Depression Tendency Screening Use Text Based Emotional Analysis Technique

Chujun Yang\textsuperscript{1}, Xiangwei Lai\textsuperscript{2*}, Zhe Hu, Yanni Liu, Peng Shen

Factual of computer and information science, Southwest University, Chongqing 400715, China
lxw@swu.edu.cn

Abstract. This paper proposes a text recognition model for semantic analysis of the interview records related to depression, which can effectively identify whether the interviewee is a patient with depression tendency. It mainly consists of two components: 1) The framework of Support Vector Machine (SVM) for Classification of depression related questions; 2) the framework of Doc2vec and Text Convolutional Neural Network (TextCNN) for classification of whether the interviewee has a tendency to depression. Finally, the results obtained by the two classification methods are combined to establish a text classification model that is easy to analyze the tendency of depression.

1. Introduction
Depression is not only the most common mental disorder, but also a major health problem. The World Health Organization recognized depression in 2002 as one of the most serious diseases in the world\cite{1}.

Depression usually makes a person more likely to commit suicide. It is estimated that up to 50% of suicides have been clinically diagnosed as patients with depression. Given the relationship between depression and suicide, effective diagnosis and treatment of depression have become the key to suicide prevention\cite{2}.

At present, there are still some difficulties in the diagnosis of depression. The biggest difficulty is the quantitative problem of the degree of depression in patients\cite{3}. Therefore, in response to depression, there are no specific diagnostic criteria recognized by the medical profession, and still rely mainly on scales, questionnaires and other methods for diagnosis. The methods Commonly used are: Hamilton Depression Scale (HAMD), Beck Depression Scale (BDI), Depression Rapid Self-Assessment Scale (QIDS), and Patient Health Questionnaire (PHQ-8).

With the development of information technology, more and more methods and techniques are used for depression research, such as COVAREP, OpenFace, DepAudioNet and other software. By collecting human facial expression information, these software use psychology to study human facial expressions, transform the collected information into data forms, and then combine the collected large amounts of data with neural networks, machine learning and other algorithms to make depression research more informative.

At present, the various analysis methods for depression used at home and abroad are shown in Table.1
Table 1 Research status

| Research status | Researcher | Signal used | Analytical method | Result |
|-----------------|------------|-------------|-------------------|--------|
| 1               | Le Yang, Dongmei Jiang | AU facial features, PHQ-8 | Decision tree | They established a decision tree to classify whether people with PTSD had depression[4]. |
| 2               | Gilles Degotteux, John Kane | audio frequency | Sinusoidal model, spectral envelope | A new open source speech processing algorithm library COVA-REP is proposed[5]. |
| 3               | Xingchen Ma, Hongyu Yang, Qiang Chen, Di Huang, | audio frequency | Convolutional Neural Network, Long Short-Term Memory | A deep neural network, DepAudioNet, is proposed, to provide a more comprehensive audio representation[6]. |
| 4               | Yan Liu, Youjun Li, Meng Chen | Resting state EEG signal | Intrinsic mode decomposition algorithm, support vector machine | The SVM algorithm was used to classify the EEG features of depression patients and normal controls, and the classification accuracy rate reached 93.3%. |
| 5               | Zhen Zhou, Hong Wang, Chunhong Liu | Brain magnetic resonance, Left and right frontal white matter | Raw Data Analysis (RDA), Principal Component Analysis (PCA) | They found there are differences in the texture characteristics of the white matter region of the prefrontal cortex in the brain magnetic resonance images of patients with depression and healthy people[7]. |
| 6               | Tadas Baltrusaitis, Peter Robinson, Louis-Philippe Morency | facial features | Conditional Local Neural Network (CLNF), Convolutional Neural Network (CNN) | OpenFace, for facial marker detection, head pose estimation, facial motion unit recognition and eye gaze estimation for computer vision, emotional computing, etc.[8]. |
| 7               | Harry Hollien | Speaking fundamental frequency (SFF) | Psychological Stress Evaluator (PSE) | There are five characteristics of speech patterns in patients with depression: reduced speech intensity, reduced pitch range, slow speech rate, reduced intonation, a lack of desire to speak[9]. |

Natural Language Processing (NLP) is often used to deal with sentiment analysis problems. In the classification method that extracts the content and sentiment of text, natural language processing is used to quantify the language and perform classification operations.

In 2013, Google proposed the Word2vec[10] method. Through the training of the Word2vec algorithm, the text content can be simplified into an n-dimensional word vector. The similarity between vectors is used to indicate the semantic similarity of the text. Therefore, it can be used to find synonyms, part of speech analysis and synonym clustering. Context refers to a specific word phrase in front of the word. First, the partial probability value is calculated to establish a neural network model, and then the established neural network model is used to predict other probability values, and the vector corresponding to the word can be obtained. Word2vec’s training process is divided into "CBOW" and "Skip-gram" models according to different language models. The "CBOW" model, also known as the continuous word bag model, is characterized by a known context that outputs a prediction of the current word. The "Skip-gram" model predicts context words from current words.

Based on word2vec, Tomas Mikolov proposed the Doc2vec[11] method in 2014 to train high quality text.

When processing text paragraphs, the input to the machine learning algorithm requires a fixed length vector, and the most commonly used fixed length text vector representation is bag-of-words. But it has two main drawbacks: First, the word bag model ignores the word order. If two different sentences are composed of different orders of the same word, in the bag model, the two sentences will be defined as the same expression; the bag model ignores the syntax, making the distance between synonyms different from the actual situation, that is, the degree of synonym between words becomes blurred.

Doc2vec, also known as Paragraph Vector, is an unsupervised learning. For the defects of the word bag model, the Doc2vec method can train sentences with different lengths without fixing the sentence
length, and the predicted words are meaningful, which overcomes the lack of semantic defects of the word bag model. Like Word2vec, Doc2vec also has two models: the PV-DM model corresponds to CBOW; the other PV-DBOW corresponds to the skip-gram model. The difference between Doc2vec and Word2vec is that Doc2vec adds a new sentence vector Paragraph Vector to identify the statement.

The research in this paper is mainly divided into two experimental steps. The first step is to consult the relevant depression disease data, and select six questions that are most relevant to depression from the interview. The questions are input into a support vector machine (SVM) algorithm for classification to determine whether the interviewee has depression tendency. In the second step, an improved text Convolutional Neural Network (TextCNN) is input for text analysis to learn whether the interviewee suffers from Depression. Finally, the two classification results are combined to obtain a text analysis model for analyzing the tendency of depression.

2. SVM Depression Emotional Tendency Analysis Model

2.1. Data preprocessing

The database used in this paper, the Disease Analysis Interview Corpus (DAIC), from the University of Southern California, is designed to analyze whether patients with poor mental health such as anxiety, and post-traumatic stress disorder (PTSD) have depression. These interviews were conducted by human-controlled computer virtual robot Ellie to reduce the psychological burden on the interviewees. The data collected included recorded text and a PHQ-8 questionnaire. A PHQ-8 questionnaire provided in the database is used to determine whether the subject is a depressed patient. There were 189 interactions and 189 interviewees, the time is between 7-33 minutes (average 16 minutes). The textual records of 189 interviewees are shown in Table.2 below.

| Start_time | Stop_time | speaker | value |
|------------|-----------|---------|-------|
| 36.588     | 39.668    | Ellie   | Hi I’m ellie thanks for coming in today |
| 39.888     | 43.378    | Ellie   | I was created to talk to people in a safe and secure environment |
| ...        | ...       | ...     | ... |
| 62.328     | 63.178    | Participant | good |
| 65.858     | 67.528    | Ellie   | Where are you from originally |
| 68.978     | 70.288    | Participant | Atlanta georgia |

(1) Extracting six key questions as six attributes are shown in Table.3 below:

| question |
|----------|
| 1 how are you doing today? |
| 2 how are you at controlling your temper? |
| 3 how do you like your living situation? |
| 4 how easy is it for you to get a good night's sleep? |
| 5 have you been diagnosed with depression? |
| 6 have you ever been diagnosed with p_t_s_d? |

(2) Label the data.

The Disease Analysis Interview Corpus (DAIC) contains a PHQ-8 questionnaire. For each interviewee, there is a questionnaire that is internationally recognized as one of the diagnostic methods used to determine whether a patient has depression. The results of the quantitative quantification of the questionnaire are directly given in the database, as shown in Table.4 below:

| Participant_ID | PHQ8_Binary | PHQ8_Score | Gender |
|----------------|-------------|------------|--------|
| 302            | 0           | 4          | 1      |
| 307            | 0           | 4          | 0      |
Participant_ID is the number of the interviewee, corresponding to the text record; PHQ8_Binary indicates whether it is depression, 1 is depression, 0 means no depression; PHQ8_Score is the score of PHQ8 quantization table, corresponding to PHQ8_Binary, in medicine PHQ8_Score>10 corresponds to PHQ8_Binary=1, diagnosed as suffering from depression, the higher the PHQ8_Score score, the more severe the depression, PHQ8_Score<10 corresponds to PHQ8_Binary=0, it can not be diagnosed as suffering from depression according to this table, which is considered here The interviewee did not suffer from depression; Gender did not make a distinction in the experiment. Therefore, the PHQ8_Binary in this table corresponds to the interviewee, and the interviewee is tagged with 1 or 0, 1 is suffering from depression, and 0 is not suffering from depression.

2.2. Classification of depression tendency using SVM

The sentiment analysis method is determined by using the positive and negative attributes of the words. Each word in the sentence has a score, the optimistic word score is positive, and the pessimistic word is negative.

The textblob algorithm is used for each interviewee's answer corresponding to the key question extracted from the data pre-processing, so that the emotional value in the answer is quantified. The result after using the textblob algorithm is between -1 and +1, the positive score is positive emotion, and the negative score is negative emotion. The closer the positive score is to 1, the more positive the emotion, the closer the negative score is to -1, the more negative the emotion. Since the positivity of language emotion has a certain relationship with the state of mind, it is speculated here whether or not there is a tendency to have depression based on the negative or positive degree of the answer. The answers to the six key questions extracted are shown in Table.5 below, and are recorded as a1, a2, a3, a4, a5, and a6. If there is no corresponding key question in the interview, it is recorded as 0.

| Participant_ID | a1  | a2   | a3   | a4   | a5   | a6   | PHQ8_Binary |
|----------------|-----|------|------|------|------|------|-------------|
| 303            | 0.5 | 0.27 | 0    | 0.1  | 1    | 0    | 0           |
| 304            | 0.7 | 0    | 0.5  | 0.475| 1.0  | 1.0  | 0           |
| 305            | 0   | 0    | 0.8  | -0.22| 0    | 0    | 0           |
| 310            | 0.5 | 0.417| -0.41| -0.467| -0.26| 0    | 1           |
| 312            | 0.416| 0.475| 0    | -0.292| 1.0  | 1.0  | 0           |

The training was focused on the quality characteristics of these six conversations and whether it was the result of depression (1 for depression, 0 for non-depressed patients). The SVM algorithm was used to get a classification for patients with depression. The classification accuracy is 70%. See Figure.1 below. In this section, the model is more sensitive to the patient with PTSD as well as the patient with severe insomnia. If the emotional tendency of the interviewee's answer is less obvious, the classification result is less than ideal.
3. TextCNN Analysis of Depression Tendency

3.1. Cleaning data

Due to the small number of data samples, 189 interviewed interview data were split, and each text record was split into ten, as ten interviewee data, thus stretching 189 samples to 1890, of which 1590 Training set samples, 300 test set samples. The processing results are shown in Table.6 below.

Table.6 Split sample sets

| Participant ID | Text                                | PHQ8 | Binary |
|----------------|-------------------------------------|------|--------|
| 303            | ['hi', 'i'm', 'ellie', 'thanks', 'coming', 'today',...] | 0    | 0      |
| 304            | ['things', 'make', 'really', 'mad', 'dishonesty', 'maybe', 'um', 'give',...] | 0    | 0      |
| 304            | ['roommates', 'yes', 'tell', 'um', 'they're', 'they're', 'friendly',...] | 0    | 0      |
| 320            | ['well', 'longer', 'living', 'i'm', 'sorry', 'hear', 'mhmm', 'noticed', 'changes',...] | 1    | 1      |
| 322            | ['mm', 'know', 'that's', 'reason', 'sounds', 'really', 'hard', 'anything',...] | 0    | 0      |

The next works are replacing all uppercase letters in the text with lowercase letters, removing extra punctuation marks such as colon quotation marks, and removing stop words in each sentence. Among them, stop words refer to some function words such as "the", "is", "at", "on". Compared with other words, functional words have no practical meaning. They mainly help to describe nouns and expression concepts in texts. However, due to the high repetition rate of stop words and large space overhead, they have certain influence on the classification results. Therefore, in order to save space and improve the classification accuracy of sentiment analysis, the stop word vocabulary is directly used here to filter out the noise in the word segmentation structure which may affect the classification result.

The cleaning results are shown in Table.7 below.

Table.7 Samples after cleaning

| Participant ID | Text                                                |
|----------------|-----------------------------------------------------|
| 303            | hi im ellie thanks coming today created talk people safe secure environment... |
| 304            | what study school early childhood education nice still working right would love... |
| 312            | laughter give example um four sisters one brother moms side dad six brothers... |
| 314            | time makes easier thats good feel lot cause theres lot things make think daily... |
| 320            | see um something new cool change cha change scenery yeah advice would give ten... |
| ……             | ……                                                  |
3.2. Establishing the Doc2vec model
In the DAIC-WOZ disease interview database, training sets and data sets have been divided. Since Doc2vec is unsupervised learning, the Doc2vec model is directly trained on the entire pre-processed database, and two models of PV-DM and PV-DBOW are built according to the characteristics of Doc2vec. The result of the training is that the interview content has been vectorized and contains the semantics of the text itself. Then it can use Doc2vec's own function to test, and input a sentence in the sample to predict the semantics of the sample. After the vectorized training set, the SGDClassifier classifier is selected to train the results of the training set to classify the test set. The final classification accuracy of the obtained test set PV-DM model and PV-DBOW model was 68%. The SGDClassifier regression graph is shown in Figure.2

![Figure 2: SGDClassifier](image)

3.3. Establishing a TextCNN model
TextCNN [12] is a method proposed by Yoon Kim in 2014 to classify text based on convolutional neural networks. According to the method in the original paper, here is a slight improvement, combined with the word vector trained by the Doc2vec method above, as the input of the convolutional neural network. The improved TextCNN structure is shown in Figure.3

![Figure 3: TextCNN](image)

Convolutional neural networks are generally used for image analysis. When applied to textual language analysis, a sentence was interpreted as a two-dimensional map of n×k. Where n is the length of the sentence, that is, there are n words in this sentence, and k is the dimension of the word vector.

The first layer is the input layer, which inputs the n×k matrix of each sentence, n = 200, k = 200; the type of this matrix can be static or dynamic.
The second layer is the convolutional layer. Since the minimum granularity to be analyzed in the text is a single word, the semantic integrity can be guaranteed. Therefore, the convolution kernel must be convolved without dividing the words, and the convolution kernel is fixed. The width is the same as the dimension of the word vector, defined as 200, which means that the position where each window slides is a complete word. Here, the widths of 2, 3, and 4 words are selected as the convolution window sizes, and 64 convolution kernels are selected in each window. A convolution kernel convolves a sentence to get a vector.

The third layer is Max Pooling. The Max Pooling operation in CNN has the following advantages: Firstly, it can ensure that the position change of the feature has no effect on the feature extraction, because no matter where the strong feature appears, just select the maximum value. So it can be directly extracted without considering its location; Secondly, Max Pooling can reduce the number of model parameters, which is beneficial to reduce the model over-fitting problem. For NLP problems, Max Pooling has the additional advantage of arranging variable-length inputs into fixed-length inputs. After the pooling operation, one feature value is fixedly selected, so that the input of the subsequent fully connected layer is not affected. However, positional invariance is a very good advantage for image processing, but it is not a good thing for natural language processing.

In the original TextCNN method, a single max pooling is used for global pooling, that is, the feature vector output for each convolution kernel in the convolutional layer, and the maximum value is taken over the entire sentence length to obtain a scalar. The result of this processing will cause only the most important part of the sentence that is the closest to the convolution kernel to be output after the pooling process, and ignore the word order position, thus losing the structure information. The location information of the feature is very important for the text. For example, the subject-predicate usually has its own fixed order, and these positional orders are very important for the classification task. On the other hand, sometimes some strong features appear multiple times, and the more times the feature appears, the stronger the feature. However, Max Pooling only retains a maximum value, so it is impossible to highlight the number of times the feature is repeated, and the strength information of the same feature is lost.

There is an improvement for this problem, that is using k-Max Pooling to make some optimizations, given a k value, and a sequence p (p>=k), k-Max Pooling selects the first k in the sequence p. These maximum values retain the order of the original sequences, as shown in Figure 4 below where k is taken as 2. The original Max Pooling only retained a maximum value and one feature information. K Max Pooling preserves the first k maximum values, and keeps the order in which these values appear, that is, arranges the k maximum values according to the position order in the text, which is actually a subsequence of the original sequence, as much as possible. In addition, since the first k maximum values are selected, k-Max Pooling can also express the same type of feature multiple occurrences. If the same value is present in the first k maximum values, then the feature values will be selected. The number of repetitions can be calculated. K is taken as 3 in the experiment.

The fourth layer is connected to a Softmax layer by means of full connection, and the feature results obtained in the previous layer are input into the Softmax layer for classification.
Finally, the Doc2vec model trained paragraph vector is input into the above TextCNN model, and the accuracy of the classification result of the test set with depression tendency is 72%. In this section, the Doc2vec method is to learn the global text, it is easier to judge the direct nuances of the interviewee, but because there are some questions and answers unrelated to depression in the interview content, these noises interfere with the classification results.

4. Conclusion

Finally, the two results of SVM and TextCNN algorithm were multiplied. The final classification of patients with depression tendency was 75%. The text analysis method provided in this paper is mainly aimed at depression-related texts, which is more targeted than the general text analysis model. Firstly, the Support Vector Machine algorithm is used to quantify the local text, and then the Convolutional neural Network is used to learn vectorized sentiment analysis on the whole text, which makes the analysis result more comprehensive. However, there are still some flaws in the method, so that the classification accuracy is not high enough. In response to this result, it is mainly due to:

(1) The SVM algorithm is used to analyze the sentiment tendency of local text. If the interviewee answers the questions selected in the text more neutrally than emotionally, it is difficult to tell if he has depression, so this method is more subjective, and the data sample is too small, resulting in a final classification accuracy that is not high enough;

(2) The TextCNN method is a semantic analysis of the global text, so the content is more complete, and the results of sentiment analysis are more objective. However, since all the text content is not necessarily all related to depression, so the text has some noise. In addition, in order to expand the training sample, the final training data set is obtained by splitting the original data set, which will also have an impact on the classification result.

At present, the difficulties in this research are as follows: (1) The database model is small, so the amount of data used for learning and testing is not large enough; (2) Only through the text model analysis, the study of depression tendency is slightly thin. In response to these problems, follow-up research will also make the following improvements: (1) Compress the text to remove the content that has nothing to do with depression, and improve the Doc2vec model, strengthen the study of long sentences and long paragraphs; (2) use more features subsequently, such as audio features, facial features, combined with text features to study together.

Acknowledgments

This work was supported by "Fundamental Research Funds for the Central Universities" No. XDJK2018C039.

References

[1] Scherer, S., Stratou, G., Lucas, G., Mahmoud, M., Boberg, J., Gratch, J., ... Morency, L. P. (2014). Automatic audiovisual behavior descriptors for psychological disorder analysis. Image and Vision Computing, 32(10), 648–658.2014.06.001

[2] Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B. W., Berry, J. T., & Mokdad, A. H. (2009). The PHQ-8 as a measure of current depression in the general population. Journal of Affective Disorders, 114(1–3), 163–173. https://doi.org/10.1016/j.jad.2008.06.026

[3] Scherer, S., Lucas, G. M., Gratch, J., Rizzo, A., & Morency, L. P. (2016). Self-Reported Symptoms of Depression and PTSD Are Associated with Reduced Vowel Space in Screening Interviews. IEEE Transactions on Affective Computing, 7(1), 59–73. https://doi.org/10.1109/TAFFC.2015.2440264

[4] Yang, L., Jiang, D., He, L., Pei, E., Oveneke, M. C., & Sahli, H. (2016). Decision Tree Based Depression Classification from Audio Video and Language Information. Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge - AVEC '16, 89–96. https://doi.org/10.1145/2988257.2988269

[5] Degottex G, Kane J, Drugman T, et al. COVAREP—A collaborative voice analysis repository for
speech technologies[C]//Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. IEEE, 2014: 960-964.

[6] Ma, X., Yang, H., Chen, Q., Huang, D., & Wang, Y. (2016). DepAudioNet. Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge - AVEC ’16, 35–42. https://doi.org/10.1145/2988257.2988267

[7] Zhen Z., Hong W., Chunhong L. Changes of texture features of brain magnetic resonance images in bipolar depression patients[J]. Beijing Biomedical Engineering, 2017.

[8] Baltrušaitis, T., & Robinson, P. (n.d.). OpenFace: an open source facial behavior analysis toolkit.

[9] Hollien H. Vocal indicators of psychological stress[J]. Annals of the New York Academy of Sciences, 1980, 347 (1): 47-72.

[10] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]//Advances in neural information processing systems. 2013: 3111-3119.

[11] Le Q, Mikolov T. Distributed representations of sentences and documents[C]//International Conference on Machine Learning. 2014: 1188-1196.

[12] Kim Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv:1408.5882, 2014.