Depression Identification of Students Based on Campus Social Platform Data and Deep Learning

Guang Feng Zhao1 and Lin Fang Sun2

1College of Sports and Health, Shandong Sport University, Jinan 250102, Shandong, China
2College of Sports and Art, Shandong Sport University, Jinan 250102, Shandong, China

Correspondence should be addressed to Lin Fang Sun; sunlinfang@sdpei.edu.cn

Received 19 January 2022; Revised 19 February 2022; Accepted 1 March 2022; Published 27 April 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Guang Feng Zhao and Lin Fang Sun. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Depression is one of the most common psychological problems faced by human society. Because of less social experience, low psychological endurance, and the multiple responsibilities of future families and society, college students have become one of the most vulnerable groups to suffer from depression. This paper explores an automatic identification method to identify patients with early depression tendency through deep mining of online information of campus social platform users. First, we comprehensively analyze the common characteristics of emotion and behavior on the campus social platform for depression. Secondly, the experimental corpus is formed by preprocessing operations such as deprivation, word segmentation, and denoising of the original data. Finally, the depression recognition is transformed into a text classification problem, and a shallow support vector machine and a deep convolutional neural network model are, respectively, constructed based on the experimental corpus. Combined with the features of depression blog, the algorithm was further improved, and a dual-input convolutional neural network algorithm compatible with multiple features was proposed. The experiments showed that the recognition rate was effectively improved.

1. Introduction

Depression is one of the most common psychological problems faced by human society. The harm and impact of depression are very serious. Suicide rates among people with depression are very high, with 3,000 people with depression committing suicide every day worldwide [1]. Considering the personal suffering of depression patients, the impact on relatives and friends, and the resources spent on treatment, it can be said that depression has become a heavy burden on human society. In college students, a special group with less social experience and low psychological endurance, but also bearing multiple responsibilities of future family and society, the incidence of depression is significantly higher than that of the general population.

In order to fight against depression, people have done a lot of research on its causes, diagnosis, and treatment and put forward diagnostic methods based on psychology and physiology, such as various scales. Traditional diagnostic methods mainly obtain basic information about people’s psychological and physiological states through verbal communication, questionnaires, physical examinations, etc. From this information, the basis for diagnosis is obtained. In the process of diagnosis, the acquisition, processing, and analysis of information consume a lot of time, money, and material. The traditional method also faces a problem: very low medical treatment rate. Due to the lack of understanding of psychological problems, many people do not realize that they are suffering from depression and do not know whether to seek help when they have problems with their physical and mental health. Some people have a biased understanding of psychological problems such as depression, and they avoid treatment because they feel that “family ugliness cannot be made public.” There are also some people who cannot seek medical treatment due to the relative lack of medical resources. Coupled with the influence of factors such as a certain missed diagnosis rate in medical institutions, the overall actual consultation rate of patients with
depression is very low. Traditional diagnostic methods can only detect a subset of the many people with depression—the ones who come to a medical facility for help. That is, other people will face depression helplessly. Passively waiting for depression patients to seek help as is currently done will result in very low rates of consultation and depression will continue to rage. If medical institutions can take the initiative to look for patients with depression, such as directly conducting psychological surveys on college students, the rate of seeing a doctor will be greatly improved. Of course, the cost of doing so is relatively high. Consider that when a person seeks medical help for depression, the depression has already taken a toll on him, and treatment is more difficult. We need “active defense,” which is to detect the person when the depressive tendencies are not too severe, so that help or treatment can be given in a timely manner.

The advent of the Web2.0 era and the emergence of online social media, such as blogs, campus social platforms, and other social networks, provide a place for many people with depression tendencies to vent their emotions. At the same time, based on the analysis of massive interactive information on social networks, we may provide a platform for actively discovering people who are prone to depression [2]. Especially as each high school has its own social platform, the students of the school publish information on the social network, and the campus management department can master a large amount of social network data. These data record students’ thoughts, opinions, and details about their lives. By mining the massive data in the campus social platform, a lot of important information and knowledge can be found. In our recent online information observation on campus network online users, we found that some users, due to the high pressure of life, showed many psychological problems [3]. In severe cases, the language manifestations of depression and suicide occurred.

This paper will propose a classification model based on natural language processing and machine learning that takes into account both generality and accuracy, to actively and efficiently discover students with depression tendency from campus social platforms. The method proposed in this paper allows us not to wait for depressed patients to seek help, but to actively search for people with depression tendency from the crowd, which will enable medical institutions and nonprofit organizations to gain the initiative in fighting depression. Since the recognition process is completed automatically by computer, the speed and accuracy have unique advantages compared with manual work, so that we can quickly and timely find people who are prone to depression and can find the target in time, which greatly improves the efficiency of intervention [4]. For college students with high network activity, it is appropriate to use the method proposed in this paper to assess the state of depression. This method will greatly improve the ability of universities and other institutions to deal with students’ depression, thereby reducing the harm caused by depression to the student population.

2. Identification Process

2.1. Model Design Idea. The campus social platform of colleges and universities is an important platform for students to communicate. It has the characteristics of huge user scale and convenient data acquisition. It is an important data source for campus management. Therefore, this paper selects users of a university campus social platform as the research object. Based on the data of campus social platforms, an effective method to automatically identify depression-prone users in campus social platforms is established. The process of model design is shown in Figure 1.

2.2. Annotated Account. This study is based on the assumption that “the language and behavior of campus social platform users who are prone to depression are different from normal users” [5]. To test this hypothesis, we first need to find a certain number of users who are prone to depression and normal users to establish a sample data set.

We randomly obtained a certain number of user IDs and manually labeled these accounts into two categories: “depressive tendencies” and “normal.” In order to reduce the impact of class imbalance problem on the classifier, we do not keep the UID of all “normal” users. Then, the campus social platform information of these users is captured through the API provided by the Sina campus social platform, and after a certain preprocessing, it is used for training and testing the machine learning model. The labeling of the samples is done manually, and the process is as follows:

(1) Receive certain training, such as learning the basic knowledge of depression, diagnostic scales, and judgment of depression tendency.
(2) Groups that users with depression tendencies from campus social platforms may gather.
(3) Two of them independently annotated the obtained user set and divided the users into two categories: “with depression tendency” and “without depression tendency.”
(4) The labeling results of the first two people were checked by a third person (the labeling rate was found to be more than 90% identical), and the labeling results were independently corrected.

2.3. Experimental Data Acquisition and Preprocessing. Data preprocessing is to process the original experimental data into experimental corpus that can be directly operated by the computer, and its operation mainly depends on the experimental data and algorithms. In this study, a supervised deep learning algorithm was used to perform binary
classification of text on campus social platforms, thereby realizing depression identification. Its preliminary work mainly includes labeling, word segmentation, denoising, feature, and algorithm selection; these operations have a direct impact on the subsequent experimental results. Reasonable preprocessing operations are a crucial step in text classification, and these operations are described in detail in this section.

Text preprocessing mainly includes word segmentation and denoising, and the specific process is shown in Figure 2. Since Chinese words do not have obvious delimiters like English words, it is necessary to divide sentences into words one by one through word segmentation. The language expressions of campus social platforms are relatively casual and do not strictly follow traditional language norms and often insert special symbols [6]. It belongs to the text with high noise content, and denoising processing is very necessary.

This experiment uses Jieba word segmentation implemented in Python to segment the text. Jieba comes with a dictionary containing more than 20,000 Chinese words, and its word segmentation speed and effect are relatively good. This experiment adopts Jieba’s precise mode, which combined with the domain dictionary can cut the experimental data more accurately, which is suitable for text analysis of campus social platforms. In order to further maintain the integrity of the semantics, some domain terms and network neologisms are avoided to be excessively segmented [7].

2.4. Feature Selection. Data features are the main internal factors affecting text classification, and reasonable feature selection can effectively improve the accuracy of experiments. By researching and analyzing the characteristics of the “tree hole” campus social platform, referring to depression-related materials, we conducted in-depth mining from two aspects of the text itself and extended information and comprehensively extracted the characteristics closely related to the depression campus social platform. In terms of content, it mainly includes two features, namely, the semantic features of the text itself and the dictionary features. In addition, some features expanded from the content can also reflect the behavior of the campus social platform of depression [8]. To sum up, the features selected in this paper mainly include three modules: semantic features, extended features, and dictionary features. The specific description is shown in Table 1.

These three characteristics are described in detail below.

Semantic features refer to the semantic information contained in the text itself, which can reflect the structure and contextual semantic relationship and reflect the publisher’s expression form and overall emotional trend, which is of great significance for recognition. This part of the features will be automatically extracted by deep learning algorithms. Dictionary feature is whether the text contains dictionary words. The anomalies of these campus social platform texts are mainly reflected in two aspects: emotion and behavior. For example, it contains a large number of negative words such as “pain,” “life is better than death,” and various kinds of thought expression and suicide methods such as “want to die,” “cut your wrist,” and “burn charcoal” are also frequent sentences in the “tree hole.” These sensitive words fully reflect the patient’s condition, so the dictionary features of these words are very important.

In this paper, a dictionary database is constructed for the field of depression campus social platform. The specific dictionary features are described in Table 2.

Affected by depression, the campus social platform behaviors of depressed patients have some notable characteristics, which are extended from the campus social platform content, so they are called extended characteristics, for example, the release time of the campus social platform, the length of the text, etc. These extended features are closely related to the symptoms of depression, and only by fully mining these valuable features can the identification of depression be optimal [9]. Through a comprehensive analysis of the symptoms of depression and the characteristics of the campus social platform, it is found that the text of the campus social platform for depression has many individual and abnormal characteristics and further artificially extracts each extended feature. The main features are as follows:

(1) Through statistical analysis, the most frequent update time is from 10:00 pm to 2:00 am.
(2) The language expressions of depression patients on campus social platforms are more casual, the format is not standardized, and the length is generally shorter.
(3) Depressed patients pay less attention to the outside world, and the general campus social platform is original.
(4) Depression is more self-focused, interacts with others, and pays less attention to others, while the first-person singular “I” is often used in language, and the plural “we” is used less.
(5) Patients frequently use emoji and some suicide-related symbols on campus social platforms.
(6) Depression is more inclined to be puzzled and rhetorical about life experiences on campus social platforms and use questions more frequently.

3. Depression Recognition Model

This paper designs three different models for student depression identification.
Table 1: Features and definitions.

<table>
<thead>
<tr>
<th>Features</th>
<th>Definitions</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantic features</td>
<td>Semantic information of text expressions on campus social platforms</td>
<td>Text itself statement</td>
</tr>
<tr>
<td>Extended features</td>
<td>Expanded features strongly associated with depression in campus social platforms</td>
<td>Time, length, originality, etc.</td>
</tr>
<tr>
<td>Dictionary features</td>
<td>Sensitive vocabulary that reflects depression tendency contained in campus social platforms</td>
<td>Sentiment dictionary, keyword dictionary, etc.</td>
</tr>
</tbody>
</table>

Table 2: Dictionary features.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Feature definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentiment dictionary features</td>
<td>Whether contains the vocabulary of the emotional dictionary</td>
</tr>
<tr>
<td>Emoji dictionary features</td>
<td>Whether contains the vocabulary of the emoji dictionary</td>
</tr>
<tr>
<td>Mood particle dictionary features</td>
<td>Whether contains the vocabulary of the modal particle dictionary</td>
</tr>
<tr>
<td>New word dictionary features</td>
<td>Whether contains the vocabulary of the online new word dictionary</td>
</tr>
<tr>
<td>Keyword dictionary features</td>
<td>Whether contains the vocabulary of the keyword dictionary</td>
</tr>
<tr>
<td>Suicide tool dictionary features</td>
<td>Whether contains the vocabulary of suicide tools dictionary</td>
</tr>
<tr>
<td>Dictionary of behavioral symbols</td>
<td>Whether contains the vocabulary of the behavioral symbol dictionary</td>
</tr>
</tbody>
</table>

Figure 3: Identification algorithm based on SVM.
3.1. Depression Recognition Algorithm Based on Support Vector Machine. The input to the support vector machine algorithm is a vector, so the features need to be vectorized first [10]. In this paper, the support vector machine algorithm is used to classify text based on extended features and dictionary features, so it is necessary to vectorize these two features. The specific vectorization rules are as follows: In the extended feature, the description length of the campus social platform, the number of interactions on the campus social platform, the degree of social activity, the degree of collective attention, and the degree of self-attention are themselves numbers. Whether it is original, whether to use a positive expression map, and whether to use a negative expression map, the value of the feature value is 1, otherwise 0. All dictionary features are the number of corresponding dictionaries; if not, take 0.

The algorithm implementation process based on SVM is shown in Figure 3. The SVM-based student depression recognition algorithm is as follows:

1. Feature selection and vectorization: According to the feature selection and vectorization rules introduced above, on the preprocessed experimental corpus, Python language programming is used to realize the selection and vectorization of extended features, and the dictionary features are obtained by scanning the dictionary library.

2. Model configuration: The experimental tool uses LIBSVM (A Library for Support Vector Machines) integrated tool for text classification experiments, which is the most widely used SVM algorithm tool developed by the Institute of Information Engineering, National Taiwan University. The tool is written based on C language, including standard SVM algorithm, probability output, support vector regression, multiclass SVM, and other functions and has call interfaces in JAVA, Python, R, MATLAB, and other languages. It is a relatively hot tool at present, and the experimental speed and effect are relatively good. This experiment calls the Python interface for model training.

3. Model training and testing: The corpus is divided into training corpus and test corpus according to the ratio of 7:3, in which three types of dictionary features, extended features, dictionary features, and extended feature fusion, are trained and tested. The optimized parameters and training corpus are used as input for model training, and then the generated model is tested through the test corpus to output the experimental results.

3.2. Depression Recognition Algorithm Based on Convolutional Neural Network. The convolutional neural network model structure mainly contains five layers, each layer has its own function and is connected before and after. The input layer is used to input the processed matrix vector, which is converted from the word vector of the text. The main component of the convolution layer is a feature extractor (convolution kernel), which is mainly used to extract features and output feature maps. The convolution layer can contain multiple layers, and the front and rear layers are connected to each other. The pooling layer is mainly used to process a large number of feature maps output by the convolutional layer to reduce the amount of data while maintaining important feature information. The fully connected layer converts the features extracted by the previous layers into one-dimensional features [11]. The output layer takes the one-dimensional features of the fully connected layer as input and then uses a classification algorithm for classification, such as softmax logistic regression. The specific description of each layer is as follows:

The input layer is used to obtain experimental data, and the input of the convolutional neural network for text classification is a matrix vector. Therefore, it is necessary to convert the vocabulary into a vector and use the word vector of each word in the sentence as a row to form a matrix vector of the sentence. Convolutional layers are the fundamental operations in convolutional neural networks. The convolution operation is actually a mathematical operation. This operation generally includes input, kernel function, and output feature map. Convolution is a local operation, and the local feature information of the data is obtained by applying a certain size of convolution kernel to the local area of the input data. Pooling layers are nonlinear downsampling algorithms. The pooling function replaces the data at the position with the overall statistical value of the adjacent data at the current position, which plays a dimensionality reduction role and ensures that the output data does not change much. Commonly used pooling functions are the maximum pooling function and the average pooling function. The fully connected layer is to stitch together the two-dimensional feature vectors output by the pooling layer and output the probability of each category through the softmax layer.

This paper uses the word2vec tool to generate word vectors, and reasonable parameter selection can effectively improve the training efficiency. In this paper, based on the specific conditions of the experiment, several important parameters are selected as follows: Compared with the CBOW model, the training time of the Skip-gram model is longer, but the accuracy is generally better than that of the CBOW model. Considering that the experimental corpus is relatively moderate, in order to obtain better experimental results, the Skip-gram model is selected for this experiment. Depending on the size of the context window of the word, the reflected information is also different. Generally, a small context window is more conducive to learning contextual semantic features and relationships. According to the characteristics of the experimental corpus, this paper adopts two values of 5 and 10, respectively. The training algorithm of the experiment adopts softmax, which has better effect with rare words and is suitable for the scene of this experiment. The specific parameters are shown in Table 3.

The experimental process based on convolutional neural network is as follows:
(1) Word vector generation.

(2) Model parameter configuration: The algorithm used in this experiment is TextCNN, which is implemented in Python language based on the TensorFlow framework.

(3) Model training and testing: Divide the preprocessed experimental corpus into training corpus and test corpus in a ratio of 7:3, and use convolution kernels of 2, 3, and 4 to conduct experiments. First, the TextCNN model is trained through the training corpus and finally test corpus test model.

The whole recognition process based on convolutional neural network is shown in Figure 4.

### 3.3. Depression Recognition Algorithm Based on Dual-Input Convolutional Neural Network

As a shallow machine learning, support vector machine has better classification effect based on shallow extended features and dictionary features [12]. Convolutional neural network, as a deep machine learning algorithm, can automatically extract semantic features of text sequences for fast learning and efficient classification. However, neither of these two algorithms is compatible with all features at the same time, which reduces the recognition rate of depression to a certain extent. Expanding compatibility is the core of solving the above problems. This paper improves the convolutional neural network algorithm to achieve compatibility with all features.

By analyzing the model structure of the convolutional neural network, it can be found that its convolutional layer and pooling layer automatically extract features and feature dimension reduction for the input data, respectively. Finally, the extracted features are processed based on the full connection layer, and the predicted values are output through the full connection layer. The output is equivalent to all the features automatically extracted by the convolutional neural network and finally classified by the output layer. Based on this, this paper proposes a dual-input convolutional neural network (dual-input-CNN, DI-CNN for short) algorithm. Its core idea is that the first four layers of convolutional neural network are unchanged; only the results of the full connection layer are integrated with the feature vector formed by the fusion of extended features and dictionary features and then classified through the softmax layer, so as to realize all feature fusion as input. The model structure is shown in Figure 5.

### 3.4. Experimental Results

In order to verify the influence of depression dictionary on depression recognition, the experimental evaluation standards used in this paper are precision rate P (precision), recall rate R (recall), and F-measure (F-measure) as evaluation standards. All experiments in this paper use this criterion, and they are defined as follows:

\[
P = \frac{T}{E},
\]

\[
R = \frac{T}{N},
\]

\[
F = \frac{2 \cdot P \cdot R}{(P + R)},
\]

where T is the number of correctly classified samples; N is the actual number of samples of a certain category; E is the

### Table 3: Model parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentences</td>
<td>Can be a list, transformed from the original corpus</td>
<td>Depression campus social platform corpus text</td>
</tr>
<tr>
<td>Sg</td>
<td>Used to set the training algorithm, 0 corresponds to the CBOW algorithm, 1 corresponds to the Skip-gram algorithm</td>
<td>1</td>
</tr>
<tr>
<td>Size</td>
<td>Refers to the dimension of the feature vector</td>
<td>200</td>
</tr>
<tr>
<td>Window</td>
<td>Indicates the maximum distance between the current word and the predicted word in a sentence</td>
<td>5, 10</td>
</tr>
<tr>
<td>Hs</td>
<td>1 corresponds to hierarchical softmax algorithm, 0 corresponds to negative sampling algorithm</td>
<td>1</td>
</tr>
</tbody>
</table>
number of samples predicted by the classification model as a certain category. The experimental results are shown in Table 4.

The experimental results and graphs show that the recognition rates of the three algorithms are DI-CNN, SVM, and CNN, respectively. At the same time, comparing SVM and CNN separately, the effect of SVM is better, which further shows that the artificially extracted shallow features have a better effect on the recognition rate of depression.

4. Application Prospects of the Model

The first threshold of depression treatment and research is identification. The identification of depression in this paper is only a preliminary judgment on whether there is a tendency to depression. However, in practical applications, this is only the initial stage, and the relevant medical staff still need to dig deep into the patient’s situation. It mainly involves the following aspects: the cause of depression, the type of disease, the severity of the disease, etc. In fact, on the social platform, the campus social platform that is updated at high speed every day contains the daily life trajectories and emotional ups and downs of users. Fully mining this information can greatly reduce the workload of people and assist medical staff to implement treatment quickly and accurately. Therefore, in-depth analysis of the campus social platform will lead to fragmented information, which makes up for the lack of dictionaries and comprehensive analysis of information. Dictionaries are gathered into a comprehensive dictionary library for the depression campus social platform, which makes up for the lack of dictionaries in this field.

To sum up, further build a depression rescue chain with the school social platform as the core, fully tap the text information on the platform with the help of natural language technology, and monitor and extract the following user information: 1. The user’s symptoms, etiology, degree of illness, and whether there are suicidal tendencies; 2. Personal information, social relations, time, and space. Based on this, the patient’s condition and information knowledge map are, respectively, constructed to realize the full integration of technicians, medical personnel, psychologists, and rescuers. The knowledge map is mined and constructed by technical personnel, analyzed and diagnosed by medical personnel, and treated and psychologically counseled by medical personnel or psychologists according to different situations. If there is suicidal behavior, rescuers can quickly locate and rescue through the user information knowledge map. This will be the development direction of college students’ depression identification and treatment.

5. Conclusion

In view of the current problems in depression identification, this paper proposes a depression identification method based on campus social platform text and deep learning. It not only effectively avoids the problem of patients not taking the initiative or cooperating, but also can obtain sufficient data for research. This method of turning passive into active discovery reduces the harm caused by direct diagnosis of patients and provides support for medical staff to quickly identify and treat patients. At the same time, in order to improve the feasibility and recognition rate of the algorithm, this paper has made the following improvements:

1. The construction of the dictionary: fully excavate the vocabulary of depression on social platforms, and combine the characteristics of depression patients with negative emotional tendencies and suicidal tendencies in actions to construct three major emotional dictionaries, behavioral keyword dictionaries, and behavioral dictionaries. Suicide-related dictionaries are gathered into a comprehensive dictionary library for the depression campus social platform, which makes up for the lack of dictionaries in this field.

2. Feature selection: Depression blogs have their own features, not only in the text itself, but also in some extended features that are closely related to
depression. This paper fully integrates semantic features, dictionary features, and extended features to ensure the optimal recognition rate of depression.

(3) Algorithm improvement: Based on the three selected feature modules, support vector machines and convolutional neural networks were used for experiments, but these two algorithms were not compatible with all features at the same time. Therefore, this paper proposes an improved two-input convolutional neural network, which realizes multifeature fusion input, and further improves the experimental effect.

Through comprehensive analysis and experiments, this paper still has some shortcomings, mainly involving the following two aspects:

(1) Although this paper extracts relevant features more comprehensively, the influence of each feature is different; for example, dictionary features, time in extended features, and other features can better reflect whether the user suffers from depression. Therefore, it is very necessary to weight each feature, which is also the next research direction.

(2) The convolutional neural network used in this paper is suitable for short text classification. Although most of the campus social platforms are short texts, there are also users who use long sentences to describe their situation comprehensively. At this time, it is easy to cause the mining of these information. Therefore, follow-up research can select different deep learning algorithms according to the length of the blog. If it is a long blog we can use a recurrent neural network.

Data Availability
The dataset can be accessed upon request.

Conflicts of Interest
The authors declare that there are no conflicts of interest.

Acknowledgments
This research was supported by Shandong art and science key project research on innovation and development of community dance in Shandong province (no. L2021Q0708033).

References