Application of Deep Learning in Sentiment Analysis Based on Movie Review Datasets

Ziyi Yang

Information Management and Information Systems, China University of Petroleum (Beijing), Beijing, 102200, China

2021011860@student.cup.edu.cn

Abstract. With the rapid development of the Internet and social networks. user-generated content (UGC) has exploded on the Internet as an important form of user feedback, especially movie reviews, which are of great value to the film industry and market analysis, and can help relevant groups understand the reputation and evaluation of movies. As a natural language processing technology, sentiment analysis can effectively identify and extract emotional biases in text, providing a new perspective and method for sentiment analysis from film criticism. The results show that using deep learning can significantly speed up this work and make sentiment analysis more efficient and faster. The purpose of this article is to explore the application and specific process of deep learning in sentiment analysis of film criticism datasets, to find the most appropriate model by comparing and analyzing multiple models, and to present the rating data of some films from the perspective of display technology.

Keywords: Deep learning, sentiment analysis, LSTM, movie.

1. Introduction

With the fast development of internet technology, the amount of the internet information has increased dramatically, especially on social media platforms, where user-generated content such as movie reviews and microblogs have become an important source for analyzing the emotional moods of the public. As an important branch in the field of natural language processing, sentiment analysis aims to automatically identify and extract subjective emotional information in text using computer technology to achieve fast and accurate analysis of text data on a large scale. In recent years, deep learning technology has achieved remarkable results in the field of sentiment analysis and has become a research hotspot.

The purpose of this study is to explore the application of deep learning technology in sentiment analysis based on movie review datasets, as well as to provide an effective and practical method for analyzing the sentiment of movie reviews by comparing and analyzing the performance of Doc2vec and deep learning models in sentiment analysis tasks. In addition, this study will also examine the impact of key aspects such as data preprocessing and model parameter optimization on sentiment analysis performance to provide a useful reference for research in related fields.

As an important method in the field of natural language management, sentiment analysis aims to distinguish sentiment in texts, including social media analysis, customer service, public affairs, etc. There are many procedures, including opinion monitoring, opinion monitoring, etc. Traditional Proceedings of CONF-MLA 2024 Workshop: Securing the Future: Empowering Cyber Defense with Machine Learning and Deep Learning DOI: 10.54254/2755-2721/94/2024MELB0075

sentiment analysis is primarily based on salient features and creating rules. When dealing with large amounts of data and highly complex text, accuracy is often reduced.

In the course of many years of research and exploration by many researchers, sentiment analysis has gradually achieved a series of remarkable research results. By using two different textual information analysis methods, qualitative and quantitative, the methods of solving problems are different, and the prediction and recognition of the emotional tendency of user reviews has naturally become a research direction that attracts many researchers in the field of natural language processing [1]. Tureny et al. analyzed and calculated the similarity of words for the purpose of introducing point mutual information, and reconstructed the composition of the dictionary by introducing polarization semantics, which increased the richness of the dictionary part-of-speech construction [2]. On the basis of the preestablished sentiment dictionary, Yang et al. introduced the LDA (Language Acquisition Device) model tool to further extract the subject words in the dictionary, expand the richness of the expression of emotional trends, and provide a more adequate supplement to the polysemy of the subject words [3]. Zhou et al. They withered information, Chinese textual information identified on Weibo, Lege's innovations in the etching SO-PMI algorithm, emotional topic words in the text were further screened, and a new emotion dictionary was compiled for new contacts with contacts [4]. Li et al. used the priority weighting method to extract the central word, integrate traditional machine learning methods, and effectively integrate the weighted priority to obtain the semantic direction of the text [5]. Pang et al. first introduced machine learning methods into solving the binary classification task of sentiment analysis, combined with the traditional bag-of-word method to improve the classification effect [6]. A large number of comments on Twitter were used as a dataset, and then sentiment was classified, and the expressions and rich features of the words in the dataset were included in the sentiment trend range to improve the performance and accuracy of the algorithm [7]. Wang et al. conducted in-depth data analysis of movie reviews based on text mining technology, used TF-IDF algorithm to visualize high-frequency words, and used Bayesian classifier to divide the film short review dataset into a positive and negative review dataset, obtained the probability of related subject words, and found out the deep information of the film review text data [8]. Li et al. introduced Bayesian algorithm, which provides an effective solution to the problem of emotion classification [9]. To a certain extent, the accuracy of Chinese sentiment classification in the target domain and the adaptability of the classifier in the target domain were improved. Govindarajan et al. have made great contributions to the organic integration of genetic algorithms and traditional machine learning algorithms, and have made certain improvements in the effect of dichotomization by using the fused algorithm as a new solution to solve the problem of emotional binary classification [10]. Jiang et al. adopted a solution of integrating multiple processing strategies with traditional machine learning, and added richer semantic rules to the sentiment trend of words, so as to improve the prediction performance of text sentiment words [11].

2. Data

2.1. Data source

In recent years, with the rapid development of Internet technology, the data of movie reviews on the Internet has shown an explosive growth. In this study, a large-scale film review dataset is selected from a well-known movie review website, which contains a large number of user reviews of movies. The dataset covers films of different genres and generations, as well as the perceptions of movies by different user groups.

First, the dataset contains fields such as the text content of the review, the rating, the time of the comment, and the user information. Among them, the text content of the review is the main analysis object of sentiment analysis, and the score is sentiment labels, which are used to train and validate the model. Fields such as comment time and user information were not the focus of analysis in this study.

Second, the dataset is large, and these review data have the following characteristics:

Abundant data: It contains a large number of movie reviews, which is helpful for the learning and generalization ability of the model. Data diversity: It covers movies of different genres and generations,

as well as the perceptions of movies by different user groups, which helps to improve the robustness of the model. High data quality: The content of the comments is highly readable, which is conducive to subsequent text processing and analysis.

2.2. Data cleaning and preprocessing

In the data preprocessing stage, data cleaning is performed first, including removing invalid data and processing missing values. Here's how:

Remove invalid data: Delete records with empty reviews, invalid ratings, or missing ratings. Handle missing values: For records with missing scores, fill with an average. Data normalization: Convert scores to floating-point numbers between 0-1 to facilitate model training and validation. Next, text preprocessing is carried out, including word segmentation, deactivation of words, part-of-speech annotation, etc. Tokenization: Tokenize the comment text and convert the long text into a word sequence. De-stop words: Remove stop words that do not contribute to sentiment analysis. Part-of-speech annotation: Part-of-speech annotation is performed on the segmented text to facilitate subsequent text representation and model training.

3. Method introduction

3.1. Doc2vec

In recent years, deep learning technology has made great strides in the field of natural language processing (NLP), and Doc2vec is widely used as an effective method of text representation. Doc2vec, proposed by Le and Mikolov in 2014, is a Paragraph Vector-based model that enables efficient representation of text by mapping the entire document into a vector space of fixed dimensions.

The core idea of the Doc2vec model is to treat a document as a sequence of multiple words, and introduce a document vector to represent the entire document. During the training process, the model learns discriminative document vectors by maximizing the similarity of words in documents and the differences between documents. The distributed memory model is jointly trained on each word and document vector, while the distributed bag-of-words model only considers the co-occurrence information of words and documents

3.2. Parameter setting and model training

Parameter setting: In this paper, the following parameters are selected for model training: Dimension: the dimension to which the document vector is mapped, which is set to 100 in this example. Window size: Considering the range of the word context, this article is set to 5; Learning rate: the initial value of the learning rate during model training, which is set to 0.025 in this document. Iterations: the number of iterations of model training, which is set to 10 in this example. Context Size: The context size of the words in the document, which is set to 10 in this document.

Footnotes should be avoided whenever possible. If required they should be used only for brief notes that do not fit conveniently into the text.

Model training: Firstly, each comment in the movie review dataset was taken as a document, and the words in each document were taken as input, and the distributed memory training strategy was used for model training.

3.3. Text vector representation results analysis

After the Doc2vec model is trained, the vector representation of each movie review is obtained. Here's an analysis of the results of a text vector representation:

Vector dimension: Through the Doc2vec model, the author maps each comment to a 100-dimensional vector space, which is conducive to the subsequent text classification task.

Vector similarity: The vector similarity between different comments is calculated, and it is found that the comments with higher similarity tend to have similar emotional tendencies, indicating that the Doc2vec model can better capture the emotional information of the text.

3.4. LSTM model

Long-Term and Short-Term Memory (LSTM) is a specialized recurrent neural network (RNN). Compared with traditional RNNs, LSTMs have stronger learning abilities when dealing with long-term data, and effectively solve gradient loss and gradient burst. The basic idea behind an LSTM is to control the flow of information through three gateway architectures (in, forget, out) and to enable the transmission of information over long distances. By leveraging these three gateway architectures, LSTM models can capture patterns in sequential data on different timescales.

3.5. Model structure design

In this study, an LSTM-based sentiment analysis model was designed for sentiment classification on the film review dataset. The model structure mainly includes the following parts:

Embedding layer: Converts text data into high-dimensional vector representations.,use the pretrained Word2Vec model to obtain the vector representation of the vocabulary and take the vector representation of each vocabulary as input to the LSTM model.

LSTM layer: According to the input text sequence, the LSTM layer can capture the long-distance dependencies in the sequence data. In this study, the author used bidirectional LSTM (Bi-LSTM) to improve the model's comprehension of text data. The Bi-LSTM consists of two LSTM layers, which perform forward and backward propagation of the input sequence, and finally stitch the outputs in both directions.

Fully connected layer: The output of the LSTM layer is connected to a fully connected layer for sentiment classification of text. The author uses the Sigmoid activation function to compress the output of the fully connected layer to the [0, 1] range, which represents the probability that the text belongs to positive sentiment.

3.6. Model parameter optimization

In order to improve the performance of the model, the model parameters need to be optimized. In this study, the following parameters were mainly adjusted:

Learning rate: Learning rate is one of the most important parameters in the training process of deep learning models. Control the speed and convergence of model training by adjusting the learning rate. In this study, a learning rate decay strategy was used, i.e., the learning rate was gradually reduced as training progressed.

Batch size: The batch size determines the number of samples updated each time during model training. The smaller batch size can improve the generalization ability of model training, but the computational efficiency is low. Larger batch sizes can improve computational efficiency, but can lead to model overfitting. In this study, the author balanceed the generalization power and computational efficiency of the model by adjusting the batch size.

Number of iterations: The number of iterations indicates the number of iterations during model training. Too many iterations can lead to overfitting the model, while too few iterations will lead to underfitting the model. Therefore, needing to adjust the number of iterations to find the appropriate model training degree.

3.7. Model training and validation

During model training, an assessment method is used to evaluate the effectiveness of the model. Specifically, the division of film review data into training sets, test sets and test sets. During the training period, the training series is used to train the model and improve the model in the audit series. When the performance of the model in the test set is at its best, use the test kit to evaluate the general performance of the model.

Metrics such as accuracy, completeness, and F1 value are used to evaluate model performance. Accuracy refers to the percentage of samples that are correctly classified by the model. Accuracy: This is the proportion of the sample and the pattern can correctly identify positive emotions. Completeness is the proportion of the sample, and the pattern accurately identifies negative thoughts. The average accuracy and completeness used in the overall performance evaluation are disproportionate.

Through the experiment, the following results were obtained:

The accuracy, precision, recall and F1 values of the LSTM model on the film review dataset were 87.5%, 87.0%, 87.0% and 87.0%, respectively.

Compared to traditional machine learning methods and other deep learning models, the LSTM model performs better in the tasks of emotional study.

The performance of the LSTM model on sentiment analysis tasks can be further improved by adjusting the model parameters.

In summary, the LSTM-based sentiment analysis model has good performance on the film review dataset, and has certain practical value and promotion significance. In future research, more advanced deep learning models can be further explored to improve the performance of sentiment analysis tasks.

4. Results and discussion

4.1. Experimental setup

In order to test the effectiveness of the proposed deep learning-based sentiment analysis model for movie reviews, setting up an experiment is described in detail in this section. The experiment was mainly divided into two parts: model training and model evaluation.

Firstly, in the model training stage, part of the data in the movie review dataset was selected as the training set and the other part as the validation set. By adjusting the model parameters, the optimal model parameters were selected based on the performance on the verification set as a reference. Specifically, the Doc2vec algorithm was used for text vector representation, and the LSTM network was used for sentiment classification. During model training, the cross-entropy loss work was approved and the specification was updated with the best Adams algorithm.

Secondly, in the model evaluation stage, the accuracy, precision, recall and F1 value were used as the evaluation indexes. By comparing the performance of different models, the effectiveness of the proposed model in sentiment analysis task was evaluated.

4.2. Comparison of experimental results

This section compares the proposed deep learning-based sentiment analysis model with several other common sentiment analysis models, including naïve Bayes, support vector machines, and convolutional neural networks (CNNs). The experimental results are shown in Table 1.

| _ | - | | | | |
|-------------------------|----------|-----------|--------|----------|--|
| model | Accuracy | Precision | Recall | F1 value | |
| Naive Bayes | 0.810 | 0.805 | 0.805 | 0.805 | |
| Support vector machines | 0.845 | 0.840 | 0.840 | 0.840 | |
| CNN | 0.865 | 0.860 | 0.860 | 0.860 | |
| Doc2vec+LSTM | 0.875 | 0.870 | 0.870 | 0.870 | |

Table 1. Comparison of experimental results of different sentiment analysis models

As can be seen from Table 1, the proposed model based on Doc2vec and LSTM is superior to several other models in terms of accuracy, precision, recall, and F1 value. This indicates that the proposed model has good performance in the sentiment analysis task of film review.

4.3. Discussion of results

By analyzing the test results, you can draw the following conclusions:

Compared with traditional machine learning models, deep learning models have better performance in sentiment analysis tasks. This is mainly because deep learning models can automatically learn complex feature representations of texts, thereby improving the classification effect. Among the deep learning models, the combined model of Doc2vec and LSTM has the best performance. This is because Doc2vec is able to efficiently convert text into vector representations, while LSTM is able to capture long-distance dependencies in text.

Experimental results show that the proposed model has high accuracy and stability in the sentiment analysis task of film reviews. However, the model still has mispositives in some cases, possibly because of noise and ambiguity in the movie review data.

4.4. Choice of visualization technology

In the current research context, data visualization technology is increasingly widely used in the field of sentiment analysis, which can present complex data in an intuitive and visual way, which is convenient for people to understand and analyze. The following visualization techniques were selected for this study:

First of all, Word Cloud technology can display keywords that appear frequently in the text in different sizes and colors, which helps to quickly capture the core content of the text. This study uses word cloud technology to show high-frequency words in film reviews to reveal the emotional tendencies of reviews.

Secondly, Radar Chart is a visualization method to show the distribution of multi-dimensional data, and this study uses radar chart to show the distribution of different emotion categories in each dimension, so as to conduct a more in-depth analysis of sentiment categories.

4.5. Visualization of results analysis

By visualizing a dataset of movie reviews, the following results are presented: First, the word cloud results show that high-frequency words such as "good-looking", "wonderful" and "like" in positive emotional comments and high-frequency words such as "disappointed", "bad review" and "boring" in negative emotional comments are high-frequency words such as "disappointed", "bad review" and "boring". These high-frequency words intuitively reflect the emotional tendencies of film reviews (Figure 1).



Figure 1. Word cloud map of the movie "Prize Claw"

Secondly, the radar chart analysis shows that the distribution of positive emotions in each dimension is relatively balanced, while the negative emotions are prominent in some dimensions, such as "plot" and "actors". This suggests that negative sentiment reviews are more problematic in these areas (Figure 2).

Proceedings of CONF-MLA 2024 Workshop: Securing the Future: Empowering Cyber Defense with Machine Learning and Deep Learning DOI: 10.54254/2755-2721/94/2024MELB0075



Figure 2. Movie evaluation radar chart

4.6. Comparative analysis

In order to further verify the validity of the visualization results, this study compares the sentiment analysis results under different visualization techniques. The results showed:

Word cloud technology can intuitively display the emotional tendency of text data, and pay more attention to the display of keywords. Radar charts have high accuracy in analyzing the distribution of sentiment categories, but they cannot visually display the distribution of text features. In contrast, word cloud technology can better make up for this shortcoming.

Based on the comparative analysis, this study concludes that the application of visualization technology in sentiment analysis has certain practicability and effectiveness, but different visualization techniques have their own advantages and disadvantages, and the appropriate technology should be selected according to the actual needs. At the same time, visualization technology needs to be continuously optimized and improved in practical applications to improve the accuracy and efficiency of sentiment analysis. Based on the comparative analysis, this study concludes that the application of observation techniques is simple and effective in mind analysis, but each observation mechanism has its own advantages and disadvantages, so it is necessary to choose the appropriate technique according to the actual needs. At the same time, vision technology needs to be continuously improved in practical applications to improve the accuracy and efficiency of the actual needs. At the same time, vision technology needs to be continuously improved in practical applications to improve the accuracy and efficiency of the actual needs. At the same time, vision technology needs to be continuously improved in practical applications to improve the accuracy and efficiency of opinion analysis.

5. Conclusion

Through an in-depth discussion on the application of deep learning in sentiment analysis of movie review datasets, this study draws the following conclusions: firstly, deep learning has significant application value in the field of sentiment analysis, especially the Doc2vec and LSTM models perform well in text vectorization and sentiment classification tasks. Secondly, after data preprocessing, text vectorization, model training and validation, the author successfully built a sentiment analysis model based on deep learning, which outperformed traditional machine learning models on experimental datasets. Finally, through data visualization technology, the author visually displays the results of the model, which further verifies the feasibility and effectiveness of the model.

Proceedings of CONF-MLA 2024 Workshop: Securing the Future: Empowering Cyber Defense with Machine Learning and Deep Learning DOI: 10.54254/2755-2721/94/2024MELB0075

Study deeply and be practical when checking emotions. In the future, the following research directions need to be addressed: First, deep learning models such as convolutional neurons should be studied to make emotion diagnosis effective. Studies have been conducted to integrate text, images, audio, and other data from a variety of sources to improve the accuracy and breadth of sensitivity analysis. Finally, it encourages more methods of emotional research, encouraging the pursuit of opinions and public opinion in film.

References

- Zhao Y Y, Qin B and Liu T 2010 Sentiment analysis of texts. Journal of Software, 21(8), 1834 1848.
- [2] Turney B W 2007 Anatomy in a modern medical curriculum. The Annals of The Royal College of Surgeons of England, 89(2), 104-107.
- [3] Chen L and Yang Y 2013 Emotional speaker recognition based on i-vector through atom aligned sparse representation. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 7760-7764.
- [4] Zhou Y M, Yang C and Yang A M 2013 Construction Method of Chinese Sentiment Dictionary for Text Sentiment Analysis. JOURNAL OF SHANDONG UNIVERSITY (ENGINEERING SCIENCE), 43(6), 27-33.
- [5] Li Y, et al. 2008 Research on Text Sentiment Classification Based on Phrase Patterns. Journal of Mudanjiang Normal University.
- [6] Pang B, Lee L and Vaithyanathan S 2002 Thumbs up: sentiment classification using machine learning techniques. Proceedings of the ACL-02 conference on Empirical methods in natural language processing, 10.
- [7] Wikarsa L and Thahir S N A 2015 Text mining application of emotion classifications of Twitter's users using Naive Bayes method. 2015 1st International Conference on Wireless and Telematics (ICWT). IEEE, 1-6.
- [8] Xin Y X and Wang X D 2021 Research on Sentiment Analysis of Film Reviews Based on Text Mining. Journal of Mudanjiang Normal University (Natural Science Edition), 25-28.
- [9] Liu H Q, Guo Y B nad Li W H 2020 Cross-domain sentiment analysis method based on Bayesian net[J] Computer Applications & Software, 37(12), 119-126.
- [10] Govindarajan M 2013 Sentiment analysis of movie reviews using hybrid method of naive bayes and genetic algorithm. International Journal of Advanced Computer Research, 3(4), 139.
- [11] Jiang J 2017 Text sentiment analysis of social media. Nanjing University of Science and Technology.