Speech-Based Automatic Recognition Technology for Major Depression Disorder

Zhixin Yang(✉), Hualiang Li, Li Li, Kai Zhang, Chaolin Xiong, and Yuzhong Liu

Electric Power Research Institute of Guangdong Power Grid Corporation, No. 8 Shuijungang, Dongfengdong Road, Yuexiu District, Guangzhou 510080, China
67677483@qq.com

Abstract. Depression is one of the common mental illnesses nowadays. It can greatly harm the physical and mental health of patients and cause huge losses to individuals, families and society. Because of the lack of hardware and social prejudice against depression, there are a large number of misdiagnosis and missed diagnosis in hospitals. It is necessary to find an objective and efficient way to help the identification of depression. Previous studies have demonstrated the potential value of speech in this area. The model based on speech can distinguish patients from normal people to a great extent. On this basis, we hope to further predict the severity of depression through speech. In this paper, a total of 240 subjects were recruited to participate in the experiment. Their depression scores were measured using the PHQ9 scale, and their corresponding speech data were recorded under the self-introduction situation. Then, the effective voice features were extracted and the PCA was conducted for feature dimensionality reduction. Finally, utilizing several classical machine learning method, the depression degree classification models were constructed. This study is an attempt of the interdisciplinary study of psychology and computer science. It is hoped that it will provide new ideas for the related work of mental health monitoring.

Keywords: Depression · Speech · Machine learning · Classification · Prediction

1 Introduction

With the development of science and technology and the social system, material life is gradually guaranteed. People begin to pay attention to mental health, and the common mental illness such as depression has also received more and more attention. Depression affects people's normal study, work and life to varying degrees, and even severely leads to suicide. Although there are reliable medical treatments, many factors make it impossible for people with depression to receive effective treatment at the first time. These factors include lack of resources, lack of well-trained medical staff and discrimination against mental patients in society. At the same time, the inability to accurately diagnose depression also affects whether patients are effectively treated. If

D. Milošević et al. (Eds.): HCC 2019, LNCS 11956, pp. 546–553, 2019.
https://doi.org/10.1007/978-3-030-37429-7_55

we can develop efficient, accurate and low-cost diagnostic methods, it can greatly help patients with depression to receive effective treatment.

Depression is a serious mental illness centered on depressive emotions and accompanied by symptoms such as low self-esteem, low vitality, and low interest in daily activities. Depression in the narrow sense refers to Major Depression Disorder (MDD), and the depression mentioned in this article, unless otherwise specified, refers to major depression. According to data from the World Health Organization (WHO) in 2017, more than 300 million people worldwide suffer from depression. Depression can lead to a very high rate of death. Reports indicate that th e suicide rate of depressed patients is more than 20 times that of the general population. At the same time, studies have indicated that depression increases the risk of cardiovascular disease and stroke. In addition, depression can badly affect the life, work and study of patients. Patients may cause work efficiency and economic loss due to work-related problems. There are reports that the economic loss in the United States due to depression in a year is as high as 36.6 billion dollars. The World Health Organization (WHO) believes that depression will become the world's major disease in 2030. Depression is a heavy burden for individuals, families, and society.

Although depression is a serious hazard, it can be effectively improved after proper medical treatment, psychotherapy or physical therapy. However, less than one-half of the patients receive effective treatment (according to the country's economic and cultural level, many regions are even less than one tenth). There are many factors that influence effective treatment, such as lack of resources and medical conditions, and lack of trained medical staff. Despite lack of hardware condition, in order to effectively treat patients with depression, we also need to ensure that the symptoms are correctly diagnosed. However, the misdiagnosis of depression is very serious. Due to the influence of traditional culture in China, the society has discrimination and misunderstanding of mental illness such as depression. People are afraid of medical treatment for mental illness, and they are not willing to go to the psychiatric department for treatment. Because of people's lack of sufficient knowledge in mental illness, they visit different wrong departments such as neurology, gastroenterology, cardiology. Besides, patients may not be able to express their mental discomfort through self-report, resulting in a high rate of misdiagnosis and missed diagnosis. Lilan [1] found that 54.2% of misdiagnosis and missed diagnosis was caused by the misdiagnosis of the first-time doctor. The doctors in general hospital had insufficient understanding of depression, only paid attention to the patient's physical symptoms and ignored their mental symptoms, therefore failed to detect those mental symptoms which patients could not express.

Most of the current diagnosis of depression is judged by a specialized psychiatrist with the diagnostic and statistical manual of mental disorders (DSM) [2], and the Hamilton Depression Scale (HAMD), Beck Depression Inventory (BDI) and other scales are also used to assist judgment. The DSM is published by the American Psychiatric Association and is the most commonly used instruction manual for diagnosing mental illness. Since its promulgation, it has been widely concerned. The American Psychiatric Association has continued to collect and sort out opinions to improve its clinical utility. However, because DSM can only be measured in a psychiatric clinic, its actual use is very limited. HAMD, which is commonly used in

clinical practice, requires a professional assessor to communicate with the patient or to observe the patient's condition. It takes about half an hour for a single assessment. The self-assessment scale has the advantage of being able to use at any time and place without the need of professional medical staff, such as BDI, PHQ (Patient Health Questionnaire) and other self-rating scales. As a more simplified version of the BDI scale, PHQ-9 is often used in patients' self-assessment to judge the recovery status and the severity of depression. It is a widely recognized self-rating depression scale.

If we can automatically identify depression through the patient's ecological behaviors such as voice, gait, facial expressions, etc., we can give the doctors some advice without the patient's self-report or psychiatric expert's diagnosis, suggesting whether they should pay attention to the patient's mental state. In order to establish a model to automatically detect and recognize depression, people choose speech as the material for feature extraction. Speech is an easily accessible, non-invasive message, and slight psychological or physiological changes can result in significant changes in speech characteristics. Depression is thought to cause changes in neurophysiological and neurocognitive effects that affect human behavior, language, and cognitive function, and thus affect human speech. The voice of patients with depression is described as low-pitched, slow, long pauses, lack of stagnation, and dull. The speech characteristics of patients with depression have changed, which may be related to cognition and memory. The cognitive impairment caused by depression affects the working memory of individuals, and working memory is closely related to language planning. The effect of depression on speech indicates that the feature extraction of speech can be used to predict depression. The speech feature itself has nothing to do with the content of the speech, and cannot be concealed by the speaker. It can reflect the true psychological state of the speaker and is more objective and reliable.

Use of speech to predict depression has yielded preliminary results [3–8]. Cohn et al. [9] used the cross-validation to distinguish between depression and healthy people, the prediction accuracy rate reached 79%, suggesting the correlation between speech characteristics and depression. Cummins et al. [10] used the readings of 47 subjects under neutral and depressed emotions as the original input data and extracted the features for classification training. They found that the detailed features of the speech can be well adapted to the classification task, and the classification accuracy could reach 80%. Williamson et al. [11] used the data of the 6th International Audio/Video Emotion Challenge (AVEC) to classify depression and healthy people and score of Patient Health Questionnaire (PHQ). It is predicted that the average F1 score is 0.70.

The current research on speech and depression makes people notice the change of speech in patients with depression. Self-rating depression scale can reflect the degree of depression of the person filling the form with numerical value, and it is also recognized by the society for its convenience. Suppose we can combine the two, extract the characteristics of speech, prove the correlation between speech and depression scores through machine learning and mathematical statistics, and establish a model to predict the depression scale. The score is used to diagnose whether or not the participant is suffering from depression and predict the degree of depression on the basis of the score. This can provide a supplementary means for the diagnosis of depression in reality.

In this paper, we use the PHQ-9 scale as a basis for quantifying the severity of depression. PHQ-9 is often used as a self-assessment of patients' depression status to judge the therapeutic effect and degree of depression, and the measures that should be taken afterward. The score of the scale explains the severity of depression, and whether it can be combined with speech to predict if a participant has depression is also a problem we want to explore.

2 Materials and Methods

2.1 PHQ-9 Self-rating Depression Scale

PHQ is called Patient Health Questionnaire and is often used as a diagnosis of mental health diseases, such as depression, anxiety, alcoholism, binge eating disorder and other common mental illnesses. The PHQ scale initially contained 11 multiple-choice questions, which were supplemented with continuous improvement. PHQ-9 specifically refers to 9 scales for depressive symptoms, which are measured by the past two weeks, from loss of pleasure, depression, sleep disorders, lack of energy, eating disorders, low self-evaluation, difficulty in concentration, slowness and negative attitudes. The scores of the options are: 0 points not at all, 1 point for several days, 2 points for more than half of the day, 3 points for almost every day, and a total score of 0–27 points. The specific evaluation indicators are listed in Table 1. In the experiment, each participant was asked to fill in the questionnaire, and the score was used as a criterion for judging whether or not depression occurred.

Table 1. PHQ-9 scoring rules

Score	The degree of depression
0–4	No depression
5–9	May have mild depression and suggest to consult a psychiatrist
10–14	May have moderate depression and suggest to consult a psychiatrist
15–19	May have moderate to severe depression and suggest to consult a psychiatrist
20–27	May have severe depression and should consult a psychiatrist

2.2 Speech Collection

First, brief the participants on the purpose and steps of the experiment, let participants sign the informed consent form, register the information of the participants, give the participants an outline of the speech, and leave a preparation time of 5 min. The outline of the speech is: Please introduce yourself and introduce your hometown in detail; please introduce your major in detail, your research work during your study; please tell us about your future planning and what kind of work you want. Participants were recorded in a state where they were not intentionally induced to have positive or negative emotions. During the process, we would not interfere or remind. The duration of the recording was about 1 min each. The participants then used the pad to fill in the questionnaire.

Data collection was carried out in two batches. In August 2018, 88 speech and questionnaire data were collected, of which 88 were valid data; in January 2019, 152 speech and questionnaire data were collected, one of which appeared error, and the actual effective data was 151 copies. A total of 239 data were collected, in which, there are 130 women and 109 men.

2.3 Data Processing

OpenSMILE is an open source toolkit for feature extraction in signal processing and machine learning. It mainly extracts the characteristics of the audio signal. In this study, we used the emotion recognition feature extraction configuration file—emobase.conf that came with the software to extracted 989 acoustic features, including the following low-level descriptors (LLD): intensity, loudness, 12 MFCC, pitch (F0), voiced probability, F0 envelope, 8 LSF (line spectral frequency), zero-crossing rate, and Delta regression coefficients for these LLDs.

The openSMILE used in the experiment is version 2.3.0. The extracted features are processed differently, for example, directly using raw data, using normalized data, and using PCA (Principal Component Analysis) to reduce the data.

2.4 Depression Classification Model

In the experiment, the WEKA software was used to train the extracted features using different machine learning methods, and the model was extracted. Then we used 10-fold cross-validation method to test the accuracy of the model, doing the regression and classification analysis respectively.

In the regression analysis, differently processed features were used to predict the scores of PHQ-9. During each training process, we respectively used Random Forest, Linear Regression, Sequential Minimal Optimization (SMO), and Random Tree, Simple Linear Regression and other algorithms to train the data.

According to the score of the PHQ-9, the person filling the form can be divided into five categories according to the severity of depression (Table 1). According to the research needs, we focus on the classification and identification of major depression disorder, so we divide the sample into two groups with 14. The score is greater than 14 means able to have severe depression, it is recommended to consult a doctor, and less than 14 means depression tendency is not serious or no depression. The processed data is trained using different algorithms, such as Bayes Net, Support Vector Machine (SVM), Support Vector Machine Minimum Optimization Algorithm (SMO), Random Tree (Random Tree), decision tree J48, etc.

In addition, because the speaker's speech may be affected by the content, the extracted features may be more related to the content. We try to use the deep learning method to learn the feature to reduce the influence of the speech content on the prediction result. Here, the BP (Back Propagation) algorithm in the neural network algorithm is used, that is, a multi-layer feedforward network. BP network for classification and regression analysis can be implemented in weka. After selecting the MultilayerPerceptron algorithm in the classifier in weka explorer, we can set the hidden

layer, learning rate, regularization and iteration number and other parameters in the graphical interface.

3 Results

Different classification algorithms are used for binary classification prediction. The ten-fold cross-validation is used to test the performance of classification model. The extracted data are normalized and standardized separately, and different algorithms are used for classification prediction. The results of depression classification on speech features without feature dimensionality reduction are shown in Table 2.

Table 2. The results of depression classification on speech features without feature dimensionality reduction

Algorithm	Correctly classified rate
BayesNet	72.38%
NaiveBayes	61.51%
Logistic	60.67%
SMO	62.34%
J48	69.46%
RandomTree	64.44%

PCA is performed on the normalized data, and then different algorithms are used for prediction. The results are shown in Table 3.

Table 3. The results of depression classification on speech features with feature dimensionality reduction

Algorithm	Correctly classified rate
BayesNet	77.82%
NaiveBayes	66.53%
Logistic	65.69%
J48	71.13%
RandomTree	68.20%

The BP network is used to learn the normalized data. The learning rate is set to 0.3.

The default regularization attribute is used to improve the utility of the network. The batch size is 100, and the number of training iterations is based on the actual number. The speed and effect of the operation are determined, and the initial setting is 500 to observe the actual learning effect. Set different hidden layers, where 0 means no hidden layer, and numbers indicate the number of nodes in each layer. The results are shown in Table 4.

Table 4. The results of depression classification on speech features using BP network

Hiden layers	Correctly classified rate
0	71.97%
2	70.71%
20	66.53%
50,4	64.02%

4 Discussion

The emobase.conf configuration file can extract 989 features related to emotion recognition. These features may be affected by the speaker's content and cannot reflect the acoustic characteristics completely. Too many features may cause a lot of noise without discrimination, which increases the difficulty of calculation and affects the actual training process. When using PCA to reduce dimension, the redundant features were removed and the prediction accuracy were improved. And the Bayesian network is better with 77.82% correctly classified rate.

There are still many shortcomings in this study. The first problems encountered when classifying were the sampling bias and uneven distribution. Due to the limitation of experimental conditions, it is difficult to recruit too many subjects with severe depressive symptoms. Most of the 239 subjects are normal (scores less than 5), while those who may have severe depression (with scores greater than 14) are less than one-twentieth.

5 Conclusion

In this study, we use a variety of machine learning methods to analyze the voice data acquisition in self-introduction context, and explored the association pattern between the voice and the severity of depression. The classification model of depression tendency were constructed and the classification accuracy is higher than 77%. It further validates the validity of the depression recognition method based on speech analysis.

Acknowledgements. The work was supported financially by the China Southern Power Grind (Grant No. GDKJXM20180673).

References

1. Huang, L.: Analysis of misdiagnosis of anxiety and depression in grass-roots hospitals Asia-pacific traditional medicine **8**(04), 207–208 (2012)
2. Kocsis, R.N.: Diagnostic and statistical manual of mental disorders: fifth edition (DSM-5). Int. J. Offender Ther. Comp. Criminol. **57**(12), 1546–15468 (2013)
3. Hashim, N.W., Wilkes, M., Salomon, R., et al.: Evaluation of voice acoustics as predictors of clinical depression scores. J. Voice **31**(2), 6 (2017)

4. Helfer, B.S., Quatieri, T.F., Williamson, J.R., et al.: Classification of depression state based on articulatory precision. In: Bimbot, F., Cerisara, C., Fougeron, C., et al. (eds.) 14th Annual Conference of the International Speech Communication Association, pp. 2171–2175 (2013)
5. Mundt, J.C., Vogel, A.P., Feltner, D.E., et al.: Vocal acoustic biomarkers of depression severity and treatment response. Biol. Psychiatry **72**(7), 580–587 (2012)
6. Pan, W., Wang, J., Liu, T., et al.: Depression recognition based on speech analysis. Chin. Sci. Bull. **63**(20), 2081–2092 (2018)
7. Scherer, S., Stratou, G., Gratch, J., et al.: Investigating voice quality as a speaker-independent indicator of depression and PTSD. In: Bimbot, F., Cerisara, C., Fougeron, C., et al. (eds.) 14th Annual Conference of the International Speech Communication Association, pp. 847–851 (2013)
8. Scherer, S., Stratou, G., Lucas, G., et al.: Automatic audiovisual behavior descriptors for psychological disorder analysis. Image Vis. Comput. **32**(10), 648–658 (2014)
9. Cohn, J.F., Kruez, T.S., Matthews, I., et al.: Detecting depression from facial actions and vocal prosody (2009)
10. Cummins, N., Epps, J., Breakspear, M., et al.: An Investigation of Depressed Speech Detection: Features and Normalization (2011)
11. Williamson, J.R., Godoy, E., Cha, M., et al.: Detecting Depression using Vocal, Facial and Semantic Communication Cues. Assoc Computing Machinery, New York (2016)