Depression Prediction Model based on NLP

Yuke Zhou

School of Cyberspace Security, Hangzhou Dianzi University, Zhejiang, China

22080733@hdu.edu.cn

Abstract. With the acceleration of modern society's pace, depression has become a common mental health issue. This study aims to develop a depression detection model based on natural language processing (NLP) technology to automatically identify potential depression patients. First, collect data from the Weibo platform using web scraping techniques, and then use NLP tools such as XLNet for in-depth analysis of the text data. In the model evaluation phase, the paper uses key metrics such as accuracy (ACC) to assess the effectiveness of NLP techniques in predicting depression. Experimental results indicate that the XLNet model performed the best among all tested models, achieving an accuracy of 81%, thus confirming the feasibility of using NLP technology for depression prediction. The significance of this study lies in providing a new method for mental health monitoring, which aids in the early detection and intervention of depression, with important social value and application prospects.

Keywords: NLP, Depression, XLNet.

1. Introduction

Depression is a common mental disorder in modern times, characterized by persistent low mood, loss of interest, or lack of joy [1]. According to estimates by the World Health Organization, approximately 5% of adults worldwide suffer from depression. As the most common and burdensome emotional disorder, depression is not far from us [2]. In the prevention and treatment of depression, early detection not only improves treatment outcomes and reduces the physical and mental burden on patients but also effectively prevents severe consequences, including suicide. Mild depression can sometimes be alleviated through non-pharmacological treatments, but for moderate to severe cases, medication is often necessary.

Currently, the assessment and diagnosis of mental disorders primarily rely on psychiatrists observing and conducting clinical interviews with patients to determine the extent to which they meet diagnostic criteria. This is supplemented by rating scales such as the Beck Depression Inventory (BDI), Pittsburgh Sleep Quality Index (PSQI), Hamilton Depression Rating Scale (HAMD), and Hospital Anxiety and Depression Scale (HADS) to achieve this assessment [3]. In addition, diagnostic methods based on biological factors, brain imaging, and multimodal data are used. However, all diagnostic methods require proactive seeking of medical help or self-assessment by the patient. Often, patients only realize their depression after it has progressed to a certain extent, which increases the difficulty and complexity of treatment.

Given the limitations of existing diagnostic methods, this study focuses on exploring a new approach to the early detection of depression. The paper plans to utilize natural language processing (NLP) technology to conduct an in-depth analysis of users' statements on social media. Social media, as a key platform for modern individuals to express emotions and share life experiences, contains a wealth of emotional and psychological information. By analyzing users' language patterns, emotional tendencies, and interaction dynamics, it is possible to uncover signals that may be related to depression.

The goal of this study is to develop an NLP-based depression detection model that can automatically identify users at risk of depression, providing timely feedback and recommendations. This study aims to achieve early detection and intervention of depression through this method, reducing missed diagnoses and improving the timeliness and effectiveness of treatment.

2. Data and Method

2.1. Data and processing

The dataset is publicly available on GitHub and consists of user data scraped from Weibo, collected in 2020.

Dataset Statistics: 32,395 users, including 10,198 with depression. Load the depression patient dataset and the normal dataset, merge the two datasets, remove irrelevant columns such as dates, and add a label column, marking depression patients as 1 and non-depression patients as 0. Split the dataset into training and testing sets with a ratio of 9:1.

2.2. Model

In the Support Vector Machine (SVM) algorithm, data points are depicted as coordinates in an n-dimensional space, corresponding to the number of features they possess. Then, it divides the points into two categories using a hyperplane. The Support Vector Machine is a discriminative classifier formally described by a distinct hyperplane (SVM). In other words, given labelled training data, the algorithm generates an ideal hyperplane to classify new examples (supervised learning). This hyperplane divides the plane into two sections, each containing a different class. In linear SVM, the learning of the hyperplane is completed by transforming the problem using some linear algebra [4]. This is where the kernel function comes into play. In the experiment, the paper used a 9:1 training-test split for the linear SVM. The paper employed scikit-learn's kernel functions for classification.

The Naive Bayes classifier is a probabilistic classification algorithm based on Bayes' theorem, which was proposed by Thomas Bayes. In classification tasks, its core purpose is to determine the best correspondence between new data items and a specific set of categories. To make this mapping probabilistically computable, some mathematical manipulations are performed to transform joint probabilities into the multiplications of prior probabilities and conditional probabilities[5-7].

The CNN model has a unique hidden layer structure and contains more unique layer types when applied to training[8,9]. The core components of the hidden layers in convolutional neural networks consist of three fundamental types: convolutional layers, pooling layers, and fully connected layers. These layers primarily include convolutional layers, pooling layers, and fully connected layers. With the advancement of deep learning technology, some novel structures, such as Inception modules and residual connections, have also been integrated into the design of CNNs. Convolutional and pooling layers are the core components of CNNs, which extract image features through different convolution operations. Furthermore, to enhance the performance of the model, some CNNs have introduced advanced convolution techniques, including tiled convolution, transposed convolution, and dilated convolution, to strengthen the model's ability to capture image details. The application of these technologies has led to significant results for CNNs in tasks such as image recognition and classification.

Recurrent Neural Networks (RNNs) are a type of neural network designed for sequence data. By introducing recurrent connections in the hidden layers, RNNs can maintain and propagate information, allowing the network to capture and utilize temporal dynamics across different time steps. This makes them effective for handling tasks involving time series, such as natural language processing and speech recognition. In an RNN, the state of the hidden layers is updated through an activation function and passed to the output layer to produce prediction results. When the network's structure is unfolded over time, it demonstrates its ability to process sequence data.

XLNet is an advanced pre-trained language model based on the Transformer architecture, using Transformer-XL as its foundational model structure. It is capable of effectively handling long-distance dependency issues. By incorporating relative position encodings and variable-length attention mechanisms, XLNet significantly enhances the model's ability to handle long texts.

XLNet also introduces a dual-stream self-attention mechanism, comprising the Content Stream and the Query Stream. The Content Stream is responsible for encoding the content information of words, while the Query Stream predicts the positional arrangement of the words. This design allows XLNet to effectively leverage both preceding and following context, thereby making full use of the available information.

The author added a linear layer on top of the XLNet model for classification, creating a sentence classifier to determine whether a sentence is from a depressed user. The preprocessed data is fed into the model for training. The design of this classifier is based on the following steps:

The preprocessed data, including the vectorized text results, is fed into the model. The preprocessing steps involve text cleaning, tokenization, removal of stop words, etc., to ensure the quality of the input data.

Utilizes XLNet's deep bidirectional contextual understanding to encode the input data. This step captures subtle emotional and semantic information in the text, providing rich feature representations for subsequent classification.

On top of the XLNet encoding, a linear layer is added as the classifier. This layer learns the mapping between text and depression labels, enabling the discrimination of depressive speech[10].

2.3. Evaluation Metrics

True Positives (TP): The model correctly predicts actual depressive samples as depressive. True Negatives(TN): The model correctly predicts actual non-depressive samples as non-depressive. False Positives(FP): The model incorrectly predicts actual non-depressive samples as depressive. False Negatives(FN): The model incorrectly predicts actual depressive samples as non-depressive.

Accuracy (ACC): Measures the proportion of correct predictions by the model, including both correctly predicted depressive and non-depressive samples. The formula is:

$$\text{ACC} = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

In this model, the accuracy is 80.12%, indicating that the model makes correct predictions in most cases.

Precision: Measures the proportion of actual depressive samples among all samples predicted as depressive by the model. This metric is crucial for avoiding false positives, where non-depressive samples are incorrectly classified as depressive. The formula is:

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

The model's precision is 84.01%, indicating that more than 84% of the samples identified as depressive are depressive.

Recall (or True Positive Rate/Sensitivity): Measures the proportion of actual depressive samples correctly identified as depressive among all actual depressive samples. This metric is crucial for ensuring that all depressive samples are identified. The formula is:

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

The model's recall is 79.95%, indicating that nearly 80% of all actual depressive samples are correctly identified.

3. Results Analysis

The differences between models are shown in Table 1.

Table 1. Model Comparison Table

Model name	SVM	Naive Bayes	CNN	RNN	XLNet
accuracy	0.61	0.61	0.76	0.76	0.81
precision	0.67	0.66	0.81	0.82	0.83
recall	0.52	0.53	0.71	0.71	mz

Compared to other NLP models, XLNet performs better on several metrics, with an accuracy of approximately 80%, precision of around 85%, and recall of about 77%.

4. Conclusion

Using NLP to analyze user social media statements can be one of the methods for detecting and preventing depression. Compared to other NLP models, XLNet performs better on several metrics, with an accuracy of approximately 80%, precision of around 85%, and recall of about 77%. Due to device memory and time constraints, the batch size for the training process was set to a smaller value of 8. It may be beneficial to spend more time on training to achieve a better model.

The XLNet model is based on Transformer architecture and requires input as fixed-length token sequences. For short sentences, padding can be applied, but for long sentences, truncation is necessary. This limitation is not prominent in domains like social media, which primarily use short sentences. However, when users write long texts, the prediction results may not be optimal.

The existing dataset consists of user data from 2020, and the prediction results may not be ideal for the current social media environment.

Different stages of depression exhibit distinct characteristics, but the dataset does not differentiate between severe and mild depression. Observations of the dataset reveal that some depressed individuals make positive statements on social platforms, while some non-depressed users may express negativity. Future research could focus on this aspect to further refine the model.

References

[1] Xie L Y 2024 A Review of Depression Diagnosis Methods and Related Intervention Research Frontier of Social Sciences 13(1) 1–7

[2] Yang Z, Dai Z, Yang Y et al 2020 XLNet: Generalized Autoregressive Pretraining for Language Understanding arXiv

[3] Aswathy K S, Rafeeque P C and Reena Murali 2019 Deep Learning Approach for the Detection of Depression in Twitter Int. Conf. on Systems Energy and Environment

[4] Hemanthkumar M and Latha 2019 Depression Detection with Sentiment Analysis of Tweets Int. Research J. of Engineering and Technology (IRJET)

[5] Yang F-J 2018 An Implementation of Naive Bayes Classifier Int. Conf. on Computational Science and Computational Intelligence (CSCI) (Las Vegas, NV, USA) pp 301–306, doi: 10.1109/CSCI46756.2018.00065

[6] Bayes T 1763 An Essay Towards Solving a Problem in the Doctrine of Chances Philosophical Transactions of the Royal Society 53(1) 370–418

[7] Tabak J 2004 Probability and Statistics: The Science of Uncertainty (New York: Facts On File, Inc.) pp 46–50

[8] Dai D 2021 An Introduction of CNN: Models and Training on Neural Network Models Int. Conf. on Big Data, Artificial Intelligence and Risk Management (ICBAR) (Shanghai, China) pp 135–138, doi: 10.1109/ICBAR55169.2021.00037

[9] Bora K et al 2023 Brain Tumor Detector Int. J. for Research in Applied Science and Engineering Technology 11(4) 141–145

[10] Li Z 2022 Deep Learning-Based Prediction of Survival in GBM Brain Cancer Patients Yunnan University