies L2-norm regularization to linear least squares, with the use of kernels [6]. This integration allows KRR to model a function within a transformed space determined by the specific kernel. By employing various kernels, KRR can capture complex relationships within the original feature space. In this experiment, the value of alpha is set to 0.6 'polynomial' is chosen as the kernel function.The degree parameter for the polynomial kernel.is set to 2. The constant term in the polynomial kernel function is set to 2.5.

- **Random Forest:**

Random Forest is an algorithm that uses ensemble learning to combine multiple decision trees and perform predictions. It leverages the concept of bagging, where each tree is trained on a randomly selected subset of the training data with replacement. Additionally, Random Forest introduces additional randomness by randomly selecting a subset of features at each split point.

The algorithm's main objective is to reduce overfitting and improve prediction accuracy. By averaging individual tree forecasts, Random Forest decreases variance and provides more robust predictions compared to a single decision tree.

To assess the importance of variable x_j in predicting Y within a Random Forest, the computation involves summing the weighted impurity decrease $p(t)\Delta i(s_t,t)$for all nodes t where x_j is utilized [7]. This calculation is then averaged across all trees φ_m (indexed as m=1,...,M) in the forest:

$$\mathrm{Imp}(X_j) = \frac{1}{M}\sum_{m=1}^{M}\sum_{t\in\varphi_m} 1(j_t = j)[p(t)\Delta i(s_t,t)] \quad (3)$$

- **Gradient Boosting**

Gradient Boosting is an overall learning algorithm which combines several low learners to create a powerful predictive model [8].

The algorithm's primary objective is to minimize a predefined loss function by iteratively fitting new models to the residual errors of the ensemble. Each iteration focuses on the samples that were predicted incorrectly, assigning higher weights to these samples to prioritize their correct classification or prediction. This iterative process steadily enhances the predictive performance of the overall model by diminishing the remaining errors.

The optimization objective in Gradient Boosting can be formulated as follows:

$$F(x) = F_{M-1}(x) + \eta \times f\_m(x) \quad (4)$$

In this experiment, learning rate is set to 0.05, indicating a relatively small learning rate. The number of boosting stages or regression trees is set to 3000. $F(x)$ represents the ensemble model's prediction, $F_{M-1}(x)$ is the prediction from the previous ensemble (M-1) models, η is the learning rate that controls the contribution of each weak model, and $f\_m(x)$ is the weak learner's prediction for the current iteration (m).

- **XGBoost Regression**

XGBoost Regression, an efficient implementation of gradient boosting, has garnered substantial acclaim in both machine learning competitions and practical use cases [9].

Similar to Gradient Boosting, XGBoost Regression builds an ensemble model by sequentially adding weak learners, typically decision trees, to correct the errors made by previous models. However, XGBoost Regression introduces several enhancements, including regularization techniques, parallel processing, and tree pruning, to improve both the accuracy and efficiency of the model [10].

The optimization objective in XGBoost Regression is defined as follows:

$$L(\varphi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (5)$$

In XGBoost Regression, the optimization objective is defined using a customizable differentiable convex loss function$l$that quantifies the difference between the predicted values $\hat{y}_i$ and the target values y_i. The regularization term Ω is introduced to penalize the complexity of the model, discouraging overfitting. In this experiment, the value of the L1 regularization term on the weights is 0.4640, while the L2 regularization term on the weights is set to 0.8571.

DOI: 10.54254/2755-2721/32/20230222

Due to the involvement of functions as parameters, XGBoost trains equation in an additive manner, as traditional optimization methods in Euclidean space are not applicable for solving this problem [8].

- **AveragingModels**

AveragingModels is an ensemble learning technique that combines multiple individual models by averaging their predictions. By training multiple models, AveragingModels captures different aspects and perspectives of the data, reducing the risk of relying too heavily on the biases of a single model.

In this paper, AveragingModels is applied by averaging the predictions of three base models: ElasticNet (ENet), Gradient Boosting (GBoost), and Kernel Ridge Regression (KRR). The individual predictions of these base models are combined using simple averaging. Additionally, a meta-model, specifically Lasso, is used to further refine the final prediction. This ensemble approach leverages the diverse strengths of the base models and combines their predictions to achieve improved overall performance.

- **StackingAveragedModels**

StackingAveragedModels is an ensemble learning technique that integrates multiple individual models hierarchically. It aims to leverage the strengths of different models by using their predictions as inputs for a meta-model, which produces the final prediction.

In StackingAveragedModels, multiple base models, such as ElasticNet (ENet), Gradient Boosting (GBoost), and Kernel Ridge Regression (KRR), are trained on the same dataset or different subsets of the data. Each base model generates predictions for the target variable. The following inputs need to be provided for the implementation of StackingAveragedModels.: base_models: A list containing multiple base models that will be used to generate predictions. meta_model: The meta-model used to combine the predictions from the base models.n_folds: The number of folds for cross-validation, with a default value of 5.

3. Result

3.1. Experimental Details

The experiment encompassed the importation of the Kaggle dataset into the Jupyter platform, followed by the execution of a sequence of Python code operations to obtain outcomes for multiple machine learning models. The experiment was conducted employing an NVDIA GeForce RTX 3060 Laptop GPU and an AMD Ryzen 7 5800H CPU with Radeon Graphics.

The training parameters were determined using cross-validation, a common technique in machine learning, to ensure better performance and generalization of the house price prediction models. For this experiment, a cross-validation fold of 5 was utilized (n_folds = 5), indicating that the dataset was divided into five subsets for training and validation purposes. The choice of this value aimed to strike a balance between model complexity and computational efficiency.

3.2. Evaluation metrics

In evaluating these models, the chosen parameters for assessment are RMSLE (Root Mean Squared Logarithmic Error) and R^2 (Coefficient of Determination). The logarithmic transformation of the original data renders RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percentage Error) less informative on the logarithmic scale. Therefore, only RMSLE and R^2 were selected as the evaluation metrics for the models.

The two evaluation methods are introduced:

Root Mean Squared Log Error (RMSLE):

$$\text{RMSLE} = \sqrt{\frac{1}{\text{N}}\sum_{i=1}^{\text{N}}[\log(y_i + 1) - \log(\hat{y}_i + 1)]^2} \tag{6}$$

RMSLE serves as a widely adopted metric for evaluating regression model performance. In this equation, 'N' represents the number of samples, 'y_i' represents the predicted values by the model, and '$\hat{y}_i$' represents the actual target variable values. A lower value of RMSLE indicates better predictive capability of the model

DOI: 10.54254/2755-2721/32/20230222

R^2:

R2 (R-squared) serves as a commonly adopted metric to assess the fitting of regression models to the data, representing the proportion of the variance in the dependent variable explained by the model.

$$R^2 = \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

where $\sum_{i=1}^{N}(y_i - \bar{y})^2$, $\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2$, $\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$ denote the total sum of squares (TSS), the explained sum of squares (ESS) and residual sum of squares (RSS) respectively.

3.3. Model Evaluation

Table 2. the performance of different models

Model	**RMSLE**	**R^2**
Lasso	0.2592	0.7293
ElasticNet	0.2592	0.7294
KernelRidge	0.2540	0.7383
Random Forest	0.2484	0.7473
Gradient Boosting	0.2343	0.7746
XGBoost Regression	0.2353	0.7765
AveragingModels	0.2434	0.7593
StackingAveragedModels	0.2328	0.7771

Among all the models, due to the application of the Box-Cox transformation on the dataset, the Linear Regression model exhibited overfitting, resulting in significantly large values for R-squared (R2) and RMSLE. To address this issue, these models introduce constraints through regularization techniques to alleviate overfitting (Table 2).

Comparing the results, we can make the following observations:

Based on the evaluation results (), StackingAveragedModels stand out as the top-performing models with the lowest RMSLE values (0.2328) and highest R^2 values (0.7771), indicating better predictive performance in terms of minimizing the logarithmic error. By leveraging the strengths and complementary characteristics of these diverse models, StackingAveragedModels can achieve improved predictive performance.

Lasso, ElasticNet, and KernelRidge demonstrate similar performance, with slightly higher RMSLE values but still within a close range. The relatively underperforming nature of Lasso, ElasticNet, and KernelRidge could be attributed to several factors. One possible reason is that Lasso and ElasticNet rely on L1 regularization, which can lead to feature selection and result in a more simplified model that may not capture all the important relationships in the data. This feature selection process could potentially exclude relevant variables, leading to suboptimal performance. On the other hand, KernelRidge may not be as flexible in capturing complex non-linear relationships present in the data, which can limit its predictive capability compared to other algorithms.As an ensemble model, Random Forest exhibit relatively higher RMSLE values, implying comparatively lower accuracy in predicting house prices. However, they still achieve reasonable results in terms of R-squared scores, indicating a moderate ability to explain the variance in the data. The subpar performance of the model may be attributed to the utilization of default hyperparameter configurations without exploring different parameter combinations to identify the optimal settings.

With relatively low RMSLE values and higher R-squared scores, Gradient Boosting and XGBoost Regression exhibit a satisfactory level of predictive capability. As a combination model, AveragingModels performs slightly worse than the top-performing models, with a higher RMSLE value. The main reason is that these model averages the results of KRR (KernelRidge) and ElasticNet models, which individually do not yield very optimal results. However, compared to the results achieved by KRR and ElasticNet as standalone models, AveragingModels optimizes the outcomes.

DOI: 10.54254/2755-2721/32/20230222

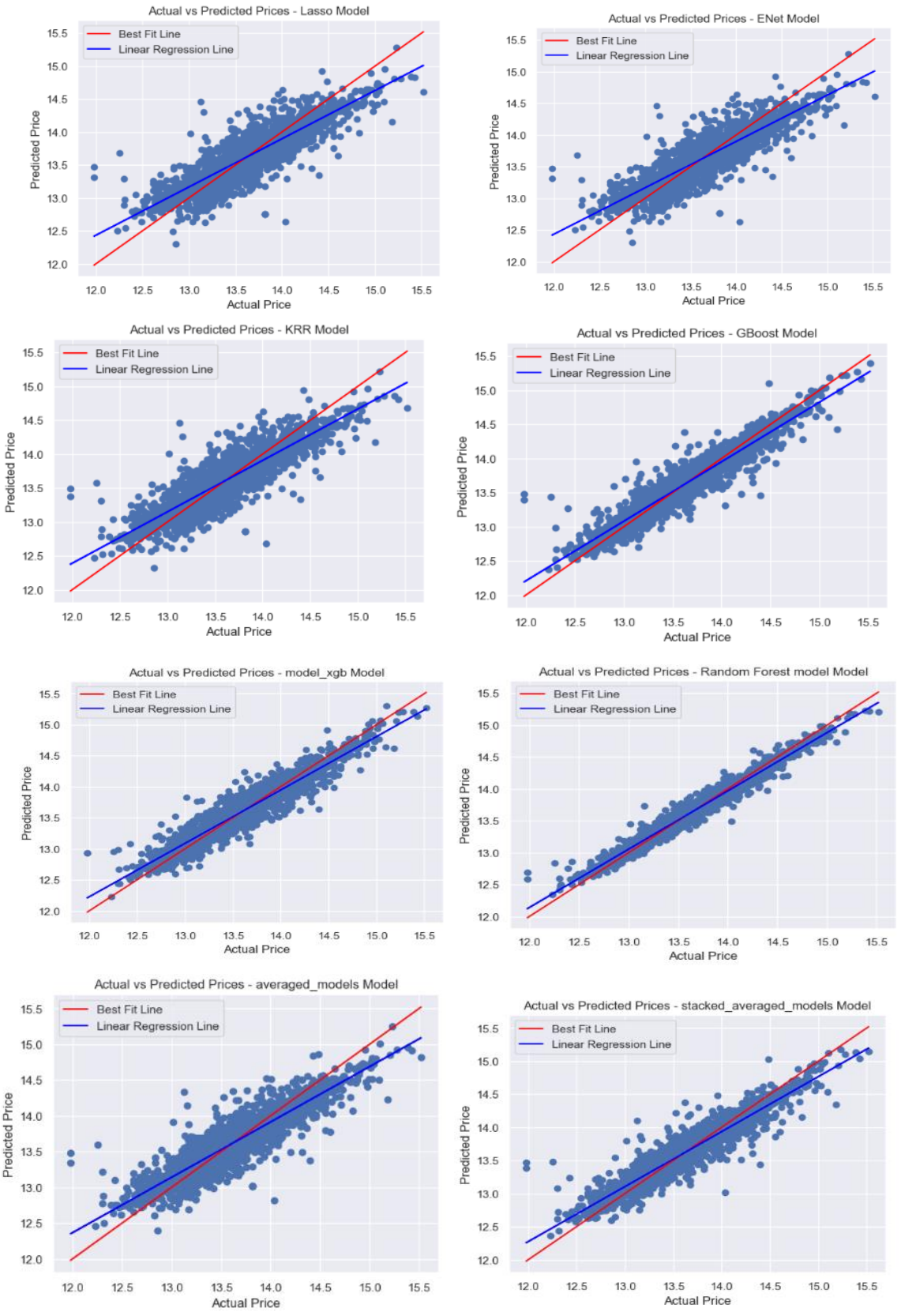


Figure 5. Actual Price vs. Predicted Price for All Models

DOI: 10.54254/2755-2721/32/20230222

3.4. Visual Analysis of Model Performance

To further analyze the performance of each model, a visual examination of the relationship between the actual and predicted values was conducted. In Figure 5, each point represents a record from the test set. The x-coordinate of the point represents the actual home value, while the y-coordinate represents the corresponding predicted home price. The red line in the plot represents the best fit line, indicating accurate predictions when the points overlap closely. The blue lines in each chart depict the linear regression lines that show the overall trend of the results, reflecting the relationship between the predicted and true values. This visual analysis allows for a comprehensive understanding of how well the models capture the actual variation in home prices. By examining the alignment of the points with the best fit line and the overall trend portrayed by the blue lines, insights can be gained regarding the accuracy and consistency of the predictions made by each model.

4. Conclusion

In this study, the aim was to predict housing prices in Seattle by employing eight different machine learning methods: Lasso, ElasticNet, KernelRidge, Random Forest, Gradient Boosting, XGBoost Regression, AveragingModels, and StackingAveragedModels. The performance of these models was compared using RMSLE and R2 as evaluation metrics, with a focus on analyzing the differences between the algorithms.

Among the evaluated models, Among the evaluated models, StackingAveragedModels, which combines the predictions of ENet, GBoost, and KRR through weighted averaging, exhibited the highest performance, achieving an RMSLE of 0.2328 and an R2 of 0.7771. Both of these metrics represent the best performance among all the experimental models. StackingAveragedModels can effectively leverage the advantages of each individual basic model, thereby improving the overall predictive capability. This is achieved by taking a weighted average of the prediction results from multiple basic models. By doing so, StackingAveragedModels combines the strengths of each model and enhances the overall prediction performance. This innovative ensemble model, which combines the predictions of multiple base models, was found to effectively reduce errors and yield superior results in predicting housing prices in Seattle.

This finding highlights the effectiveness of StackingAveragedModels in accurately predicting housing prices. This model demonstrates strong performance based on the selected evaluation metrics, which further supports their potential application in real-world scenarios. In the future, advancements in housing price prediction may benefit from applying deep learning techniques, which have the potential to provide more accurate forecasts. Therefore, further research and exploration into integrating deep learning methods into housing price prediction are crucial directions for future studies.

References

[1] Truong Q, Nguyen M, Dang H, et al. Housing price prediction via improved machine learning techniques. Procedia Computer Science, 2020, **174:** 433-442.

[2] Adetunji A B, Akande O N, Ajala F A, et al. House price prediction using random forest machine learning technique. Procedia Computer Science, 2022, **199**: 806-813.

[3] Cervero R. America's suburban centers: the land use-transportation link. Routledge, 2018.

[4] Cortunhas, S., “House Price Prediction Sattle” Kaggle Inc, (2023). https://www. kaggle.com/datasets/samuelcortinhas/house-price-prediction-seattle

[5] Huang Y, Schell C, Huber T B, et al. Traction force microscopy with optimized regularization and automated Bayesian parameter selection for comparing cells. Scientific reports, 2019, **9(1)**: 539.

[6] Suganthan P N. On non-iterative learning algorithms with closed-form solution. Applied Soft Computing, 2018, **70**: 1078-1082.

[7] Wu X, Yang B. Ensemble Learning Based Models for House Price Prediction, Case Study: Miami, US, 2022 5th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE). IEEE, 2022: 449-458.

DOI: 10.54254/2755-2721/32/20230222

[8] Sahin E K. Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. SN Applied Sciences, 2020, **2(7):** 1308.
[9] Chen T, Guestrin C. Xgboost: A scalable tree boosting system, Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016: 785-794.
[10] Bhattacharya S, Maddikunta P K R, Kaluri R, et al. A novel PCA-firefly based XGBoost classification model for intrusion detection in networks using GPU. Electronics, 2020, **9(2)**: 219.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230861

Advancements in robotics engineering: Transforming industries and society

Zuheng Bai

St. John Paul II High School, Hyannis MA, USA

tbai24@jpiihyannis.org

Abstract. Robotics engineering is a dynamic and multidisciplinary field that has undergone rapid evolution in recent years, transforming various aspects of our lives. This paper serves as a comprehensive exploration of robotics engineering, delving into its historical development, contemporary applications, and the exciting prospects that lie ahead. The history of robotics engineering is a fascinating journey, from the early automata and industrial robots to the cutting-edge autonomous systems and humanoid robots of today. It traces the evolution of technology and human ingenuity, showcasing how robotics has become an integral part of our modern world. Current applications of robotics engineering span a wide range of industries, from the manufacturing sector, where robots enhance efficiency and precision, to healthcare, where they assist in surgery and patient care. Furthermore, robots are venturing into hazardous environments, disaster response, and space exploration, expanding their utility. However, as robotics technology continues to advance, it raises ethical and societal questions. This paper will explore these concerns, such as the impact on employment, privacy, and human-robot interaction, emphasizing the importance of responsible development and use of robotic systems. In conclusion, robotics engineering is a transformative force, shaping industries and influencing society in profound ways. This paper seeks to shed light on its pivotal role in our world, inspiring discussions and responsible practices that ensure its positive impact on the future.

Keywords: Robotics Engineering, Transforming Industries, Transforming Society.

1. Introduction

Robotics engineering is a beacon of technological innovation, driving profound changes across industries and reshaping the fabric of our society. This remarkable field has evolved from its early industrial roots to cutting-edge advances in artificial intelligence (AI) and machine learning (ML). This article reviews the historical development of robotic engineering, discusses its contemporary applications, and looks forward to its promising future. This story not only chronicles the evolution of robotics but also highlights its integral role in shaping our world [1]. The roots of robotics engineering date back to the early 20th century, when inventors and engineers began experimenting with automatons and mechanical devices. However, the mid-20th century was a pivotal moment. The first industrial robots appeared. The Unimate, designed by George DeVol and Joseph Engelberger in the 1960s, was the first robot to perform precise tasks, marking the birth of manufacturing automation [2].

Johri, Singh, Sharma, and Rastogi discussed how technology advances, robots evolve from basic machines to highly complex systems [3]. The convergence of computer technology and artificial

DOI: 10.54254/2755-2721/32/20230861

intelligence paves the way for smarter, more adaptable robots. The 21st century has seen the rise of collaborative robots that work alongside humans, revolutionizing industries such as automotive and healthcare. Today, robotic engineering has applications in multiple fields and has left an indelible mark on its operations. In manufacturing, robots are key to efficiency, improving productivity and quality while reducing costs. In healthcare, surgical robots enable minimally invasive procedures, resulting in shorter recovery times and improved patient outcomes. Agriculture benefits from autonomous drones and robotic harvesters, solving the challenges of food production in a rapidly changing world. Logistics and transportation are also transformed, with robots playing a key role in warehouses and self-driving cars reshaping the future of transportation. Space exploration is no longer limited to human astronauts, rovers and autonomous probes will venture into the universe and uncover its mysteries.

2. Historical evolution of robotics engineering

2.1. Origins of robotics engineering

The history of robotics engineering is a captivating narrative that spans centuries, from the earliest sparks of human ingenuity to the cutting-edge technologies of today. This article embarks on a journey through time to explore the origins of robotics engineering, the key milestones that have shaped its development, and the interdisciplinary nature that fuels its innovation.

The seeds of robotics were sown in ancient civilizations. Ingenious inventors and engineers crafted automatons and mechanical devices that imitated human and animal movements. Among the earliest recorded automata was the "Water Clock" created by the ancient Egyptian engineer Ctesibius in the 3rd century BC [4]. This water-driven mechanism displayed a semblance of programmability, foreshadowing the concept of automation that lies at the core of robotics. Fast forward to the 20th century, and the birth of modern robotics engineering was imminent.

2.2. Key milestones in robotics development

The journey of robotics engineering has witnessed several key milestones that pushed the field forward. In the 1950s, George Devol and Joseph Engelberger introduced the world to the Unimate, the first industrial robot capable of performing repetitive tasks with precision. Unimate marks the dawn of manufacturing automation and catalyzes the robotics revolution in industry.

The 1970s saw the emergence of microprocessors and computer technology, which fueled advances in robotics. The pioneering work in robotics by researchers such as Joseph F. Engelberg, Marvin Minsky, and John McCarthy led to the development of intelligent robots capable of making decisions [5]. These robots are used not only in manufacturing but also in industries such as space exploration, healthcare, and agriculture. The era of collaborative robots, or "cobots," began at the turn of the 21st century. These machines are designed to work side by side with humans to increase productivity while ensuring safety. Collaborative robots are already being used in everything from automotive assembly lines to medical surgeries, highlighting the versatility of modern robotics technology.

2.3. The interdisciplinary nature of robotics

One of the defining characteristics of robotics engineering is its interdisciplinary nature. Robots span many disciplines, including mechanical engineering, computer science, electrical engineering, artificial intelligence, materials science, and even biological science. This interdisciplinary approach enables robotics engineers to create machines that mimic the complexity of the natural world [6]. For example, bionics is a branch of robotics inspired by nature. Engineers look to animals such as birds and insects for insights into flight and mobility, while soft robotics draws inspiration from the adaptability of biological organisms to design flexible, shape-shifting robots.

In summary, the historical evolution of robotics engineering is a testament to human innovation and the relentless pursuit of automation and efficiency. From ancient automata to today's intelligent machines, robotics engineering has traveled an extraordinary path. As we stand on the precipice of a future in which robots are likely to become even more integrated into our lives, it is important to

DOI: 10.54254/2755-2721/32/20230861

recognize the interdisciplinary nature of the field, where the convergence of different areas of knowledge will continue to drive innovation and shape the world of the future.

3. Technology military and defense

3.1. Unmanned aerial vehicles (UAVs)

The integration of robotics technology has changed the military and defense landscape, enhancing capabilities, reducing risks, and expanding the scope of operations [7]. Unmanned Aerial Vehicles (UAVs), commonly known as drones, are one of the most important contributions to this field. Military reconnaissance, surveillance, and combat strategies have been revolutionized by these drones. UAVs provide real-time intelligence by providing a bird's-eye view of the battlefield. Equipped with high-resolution cameras, sensors and even weaponry, they can locate and attack targets with extreme precision. This technology not only minimizes risk to human soldiers but also allows for rapid and effective response to evolving threats. In modern warfare, UAVs have become an indispensable asset, enabling both offensive and defensive operations.

3.2. Ground and underwater robotics

Robotics transcends the skies and enters the land and sea realms. Ground robots are designed to perform a variety of tasks, from bomb disposal to reconnaissance in urban environments [8]. Early Concepts, and What It Really Takes to Explore Alien Skies. Drifting on Alien Winds: Exploring the Skies and Weather of Other Worlds, 22-41. These machines can traverse treacherous terrain, clear obstacles, and even engage hostile forces if necessary. By doing so, they protect human soldiers and reduce the dangers associated with high-risk missions. Underwater robotics is equally transformative. Remotely Operated Vehicles (ROVs) and Autonomous Underwater Vehicles (AUVs) are playing a vital role in naval operations. These robots explore the ocean depths, inspect ship hulls, and detect and neutralize underwater mines. They can also be used to survey underwater infrastructure and collect data critical to national security.

3.3. National security implications

The application of robotics in military and national defense is of great national security significance. The ability to deploy robots for surveillance and reconnaissance enables advanced threat assessment and strategic planning. By minimizing the dangers faced by humans, these technologies reduce casualties, a key consideration in modern warfare. Additionally, robotics can enhance the capabilities of armed forces, making them more agile and effective. The precision and accuracy of robotic systems can reduce collateral damage and civilian casualties, which is not only critical for ethical reasons but also for international relations and public perception.

However, ethical and legal issues also arise with the rise of robotics in the military. Discussions about the responsible use of these technologies have been fueled by concerns that robotic autonomy could lead to unintended consequences and harm to civilians. International protocols and conventions are constantly evolving to address these issues and provide guidance on their use. Ultimately, the application of robotics in military and defense represents a major leap forward in modern warfare. Drones, ground robots and underwater systems have proven invaluable in enhancing combat capabilities and minimizing risk to human soldiers. However, as these technologies continue to evolve, countries must strike a balance between innovation and responsibility, ensuring that robotics contributes to national security while upholding ethical standards and international norms.

4. Discussion

The historical evolution of robotics engineering is not only a chronicle of technological progress, but also a reflection of mankind’s unremitting pursuit of innovation. Understanding its importance is critical to comprehending the profound impact robotics will have on our world. This chapter explores the

DOI: 10.54254/2755-2721/32/20230861

historical evolution of robotics engineering, from its origins to key milestones in development, highlighting the interdisciplinary nature of the field.

4.1. Origins of robotics engineering

Robotics engineering has its roots in our natural curiosity and desire to replicate and automate tasks. Ancient civilizations first glimpsed this tendency through the invention of mechanized devices. These early efforts revealed the fundamental human desire to make life easier and more efficient. From these humble beginnings, robotic engineering emerged as a testament to our ability to blend art, science, and engineering. It's a journey from the mechanical automata of ancient Greece to the artificial intelligence-powered robots of today. This progress underscores our ability to adapt, learn, and improve over time [9].

4.2. Key milestones in robotics development

Important milestones in the development of robots highlight not only technological achievements but also the social changes they catalyze. In the 1960s, the launch of the first industrial robot, the Unimate, heralded an automation revolution in manufacturing. It sparked discussions about the impact of automation on employment and the labor market, sparking debate that continues today. The development of robotics into intelligent machines capable of learning and decision-making is a testament to our ability to merge computational power with mechanical capabilities. It marks a paradigm shift in how we interact with machines, from issuing explicit commands to collaboration and intuitive partnerships [10].

The interdisciplinary nature of robotics underscores its ability to transcend traditional boundaries. Robotics engineering borrows from mechanical engineering, computer science, artificial intelligence, materials science, and even biology. This fusion of disciplines allows us to draw inspiration from nature, replicate its intricate designs, and create robots that adapt and evolve, much like living organisms [11]

The application of robotics in military and defense is not only a technological advancement but also a transformation in the nature of the armed forces. Discussions surrounding these applications involve strategic, ethical, and international considerations.

4.3. Unmanned Aerial Vehicles

The use of UAVs in military operations has redefined air combat. While these machines offer unprecedented advantages in surveillance and precise targeting, they also raise critical ethical questions about remote warfare and the potential for civilian casualties. Striking the right balance between military effectiveness and moral responsibility remains a challenge [12]

4.4. Ground and underwater robotics

Ground and underwater robots expand the military's combat capabilities. Concerns about autonomous decision making and the potential for misuse also arise with their use. Ethical considerations come into play when discussing the use of robots in conflict zones and their impact on international law [13].

4.5. National security implications

The impact of robotics on national security in the military and defense fields is multifaceted. These technologies offer enhanced security and strategic advantages, but they also introduce vulnerabilities, such as the risk of cyberattacks targeting robotic systems. Therefore, protecting these technologies from misuse and ensuring responsible deployment is an important aspect of national security strategy [14].

5. Conclusion

In conclusion, robotics engineering has undeniably surfaced as a pivotal and transformative force in our contemporary world. This paper has taken an insightful journey through its historical evolution, illuminating the path that has led us to the remarkable present-day applications and the boundless potential it holds for our future. The historical timeline of robotics, from its rudimentary beginnings to

DOI: 10.54254/2755-2721/32/20230861

the sophisticated systems of today, showcases the relentless human quest for technological advancement. Moreover, the wide spectrum of applications, ranging from healthcare and manufacturing to space exploration and artificial intelligence, underscores the incredible versatility of robotics. However, alongside these exciting possibilities come significant ethical and societal challenges. The increasing integration of robotics in our daily lives demands careful consideration of issues such as privacy, job displacement, and the potential for misuse. To navigate these challenges successfully, it is imperative that we promote interdisciplinary collaboration and responsible innovation. By fostering an environment that encourages open dialogue among engineers, ethicists, policymakers, and the wider society, we can ensure that robotics continues to benefit humanity. In this rapidly evolving field, our collective responsibility is to harness the full potential of robotics engineering for the betterment of our lives and the shaping of a brighter, more inclusive, and ethically sound future.

References

[1] Bauer, W., Hämmerle, M., Schlund, S., & Vocke, C. (2015). Transforming to a hyper-connected society and economy–towards an "Industry 4.0". Procedia Manufacturing, 3, 417-424.

[2] Cantú-Ortiz, F. J., Galeano Sánchez, N., Garrido, L., Terashima-Marin, H., & Brena, R. F. (2020). An artificial intelligence educational strategy for the digital transformation. International Journal on Interactive Design and Manufacturing (IJIDeM), 14, 1195-1209.

[3] Ebert, C., & Duarte, C. H. C. (2016, September). Requirements engineering for the digital transformation: Industry panel. In 2016 IEEE 24th International Requirements Engineering Conference (RE) (pp. 4-5). IEEE.

[4] Feroz, A. K., Zo, H., & Chiravuri, A. (2021). Digital transformation and environmental sustainability: A review and research agenda. Sustainability, 13(3), 1530.

[5] Gravish, N., & Lauder, G. V. (2018). Robotics-inspired biology. Journal of Experimental Biology, 221(7), jeb138438.

[6] Johnson, J. (2019). Artificial intelligence & future warfare: implications for international security. Defense & Security Analysis, 35(2), 147-169.

[7] Johri, P., Singh, J. N., Sharma, A., & Rastogi, D. (2021, December). Sustainability of coexistence of humans and machines: an evolution of industry 5.0 from industry 4.0. In 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART) (pp. 410-414). IEEE.

[8] Ventre, D. (2020). Artificial Intelligence, Cybersecurity and Cyber Defence. John Wiley & Sons.

[9] Carroll, M., & Carroll, M. (2011). Early Concepts, and What It Really Takes to Explore Alien Skies. Drifting on Alien Winds: Exploring the Skies and Weather of Other Worlds, 22-41.

[10] Jacobs, I., Jaffe, J., & Le Hégaret, P. (2012). How the open web platform is transforming industry. IEEE internet computing, 16(6), 82-86.

[11] Narayanan, K. L., Krishnan, R. S., Son, L. H., et al. (2022). Fuzzy guided autonomous nursing robot through wireless beacon network. Multimedia tools and applications, 1-29.

[12] Azcárate, A. L. V., & Sussman, H. (2017). Technopoïesis: Transmedia Mythologisation and the Unity of Knowledge. An Introduction.

[13] Lin, P., Bekey, G., & Abney, K. (2008). Autonomous military robotics: Risk, ethics, and design.

[14] Schmidt, E., Work, B., Catz, S., Chien, S., Darby, C., Ford, K., ... & Matheny, J. (2021). National security commission on artificial intelligence (ai). National Security Commission on Artificial Intellegence, Tech. Rep.

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230863

Data analysis with different variables and credit risk assessment

Ruixin Jin[1,3],Huanyu Zhou[2]

[1]Hangzhou Tianyuan College, Hangzhou, 311121, China
[2]Guangzhou NO.7 Middle School, Guangzhou, 510080, China

[3]alan2873464366@outlook.com

Abstract. Nowadays, credit payment is a very common way to pay, such as credit cards, loans, many people can use their credit as a guarantee to borrow money from the bank, however some people will default. So we have to predict whether the borrower will pay on time, it is known as credit risk assessment. In this paper, we analyze a data set on credit risk to predict whether individuals will be late on their payments, helping financial firms improve their earnings and reduce their losses. We not only made predictions on the data, but also analyzed the relationship between the variables that affect the overdue probability to find some specific associations. Specifically, we performed ANOVA analysis and found that married people borrowed significantly more than other groups, and the delinquency rate of people with higher education was lower, and the delinquency rate of married people was higher than that of unmarried people. In addition, we conducted a binary logistic regression and found that gender had no significant impact on the prediction results, but an individual's amount of bill statement, amount of previous payment, past repayment situation and Amount of the given credit had an impact on the prediction results. Other variables, such as marital status and education, can also impact the predicted results. Our research puts forward more factors affecting credit risk and also different angles that can be used to analyzes individual credit risk. This has a guiding role for financial firms like banks and other companies in the financial industry, providing more ways to help them analyze the credit risk of borrowers.

Keywords: Credit risk assessment, Factor analysis, ANOVA analysis, Binary logistic regression.

1. Introduction

Nowadays, more and more people use credit cards, which are a payment method guaranteed by personal credit. However, banks and financial companies do not know whether individuals will pay their bills on time. Inevitably, banks must take risks in giving loans and credit cards to customers because these are economic drivers [1]. But when debts are not paid on time, the cash flow of financial companies will be affected, and it can also cause some harmful effects on the company. So credit risk assessment is vital for banks; they must ensure that borrowers are able to pay their installments before allocating a loan to them [2]. For example every month almost every adult in the US and the UK is scored several times to enable a lender to decide whether to mail information about new loan products, to evaluate whether a credit card company should increase one's credit limit, and so on [3]. On a daily basis credit/financial analysts have to investigate an enormous volume of financial and non-financial data of firms [4]. Credit

DOI: 10.54254/2755-2721/32/20230863

risk assessment can help financial companies effectively manage risks, make wise investment decisions, and avoid unnecessary losses.

So the predecessors have done many research in this area. Paweł Pławiak et al .applied a novel approach based on deep genetic cascade ensemble of different support vector machine (SVM) classifiers (called Deep Genetic Cascade Ensembles of Classifiers (DGCEC)) to the Statlog Australian data [5]. Charles Guan et al .combined ML classification models trained on limited data with a well established form of "human-in-the-loop" knowledge acquisition based on Ripple-Down Rules (RDR) to construct fair and compliant rules that could also improve overall performance [6]. David West investigates the credit scoring accuracy of five neural network models: multilayer perceptron, mixture-of-experts, radial basis function, learning vector quantization, and fuzzy adaptive resonance. Found that the multilayer perceptron may not be the most accurate neural network model [7]. Stjepan Oreski et al .designed a hybrid system with genetic algorithm and artificial neural networks (GA-NN) for finding an optimum feature subset at retail credit risk assessment that enhances the classification accuracy of neural network classifier [8]. Hussain Ali Bekhet et al .Radial basis function (RBF) and logistic regression model are used to test the validity of a credit scoring model. In terms of overall accuracy, logistic regression model is slightly better than radial basis function model. However, the radial Foundation function is better at identifying those customers who are likely to default [9].

We found a data set on credit risk assessment to study the relationship between default of payment and different factors. The variables in the dataset are Gender, Age (year), Marital status, Education, Amount of the given credit, History of past payment, Amount of bill statement, Amount of previous payment and default of payment. Gender is divided into male and female, marital status is divided into married, unmarried and other, education is divided into high school, university, graduate and other. Among all variables, these three variables are different (History of past payment, Amount of bill statement, Amount of previous payment), each of them is divided into six periods, representing different states of individuals in the data set in six months. The History of past payment represents if the individual has been overdue in the past or not and, if so, for how long.

In this paper, we used factor analysis, ANOVA and binary logistic regression to analyze the data.By using factor analysis, we can reduce the dimension and compress the six periods of data of History of past payment, Amount of bill statement and Amount of previous payment to a smaller number of data and simplify the data. In order to make the data easier to observation and research, and also to let the model has a better generalization ability. To investigate the relationship between the two different data points, we performed ANOVA analyses and used the resulting data to analyze which factors might be more influential on default of payment. In our research the method was also used to investigate the relationship between other factors. We also analyze the data and use binary logistic regression to predict if the person will default his payment.

2. Factor analysis

2.1. Factor analysis on History of payment

Factor analysis is formed on the History of past payment, and compressed the data through the analysis of component eigenvalues and lithotriptic maps. After analyzing the data, this work found that the set of data can be compressed to at least two components. As can be seen from the rotated component matrix in as shown in Table 1 and 2, component 1 can better summarize phases 3, 4, 5 and 6, while component 2 can better summarize phases 1, 2 and 3. Component 2 is better at summarizing recent conditions. (the period of history of past payment smaller it refers more recent). It can be seen from the common factor variance table that the extracted two components have better generalization ability to the original data, so such compression is reasonable.

DOI: 10.54254/2755-2721/32/20230863

Table 1. Rotated component matrix.

	Component 1	Component 2
History of past payment 1	0.210	0.889
History of past payment 2	0.438	0.797
History of past payment 3	0.624	0.625
History of past payment 4	0.809	0.428
History of past payment 5	0.890	0.311
History of past payment 6	0.877	0.248

Table 2. Communality table.

	Initial	Extraction
History of past payment 1	1.000	0.849
History of past payment 2	1.000	0.821
History of past payment 3	1.000	0.777
History of past payment 4	1.000	0.839
History of past payment 5	1.000	0.884
History of past payment 6	1.000	0.824

2.2. Factor analysis on Amount of bill statement

This work also carried out factor analysis on the Amount of bill statement, and compressed this set of data through the analysis of component eigenvalues and gravel map. After analyzing the data, this work found that the set of data can be compressed to at least two components. As can be seen from the rotated component matrix in as shown in Table 3 and 4, component 1 can better summarize the 4th, 5th and 6th phases, and component 2 can better summarize the 1st, 2nd and 3rd phases. Component 2 is better at summarizing recent conditions. (the time order of 6 phase of Amount of bill statement is the same as that of History of past payment). It can be seen from the common factor variance table that the extracted two components also have good generalization ability to the original data, so such compression is reasonable.

Table 3. Rotated component matrix.

	Component 1	Component 2
Amount of bill statement 1	0.452	0.870
Amount of bill statement 2	0.510	0.844
Amount of bill statement 3	0.611	0.747
Amount of bill statement 4	0.755	0.612
Amount of bill statement 5	0.845	0.507
Amount of bill statement 6	0.868	0.453

DOI: 10.54254/2755-2721/32/20230863

Table 4. Communality table.

	Initial	Extraction
Amount of bill statement 1	1.000	0.961
Amount of bill statement 2	1.000	0.972
Amount of bill statement 3	1.000	0.932
Amount of bill statement 4	1.000	0.945
Amount of bill statement 5	1.000	0.971
Amount of bill statement 6	1.000	0.958

2.3. Factor analysis on Amount of previous payment

When analyzing the Amount of previous payment, the results show that when the data was compressed into two components, the interpretation ability was not good. When combined with the accumulation, when the data was compressed into two components, the accumulation was less than 50%.This data needs to be accumulated to the fourth factor before the accumulation approaches 80%, so this work decided to compress this set of data to four components to make the compressed components more interpretable.

Table 5. Communality table.

Component 1	Total	Cummunality%
Amount of previous payment 1	1.958	32.629
Amount of previous payment 2	0.893	47.519
Amount of previous payment 3	0.852	61.719
Amount of previous payment 4	0.837	75.669
Amount of previous payment 5	0.753	88.225
Amount of previous payment 6	0.707	100.000

This is to the data after compression, from Table 7 communality table, we can conclude four ingredients, compressed on the original data has good generalization ability. As shown in table 5 and 6, the rotated component matrix shows that component 1 can better summarize phases 1, 2 and 3, component 2 can better summarize phase 4, component 3 can better summarize phase 5 and component 4 can better summarize phase 6. (Note: the chronological order of factors from 1 to 6 is the opposite of the first two groups.).

DOI: 10.54254/2755-2721/32/20230863

Table 6. Rotated component matrix.

	Component 1	Component 2	Component 3	Component 4
Amount of previous payment 1	0.731	0.086	-0.034	0.174
Amount of previous payment 2	0.763	-0.29	0.187	0.005
Amount of previous payment 3	0.569	0.376	0.049	0.023
Amount of previous payment 4	0.104	0.946	0.075	0.077
Amount of previous payment 5	0.112	0.082	0.977	0.076
Amount of previous payment 6	0.123	0.075	0.077	0.979

Table 7. Communality table.

	Initial	Extraction
Amount of previous payment 1	1.000	0.573
Amount of previous payment 2	1.000	0.618
Amount of previous payment 3	1.000	0.468
Amount of previous payment 4	1.000	0.917
Amount of previous payment 5	1.000	0.979
Amount of previous payment 6	1.000	0.986

3. ANOVA analysis

3.1. Gender and overdue

In this data set, in the item of whether the loan is overdue, 1 means overdue, and 0 means repayment on time, so in this line chart and other chart where the Y-axis is overdue, the higher the data point, the higher the probability of the group of people being overdue. As can be seen from the Figure 1, the probability of men defaulting on loans is slightly higher than that of women, but the difference is not large, so it can be said that gender has no significant impact on whether they will default.

DOI: 10.54254/2755-2721/32/20230863

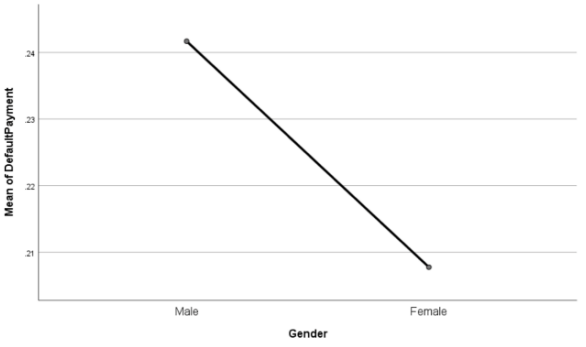


Figure 1. The relationship between Gender and overdue.

3.2. Marital status and overdue

It can be seen from the line chart in Figure 2 that married people are more likely to overdue than unmarried people. We confused why people that are married the possibility of default can be higher, there are two people repay the debts together, so we combined the line chart in Figure 3 (the relationship between marital status and loan amount) for analysis.

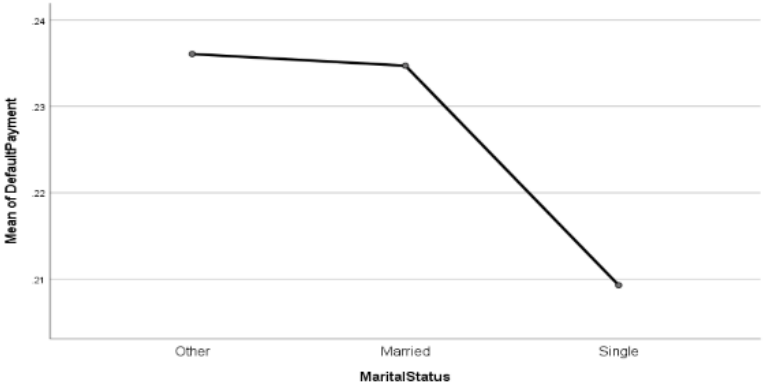


Figure 2. The relationship between marital status and overdue.

From Figure 3, this work shows that married people have the largest amount of loans, and the loan amount of unmarried people is obviously smaller than that of married people. Therefore, married people may be have a bigger possibility to overdue, and it can also be regarded as the group with more loans, the greater the probability of loan overdue. So we can infer that the majority of married people probably spend more than the majority of people in other states, and they have to take on more expenses, maybe a house, a car, or other spending on their children.

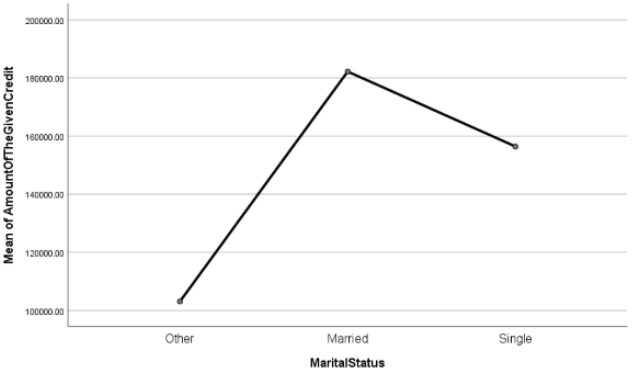


Figure 3. The relationship between marital status and loan amount.

DOI: 10.54254/2755-2721/32/20230863

In addition, this work can relate Figure 4 to the relationship between age and marital status. As Figure 4 shows, marriage rates are higher among those over 33. Combined with the above analysis, married people borrow the most, so it can be guessed that people in this age group consume more and have more expenses.

Figure 4. The relationship between marital status and age.

3.3. Education and overdue

From Figure 5, we can get an intuitive conclusion that people with higher education are less likely to default. But the difference between the default probability of a college degree and a high school degree is not big. The default probability of graduate school students is significantly lower than the previous two, which can be guessed that this group may have higher salary income after graduation, or the number of loans of this group is smaller.

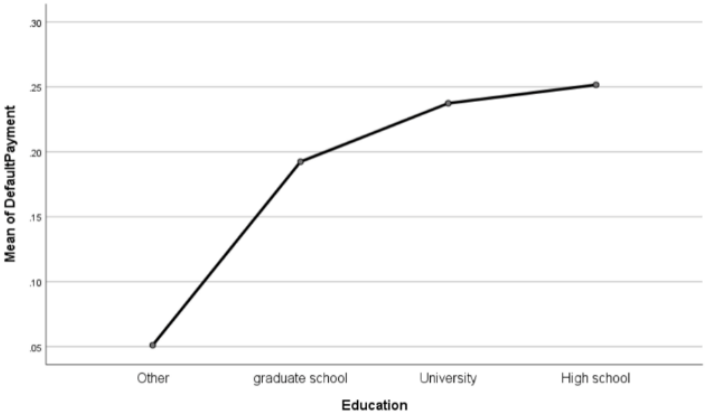


Figure 5. The relationship between education and default.

4. Binary logistic regression

From the table 8, we can see that the B coefficient of gender is -0.105, so gender is an influential factor to predict whether an individual is overdue, and its Exp(B) is 0.9, indicating that the aberration of overdue probability between men and women is not very large.

If the amount of the given credit has a B coefficient of 0, it can be considered not an influencing factor.

The B-coefficient of education is -0.054, so it is also one of the factors influencing the prediction result. Exp(B) shows that the aberration of overdue probability of different education levels is about 0.95 times.

DOI: 10.54254/2755-2721/32/20230863

Marital status is also one of the influential factors of overdue prediction, with a coefficient of -0.152, which has a greater impact on the overall prediction result, and its Exp(B) ranges from 0.806 to 0.915, indicating that the overdue probability of different maretal status differs by about 0.86 times.

The 3456 history of payment had no effect on the forecast, because the sig. is bigger than 0.05.

(In this experiment, the reason why we did not use the compressed amount of previous payment is that only when it is compressed to 4 or 5, can the data have a better generalization, so we used the original data of 6 periods for the experiment.)

Table 8. Binary logistic regression.

				95% C.l.for EXP(B)	
	B	Sig.	Exp(B)	Lower	Upper
Gender	-.105	.001	.900	.848	.956
Amount Of The Given Credit	.000	.000	1.000	1.000	1.000
Education	-.054	.020	.948	.906	.991
Marital Status	-.152	.000	.859	.806	.915
Age	.006	.001	1.006	1.003	1.010
Amount Of Previous Payment1	.000	.000	1.000	1.000	1.000
Amount Of Previous Payment2	.000	.000	1.000	1.000	1.000
Amount Of Previous Payment3	.000	.005	1.000	1.000	1.000
Amount Of Previous Payment4	.000	.000	1.000	1.000	1.000
Amount Of Previous Payment5	.000	.001	1.000	1.000	1.000
Amount Of Previous Payment6	.000	.060	1.000	1.000	1.000
History of payment 3456	.032	.123	1.032	.991	1.075
History of payment 123	-.146	.000	.864	.831	.898
Constant	-.854	.000	.426		

Form the Table 8 and 9, we can see the model has a relatively good predict ability, the overall accuracy is about 70% (In the experiment we set the dividing value as 0.220).

DOI: 10.54254/2755-2721/32/20230863

Table 9. Classification Table.

Observed			Predicted		
			Default Payment		Percentage Correct
			No	Yes	
Step 1	Default Payment	No	16112	6947	69.9
		Yes	2457	4153	62.8
	Overall Percentage				68.3

5. Conclusion

This paper mainly uses SPSS to analyze the influence of different variables on the overdue credit repayment of commercial banks, including factor analysis, binary logistic regression, ANOVA, comparative analysis and others. At the same time, for these different analysis methods, we also made a summary of the results and chart analysis. Through the analysis, we have reached some conclusions, including the higher the degree of the group of people delinquent rate is lower, the probability of married people delinquent rate is higher than unmarried people, the amount of married people loan is higher, the difference between men and women is not significant and so on. At present, with the development of the global economy, the cooperation between enterprises and banks is deepened, and the bank pursues profit as the goal, so it is inevitable that customers will delay their payment in the process. The phenomena told by these data can enable us to predict the lender's loan risk well in advance, thus avoiding many problems and accidents in advance and increasing profits. Therefore, commercial banks should always pay attention to the changes in the global economic situation, collect the information of each customer, understand the development prospects of customer business, establish and improve the credit risk assessment system, so as to reduce risks and promote the development of banks.

References

[1] Moradi, S., Mokhatab Rafiei, F., Financ Innov 5, 15 (2019), A dynamic credit risk assessment model with data mining techniques: evidence from Iranian banks, https://doi.org/10.1186/s40854-019-0121-9

[2] J.I. Gusti Ngurah Narindra Mandala, Catharina Badra Nawangpalupi, Fransiscus Rian Praktikto (2012), Assessing Credit Risk: An Application of Data Mining in a Rural Bank, https://doi.org/10.1016/s2212-5671(12)00355-3

[3] Jonathan N. Crook a, David B. Edelman b, Lyn C. (2007), Thomas c Recent developments in consumer credit risk assessment, https://doi.org/10.1016/j.ejor.2006.09.100

[4] M Doumpos a, K Kosmidou a, G Baourakis b, C Zopounidis (2002), a Credit risk assessment using a multicriteria hierarchical discrimination approach: A comparative analysis, https://doi.org/10.1016/S0377-2217(01)00254-5

[5] Paweł Pławiak a, Moloud Abdar b, U. Rajendra Acharya c (2019), Application of new deep genetic cascade ensemble of SVM classifiers to predict the Australian credit scoring, https://doi.org/10.1016/j.asoc.2019.105740

[6] Charles Guan, Hendra Suryanto, Ashesh Mahidadia, (2023), Michael Bain & Paul Compton Responsible Credit Risk Assessment with Machine Learning and Knowledge Acquisition, https://link.springer.com/article/10.1007/s44230-023-00035-1#Sec5

[7] David West (2000), Neural network credit scoring models, https://doi.org/10.1016/S0305-0548(99)00149-5

DOI: 10.54254/2755-2721/32/20230863

[8] Stjepan Oreski a, Dijana Oreski b, Goran Oreski a (2012), Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment, https://doi.org/10.1016/j.eswa.2012.05.023

[9] Hussain Ali Bekhet, Shorouq Fathi Kamel Eletter (2014), Credit risk assessment model for Jordanian commercial banks: Neural scoring approach, https://doi.org/10.1016/j.rdf.2014.03.002

Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230864

Simulation of the motion of a pendulum

Tianhui Zhang

New channel, Shenyang, 110002, China

Tt060913@qq.com

Abstract. The project's objective is to simulate movement in the study and analysis of motion simulation problems and to propose simulation algorithms based on numerical calculation methods. It has many practical application values. In engineering and science, it is often necessary to simulate and analyze the motion of the fold to study the motor and dynamic properties of the fold, providing the theoretical basis and solutions to practical problems. With the support of modern computers and numerical computing technology, the problem of motion simulation has become a popular research direction. Many scholars and engineers have proposed different numerical calculation methods and simulated algorithms to simulate the motion process of the fold and analyze its motion laws. This article introduces the basic knowledge of physics and the formulas of motion, as well as some important concepts and theories related to motion. The motion simulation algorithm was then analyzed and discussed in detail. Subsequently, numerical calculations were prepared using MATLAB software, and simulated experiments were conducted using examples to analyze dynamic changes. Finally, the prospects for the future direction of research are presented. Therefore, if the initial speed is the same, the width and length of the time will increase.

Keywords: numerical calculation methods, MATLAB, change initial speed.

1. Introduction

A pendulum is a system that produces recurrent oscillations, with one end of an unextendable string or thin rod hanging at a certain point in the gravitational field and the other end of a rigid, heavy ball fixed to form a single spin. For ease of handling, we assume that the line is a long string that can be neglected and not stretched; the radius of the ball is much smaller than the length of the line, thus ignoring the size of a ball, and it is considered a quality point.

The monocouple consists of a smaller mass object (a monocouple ball) and a lightweight, fine line, usually hanging on the support. A single bow ball is influenced by gravity and moves along an arc with cyclic and gradually weakening amplitude. The laws and cycles of the motion of the single swing are related to the weight, length, and initial angle of the ball, so the movement of the single swing is an important experiment in the study of mechanics, dynamics, and vibration.

The accuracy of the simulation results depends on the calculation methods and parameter settings used and usually requires error analysis and accurate control to ensure the reliability of the results. Multiple methods and tools can be used to simulate single swing movements, such as Eurafa, Lagrange, MATLAB, etc. The accuracy of the simulation results depends on the calculation methods and parameter settings used and usually requires error analysis and accurate control to ensure the reliability of the

DOI: 10.54254/2755-2721/32/20230864

results. Simulation of single swing movements can be used in academic research and engineering design; for example, it can be applied to mechanical control systems, astronomy, computer animation, etc.

The primary purpose of this dissertation, titled "Simulation of the Motion of a Pendulum," is to advance our understanding of pendulum dynamics through advanced computational simulations and to explore the practical applications of this knowledge across various disciplines. This research endeavors to achieve several specific objectives: to delve into the theoretical foundations of pendulum motion, encompassing classical mechanics, nonlinear dynamics, and the mathematics of oscillatory systems. By establishing a robust theoretical framework, this work aims to provide a comprehensive understanding of the underlying principles governing pendulum behavior.

To develop and implement advanced computational models and simulation algorithms capable of accurately replicating the behavior of pendulum systems under a wide range of conditions. The purpose is to harness the power of modern computing to simulate pendulum dynamics with precision and fidelity.

2. Solving the linear ordinary differential equation:

Through reading the comprehensive paper on a class of high-level Euler equations, I gained a deeper understanding of Euler's equation [1], The first and second stages of the ordinary differential equation as a carrier, the application of variable replacement in the query solution was analyzed and summarized, thereby expanding the variable substitution method in the solution of ordinary variable equation [2]. Then, I read a paper on the problem of solving a linear equation of a class of descendable secondary variable coefficients and discussed a solution of a special secondary differential coefficient of a series of linear equations. The solutionability of the linear differential equation u(x)y″+v(X)y′+w(x),y=0 is studied using the degradation method of the variable coefficient and so solved this linear ordinary differential equation[3].

$$\theta'' = -g*\sin(\theta) \ldots (1)$$

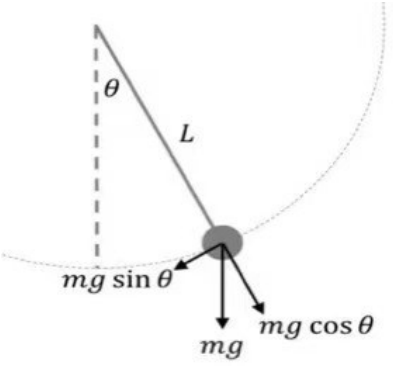


Figure 1. The motion of the pendulum.

Figure 1 shows the motion of the pendulum where θ is the angle from the negative vertical axis, and g is the gravitational constant. Initial conditions are θ(0)=45° and θ'(0)=0, and describe the initial position and the initial velocity of the pendulum.

Then use the Euler method to solve this equation, and we get the final equation, for the initial conditions are θ(0)=45° and θ'(0)=0.

Type the final equation on MATLAB, then plot the graph:

Algorithm 1.

```
1.for i=2:N
    theta(i+1)= 2*theta(i)-theta(i-1)-g*deltat*deltat*sin(theta(i));
end
2.plot(theta)
```

DOI: 10.54254/2755-2721/32/20230864

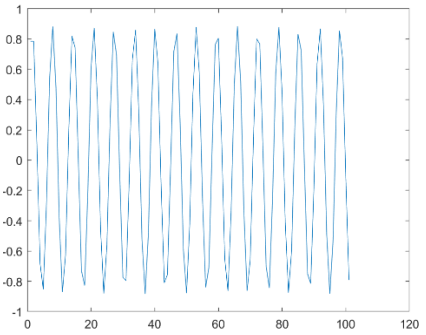


Figure 2. The motion of the pendulum at v=0.

As shown in Figure 2, the motion of the pendulum at v=0. A dynamic description of the single swing movement was made using MATLAB software to calculate the values. By comparing the theoretical calculations with the real data, it was found that they matched the real value. For this purpose, a basis was provided for determining the optimal swing angle for the actual one-swing experiment [4]. The method of study of the single swing movement cycle, written by the land column, proposes the direct method of calculating the motion cycle and gives the universal expression of the mono swing cycle so that the traditional method can be improved [5].

Changing the initial velocity to see the difference. Through specific algorithms, compare the calculation accuracy and efficiency of various high-level algorithms, give the corresponding numerical error and chart description, fully verify the advantages and shortcomings of class 2 high-grade methods, and provide a reference basis for the solution of practical application problems [6]. First, the linear elementary equation of the second variable factor is transformed two times in a row to the equivalent of a second type of vitamin-linear elemental equation. Then, the solution of the elementary Equation is solved, thus obtaining the requested problem exists a single continuous solution, and the solution is given [7].

Algorithm 2. at v=1(v is velocity):

```
1.  for i=2:N
        theta (i+1)=2*theta(i)-theta(i-1)-g*deltat*sin(theta(i));
    end
2.  plot(theta)
```

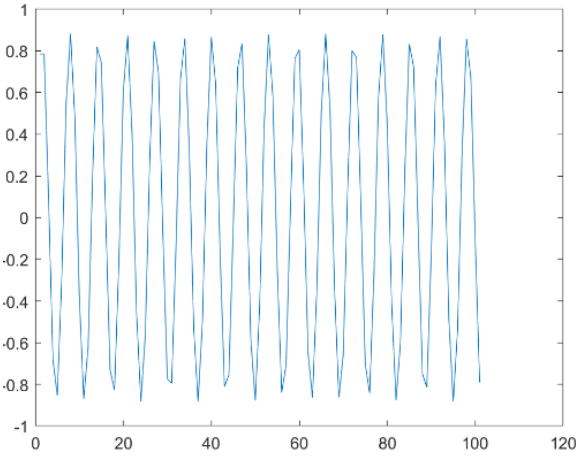


Figure 3. The motion of the pendulum at v=1.

DOI: 10.54254/2755-2721/32/20230864

Figure 3 shows how the pendulum moves at the initial velocity v=1. Then, continue the process by changing the initial velocity to V=2,3,4 and plot the graphs. Use methods and methods in combination with MATLAB software to analyze the monoclinic system, high accuracy, and small error [8]. The accurate period calculation of a single pendulum under the condition that the air resistance is neglected was given by numerical integration with Mathematica. Accordingly, some approximate period formulae described in references were compared by the curve fitting method of MATLAB [9].

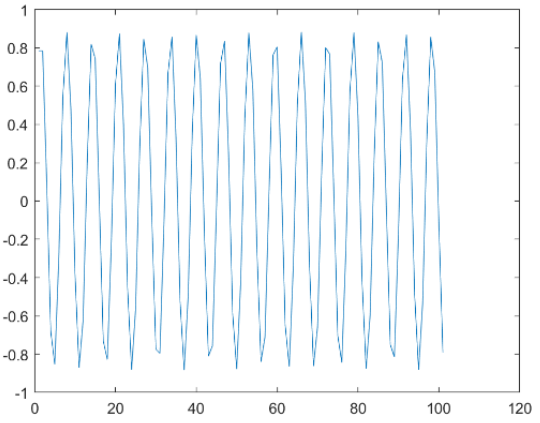


Figure 4. The motion of pendulum at v=0.

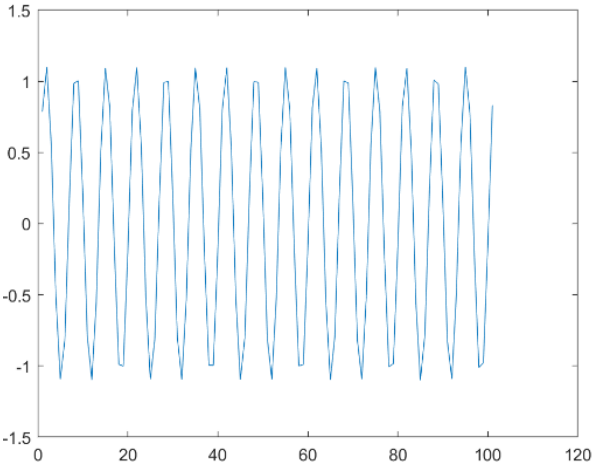


Figure 5. The motion of pendulum at v=1.

DOI: 10.54254/2755-2721/32/20230864

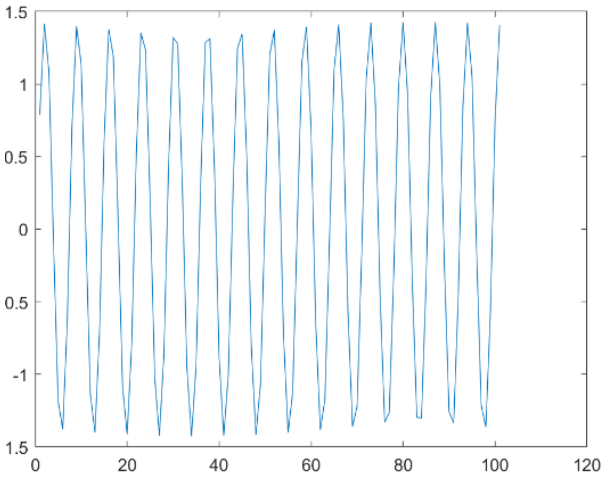


Figure 6. The motion of pendulum at v=2.

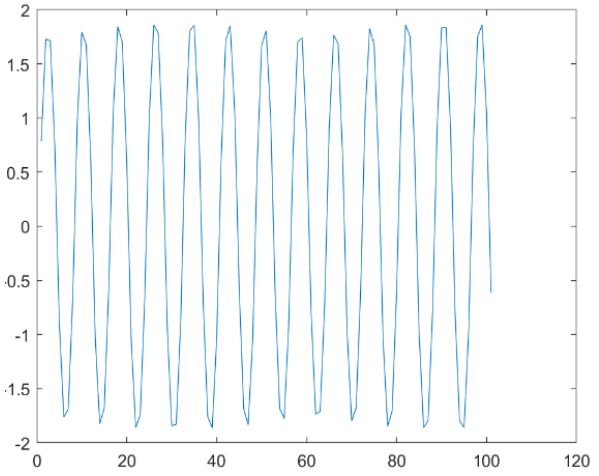


Figure 7. The motion of pendulum at v=3.

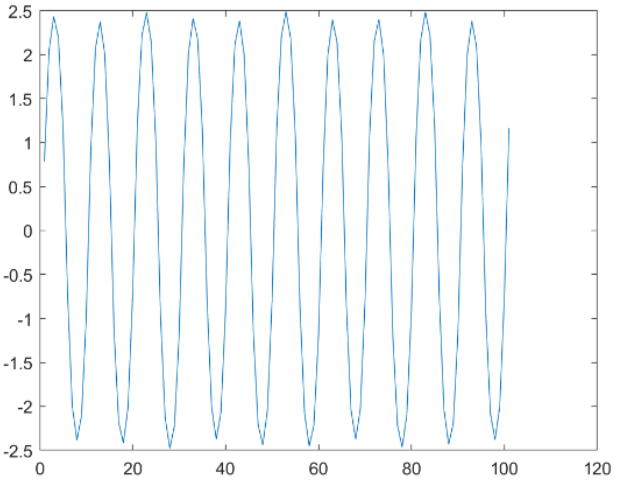


Figure 8. The motion of the pendulum at v=4.

Figures 4, 5, 6, 7, and 8 show the pendulum's motion and compare the initial velocity by graphs at v=0,1,2,3,4, respectively. At the same time, using MATLAB, accurate results in different periods are

DOI: 10.54254/2755-2721/32/20230864

calculated, and follow-up curves are drawn based on these data. By comparing formulas and images, the cycles, frequencies, and magnitude of single-swing movements can be seen from these different images. The results showed that for unimpeded cycles such as oscillation, the larger the initial angle of the single swing, the greater the cycle of the mono swing. The nonlinear twisting cycle obtained in the numerical calculation is consistent with the result of the literature's aggregate resolution, with an error of less than 0.3% [10].

3. Conclusion

The same tiny ball was allowed to fall with varied initial speeds while maintaining the initial angle constant, and the study's conclusion—that as the initial velocity grows, the amplitude and period both grow—was achieved as a result. This discovery can be applied in many fields. For example, it can be used in a deep understanding of simple pendulum movement, which is a classic problem in Classical physics. Changing the initial speed of a simple pendulum can help to deeply understand the movement law and vibration characteristics of a simple pendulum, which is conducive to academic research and teaching.

The vision of flat motion simulation has great hope for further progress and interdisciplinary exploration. When we look to the future, several key areas draw our attention. Future advances in computational methods and technologies will lead to more sophisticated and efficient simulation of rotational motion. We expect to develop new algorithms and digital technologies that enable simulations to capture the complexity of behavior with unparalleled accuracy. Engineers can use these simulations to optimize structural designs, astronomers can simulate movements of celestial bodies, and biologists can study vibrational behavior in biological systems. There is great potential for interdisciplinary development. With increased computational and graphical capabilities, it has become more feasible to incorporate real-time interactive simulations into educational materials and outreach initiatives. It can potentially revolutionize science education by attracting students and promoting a deeper understanding of physics principles.

In addition to the academic community, there is also the potential to solve real-world challenges in practical situations. Engineers can use simulations to design energy-based collection systems or improve the stability of rotating structures. These practical applications can lead to far-reaching innovations in industries such as renewable energy and civil engineering. And the simulations developed here can serve as a valuable educational tool. They can be incorporated into science courses to enhance students' understanding of complex physical phenomena, and the realization of simulated actions is applied to computer graphics, animation, video games, and special effects. It helps to create a vibrant virtual environment that enriches the visual quality and reality of entertainment experiences.

At the same time, he can conduct research and experiments, and the simulator can be used as a virtual laboratory to conduct unrealistic or expensive experiments in the physical world. Researchers can explore the behavior of complex systems under various conditions and parameter settings.

In short, the study of this paper goes far beyond the boundaries of theoretical physics. By unlocking the complexity of penis dynamics, it opens the door to innovation and solutions in engineering, energy, astronomy, biomechanics, education, recreation, and scientific research. Understanding the practical impact of this movement is expected to shape cross-disciplinary progress and contribute to building a more informed and technologically advanced society. This work demonstrates the transformative potential of simulation movement in visual and application in our evolving world.

References

[1] Deng Ruijuan, Chen Qianqian.(2021) Introduction of the Higher Grade Eula Equation [J]. Red River Academy Journal, 19 (02): 149-150

[2] Li Jianxiang, Li Ling.(2016) The application of variable substitution in the solution of differential equations [J]. Scientific advice (educational and scientific research), 2022(08):123-127.

DOI: 10.54254/2755-2721/32/20230864

[3] Huang Fei, Geng Jie.(2018) A Class of Reducible second-order homogeneous linear equations with variable coefficients [J]. Journal of Hebei North University (Natural Science Edition),34(09): 7-921.
[4] Jie Liu.(2000) Different solutions to the vibration equation of a single pendulum [J] . Journal of Liaoning Teachers College (natural science edition) , 2000(02) : 26-27.
[5] Ju Yanqing, Zhang Fenglei.(2010) Movement analysis of simple pendulum under the action of comprehensive factors [J]. Liaodong Academy Journal (Natural Science Edition), 17 (02): 151-153.
[6] Jiangshan, Zhang Yan, Sun Meiling.(2019) Comparison of the primary values of the equation of common differential differentiation with the application of the Long-Kuta method [J]. Journal of Science of Teachers′ College and University, 39(12):12-15.
[7] Chen Yan.(2019) Explicit solution of second order Linear differential equation initial value problem with variable coefficients [J]. Journal of Jiamusi Vocational College, 2019(05) : 287-288.
[8] Yang Wenjin, Wang Hongli, Liu Caiyun, etc. . (2020) using MATLAB to determine the motion characteristics of single pendulum theoretical research [J]. Journal of Southwest Normal University (natural science edition) , 45(11) : 167-170.
[9] Ju Yanqing.(2006) Comparison of several approximate formulas for the period of a simple pendulum[J]. Laboratory Research and Exploration, 2006(05):585-587.
[10] Ma Kun. (2019) numerical analysis of simple pendulum motion based on MATLAB [J] . Journal of Chizhou College, 33(03) : 37-39.