A Study on Depression Detection Using Eye Tracking

Shuai Zeng, Junhong Niu, Jing Zhu(✉), and Xiaowei Li

School of Information Science and Engineering, Lanzhou University, Lanzhou, China
{zengsh16,niujh16,zhujing,lixwei}@lzu.edu.cn

Abstract. Depression has become one of the most common mental illnesses in the past decade, affecting millions of patients and their families. However the methods of diagnosing depression almost exclusively rely on questionnaire-based interviews and clinical judgments of symptom severity, which are highly dependent on doctors' experience and makes it a labor-intensive work. Our study aims to develop an objective and convenient method to assist depression detection using the eye tracking technology. Eye movement data was collected from over 50 subjects using an emotional faces free viewing task paradigm. After data preprocessing, the highest accuracy of 76.04% was achieved by the Support Vector Machine (SVM) classifier. Results indicate that with the improvement of the classification accuracy, eye movement features hold the potential to form a feasible method for depression detection.

Keywords: Depression detection · Eye tracking · Classification

1 Introduction

Depression is a common mental disorder that already affects more than 350 million people worldwide [1]. It will not only make a bad influence on the patients but also on their families. The World Health Organization said that depression will become the second leading cause of illness by the year 2020 [2]. However, the assessment methods of diagnosing depression almost exclusively rely on the patient-reported and clinical judgments of the symptom severity [3]. Current diagnostic techniques of depression have obvious disadvantages, which are associated with the patient denial, poor sensitivity, subjective biases and inaccuracy [4]. Finding an objective, accurate and practical method for depression detection still remains a challenge.

Some diagnostic techniques of depression can achieve over 90% accuracy in depression detection, such as the EEG (Electroencephalo-graph) and sMRI (Structural Magnetic Resonance Imaging). However, they all need the professional and extremely expensive apparatuses. By contrast, eye tracking is much more accessible (even the Kinect V2 can support eye tracking and it only cost less than 200$), so it is growing in popularity amongst the researchers from different disciplines. Usability analysts, sports scientists, cognitive psychologists, reading researchers, neurophysiologists, electrical engineers, and other have a vested interest in eye tracking for different reasons [5]. Compared with EEG and sMRI the convenience and accessibility make eye tracking a

Y. Tang et al. (Eds.): HCC 2018, LNCS 11354, pp. 516–523, 2019.
https://doi.org/10.1007/978-3-030-15127-0_52

more practicable approach for depression detection in mass usage. In the future, with the development of the camera and sensor in the mobile phone we can even make an app for depression detection with eye tracking (now iPhone X can support eye tracking for some functions such as face ID). Although the classification accuracy of eye tracking for depression detection is inferior to EEG and sMRI, the generalizability and accessibility are irreplaceable advantages of eye tracking.

1.1 Related Work

Dating back to the spread of psychology's "cognitive revolution" to psychotherapy, attentional biases for emotional stimuli have been a key mechanism in theoretical accounts of affective disorders [6]. Therefore, most of the previous studies mainly focus on finding out the difference of attention bias between depressed patients and healthy controls. Peckham suggested an attention bias for dysphoric stimuli and possible neglecting of positive stimuli in depression [6] and Ms. Li also found an attention bias using a free view task paradigm [7]. Besides, as far as we know, only a few of studies pay attention to the depression detection using eye movement. Li used a free view task paradigm and KNN to got 81% accuracy for mild depression detection [8]. Sharifa used an audio-video experimental paradigm and SVM to got 75% accuracy for major depression detection [9]. The recent evidence in the mental health assessment have demonstrated that the facial appearance could be highly indicative of depressive disorder [10–12]. Over all, in this paper we used an emotional faces free viewing task paradigm to detect major depression.

1.2 Subjects

For the experimental validation, we used the real-world data, collected in a study at Lanzhou University Second Hospital in Lanzhou, China, a hospital granted with the title of Class A Grade 3 Hospital. The subjects in this study included patients who have been diagnosed with depression as well as healthy controls. Data was collected from over 28 depressed subjects who had been diagnosed with depression, and over 24 healthy controls. Data was acquired after obtaining the informed consent from the participants in accordance with the approval from the local institutional ethics committee. In this paper, to balance the gender, age, and education level, we chose 18 subjects undergoing depression and 18 healthy controls (shown in Table 1). Although we have to admit that the amount of data is small, it is a common problem in similar studies. As we continue to collect more data, the future study will be able to report on a large dataset. All participants gave informed consent before enrolment into the study, which was approved by Lanzhou University Second Hospital's ethics committee. Each participant received a reimbursement of approximately USD $16 for the participation after experiment.

1.3 Apparatus

The eye movement data was collected by an EyeLink 1000 Eye Tracker (SR Research Ltd., Mississauga, Ontario, Canada) with a sampling rate of 250 Hz. We only recorded

Table 1. Basic information for experiment subjects

Variables (Mean ± S.D.)	Depressed patients (n = 18)	Healthy controls (n = 18)	p
Gender (Males: Females)	9:9	10:8	0.747
Age (Years)	31.56 ± 8.60	31.33 ± 9.26	0.941
Education level (Years)	15.06 ± 3.04	15.50 ± 3.24	0.674
PHQ9	18.17 ± 4.00	3.06 ± 4.06	0.000*

P = p-value

the eye movement from the left eye of the subjects, because two eyes of people without eye diseases have the same movement patterns. The experimental stimuli were shown on a 17 inch liquid crystal display monitor at a resolution of 1024 × 768. The participants' eyes were kept at a distance of approximately 60 cm from the monitor and 60 cm from the eye tracker a fixed chin rest was used to keep the participants' heads steady.

1.4 Stimuli and Procedure

The stimuli we used consisted a set of images including 24 happy faces (12 males, 12 females), 24 sad faces (12 males, 12 females), 24 angry faces (12 males, 12 females), 24 astounded faces (12 males, 12 females), and 64 neutral faces (32 males, 32 females). All the pictures were selected from the NimStim Set of Facial Expressions image library [13]. All images were processed by Photoshop software and the size, gradation and resolution were made consistent (i.e. image size 190 × 250 pixels; 6.5 × 5 cm). Subjects would see four pictures in one trail and all the stimuli could be divided into four kinds (as shown in Fig. 1). The first kind included 1 emotional face expression and 3 neutral face expressions, the second kind included 2 emotional face expressions and 2 neutral face expressions, the third kind included 3 emotional face expressions and 1 neural face expression, and the fourth kind included 4 emotional face expressions. The emotion of stimuli could be happy, sad, astounded, and angry. There would not be the same emotion in one stimulus except neutral face expression because the neutral face expression was used as comparison in the whole experiment. The positions of the emotional images were counter balanced.

Firstly, a calibration would be performed to ensure that our eye tracker can catch the pupil and the eye movement data were recorded accurately (an error of below 0.5° had been achieved). Secondly, there would be 4 practice trails to make sure that subjects were familiar with the experimental procedures and reduce mistakes, the practice trails were totally the same as real trails. Each trial began with a white cross on a gray background. This cross was presented as a prompt for 1 s in the center of the screen to inform the participant that the emotional picture is to come. Participants were asked to stare at the fixation cross. Afterwards, an image with four faces was displayed for 3 s and participants were instructed to view them freely. A rest time of 2 s preceded the next trail. There were a total of 40 trails for each subject.

In one stimulus, each face will be regarded as an interest area (every stimulus has four faces, therefore there will be four interest areas) and the fixation time on each

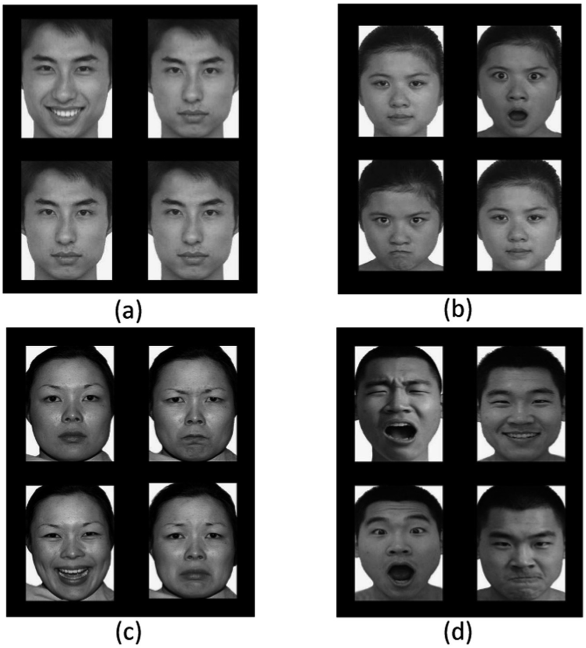

Fig. 1. Example of the facial expression stimulus materials: (a) one emotion (the emotion can be happy, sad, astounded, angry) faces and 3 neutral faces, (b) 2 emotion faces and 2 neutral faces, (c) 3 emotion faces and 1 neutral faces, (d) 4 emotion faces.

interest area was designated the total fixation time independently. Each stimulus would produce four data records, one subject would have 160 data records, for instance in Fig. 1(a) would produce one data record for happy face and three data records for neutral faces.

2 Data Processing

2.1 Data Cleaning

Because the experiment used a free view task paradigm, which would produce data record for each face in one stimulus, it might produce empty values in some specific situations. For example, if the participants blinked for a long time or blinked too frequently the eye tracker may lose the pupil of the subject and this situation will surely produce a lot of empty values in one trail, another situation is that the participants who did not focus on the interest areas can also lead to missing data.

There are different strategies to deal with tuples with empty values. If a tuple contains many attributes (more than 50% amount of attributes) with missing values, this tuple will be deserted. If a tuple just contains a few attributes with missing values then we use a measure of central tendency for the attribute (e.g., the mean or median) to fill in the missing value. For normal (symmetric) data distributions, the mean can be used, while skewed data distribution should employ the median [14]. In this paper because our data distribution was a skewed distribution, we employed the median to filling the missing value.

2.2 Data Quality Control

Outliers mixed in the experimental data will lead to inaccuracy of experimental results. Hence to obtain the correct results, outliers must be properly eliminated. The outliers are not necessarily wrong but only unrepresentative on the statistical properties. Therefore, from the representation of the statistical sense, the outliers can be removed. Pauta criterion is a kind of criterion which is often used to exclude outliers in experimental data [15]. Pauta criterion, also known as 3σ criterion, is expressed as follows,

$$|x_i - \bar{x}| > 3\sigma, (\mathrm{i} = 1, 2, \ldots, \mathrm{m}) \tag{1}$$

The x_i would be considered as an abnormal value and be rejected, where σ is the standard deviation, and $\bar{x}$ is the mean of all measurement values [16]. Detailed process is depicted in Fig. 2.

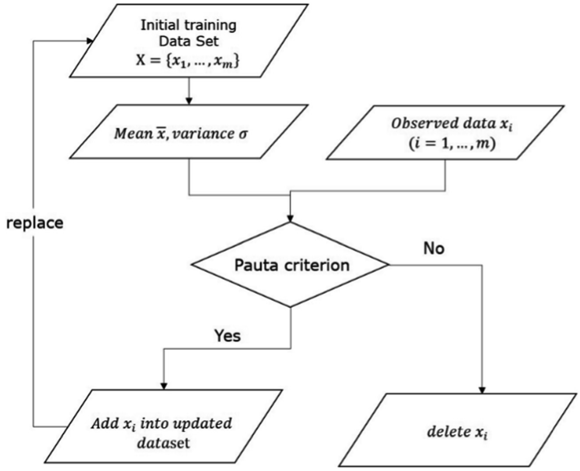

Fig. 2. The flow chart of data quality control

2.3 Data Standardization

The measurement unit can have effect on the data analysis. In general, expressing an attribute in smaller units will lead to a larger range for that attribute, and thus tend to give such an attribute greater effect or "weight". To help avoid dependence on the choice of measurement units, the data should be normalized or standardized. This involves transforming the data to fall within a smaller or common range such as [–1,1] or [0.0, 1.0]. (The terms standardize and normalize are used interchangeably in data preprocessing, although in statistics, the latter term also has other connotations.) [14].

In this paper, we used method of z-score normalization for analysis. Here we give the mathematical expression of the method. Let A be a numeric attribute with n observed values, $v_1, v_2, \ldots, v_m$. In z-score normalization (or zero-mean normalization), the values for an attribute, A, are normalized based on the mean (i.e., average) and standard deviation of A. v_i, is normalized to v_i' by computing:

$$v_i' = (v_i - \bar{A})/\sigma, (\mathrm{i} = 1, 2, \ldots, \mathrm{m}) \tag{2}$$

Where $\bar{\mathrm{A}}$ and σ_A are the mean and standard deviation, respectively, of attribute A.

2.4 Feature Used

The raw data used in this paper was generated by the EyeLink Data Viewer (SR Research Ltd., Mississauga, Ontario, Canada), which is used to display, filter, and create output reports from EyeLink 1000 EDF data files. In this study, we selected 87 attributes of eye movement data, which could directly and continuously record visual attention such as average_fixation_duration, fixation_count, fixation_duration_max, fixation_duration_max_time, fixation_duration_min, fixation_duration_min _time, pupil _size_mean, pupil_size_max, pupil_size_max_time, pupil_size_max_X, pupil_size_max_Y, pupil_size_min, pupil_size_min_time, pupil_size_min_X, pupil_size_min_Y and etc.

3 Results and Discussion

A 10-fold cross validation was used in this paper, training data and testing data were strictly separated. Data samples from one subject could not be divided into both training data and test data, lest it achieve a falsely high classification accuracy. From a large number of classifiers we chose: k-Nearest Neighbor (KNN), Support Vector Machine (SVM), Naïve Bayes, Bayes Net, Random Forest, J48, and Logistic Regression, which are often used in other cognitive researches. We also compared the results that used the feature selection algorithm and the results that directly used all features. The idea of the feature selection is to choose a subset of features that improve the performance of the classifier especially when we deal with high dimension data. In this paper, we adopted Correlation-based Feature Selection (CFS) together with the search method Best First or Greedy Stepwise, on account of selecting a subset of features which are highly correlated with the class while having low inter-correlation.

Table 2 demonstrated that the SVM classifier got the highest accuracy of 76.04% ± 3.83%, this result is in accordance with the relative researches mentioned previously. As seen in Table 2, the Random Forest, SVM and KNN classifiers are better than other classifiers, these three algorithms achieve the best classification accuracies and this phenomenon is same as the relative studies both in the eye tracking and EEG [8, 9, 15–17], in their studies the SVM, KNN and Random Forest classifiers also outperformed than other classifiers. From Table 2 we also can see that the feature selection algorithms didn't improve the accuracies remarkably, in some classifiers the results of using feature selection algorithm even worse than using all features. The statistics results of our study showed no statistically significant differences and according to our results it was difficult to demonstrate attentional bias in depression in a statistical way but this result was also same to many researchers' results [18].

Table 2. Classification results of several classifiers

Classifiers	Best first	Greedy stepwise	All features
BayesNet	69.24% ± 7.00%	71.10% ± 7.73%	70.40% ± 7.15%
Logistic	59.21% ± 8.40%	63.60% ± 7.05%	63.27% ± 7.91%
RandomForest	73.07% ± 7.75%	72.37% ± 6.25%	74.29% ± 7.68%
J48	49.09% ± 9.68%	54.23% ± 8.56%	58.55% ± 8.58%
NaiveBayes	59.02% ± 8.48%	61.10% ± 6.27%	64.13% ± 5.62%
SVM	73.53% ± 3.99%	74.01% ± 2.66%	76.04% ± 3.83%
KNN	72.34% ± 6.30%	73.99% ± 6.92%	74.55% ± 6.07%

The accuracy of the eye tracking is not as high as EEG in depression detection area. However the eye tracking possess several irreplaceable advantages. Firstly, the price of equipment used in eye tracking studies are much lower than EEG equipment. Secondly, in recent years, the eye tracking is becoming more and more easily to access than ever (even the Kinect V2 can support eye tracking and it only cost less than $200). In the future, with the development of the camera and sensor in mobile phone we can even make an app for depression detection using eye tracking. Finally, the raw data of eye tracking contains few artifacts to remove, it even can be used directly and give a diagnosis very fast. For now depression detection researches using the eye tracking still in its infancy, we believe this situation will be improved with the development of eye tracking equipment and data mining methods. With a higher accuracy, eye tracking may become an ideal technology for depression detection.

4 Future Work

In the future, we will continually focus on depression detection algorithm improving and using a different experiment paradigm to collect eye movement data to get a more reliable classification accuracy for real-time applications, which will make eye tracking a more convenient, economical, and a popular method for depression detection.

Acknowledgement. This work was supported by the National Basic Research Program of China (973 Program) [No. 2014CB744600]; the National Natural Science Foundation of China [Nos. 61632014, 61210010]; the International Cooperation Project of Ministry of Science and Technology [No. 2013DFA11140]; and the Program of Beijing Municipal Science & Technology Commission [No. Z171100000117005].

References

1. Marcus, M., Yasamy, M.T., van Ommeren, M., Chisholm, D., Saxena, S.: Depression: a global public health concern. WHO Department of Mental Health and Substance Abuse 1, 6–8 (2012)
2. Brundtland, G.H.: Mental health: new understanding, new hope. JAMA **286**, 2391 (2001)
3. Mundt, J.C., Snyder, P.J., Cannizzaro, M.S., Chappie, K., Geralts, D.S.: Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. J. Neurolinguistics **20**, 50–64 (2007)
4. Sung, M., Marci, C., Pentland, A.: Objective physiological and behavioral measures for identifying and tracking depression state in clinically depressed patients. Massachusetts Institute of Technology Media Laboratory, Cambridge, MA, Tech. Rep. TR 595 (2005)
5. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J.: Eye tracking: a comprehensive guide to methods and measures. OUP Oxford (2011)
6. Peckham, A.D., McHugh, R.K., Otto, M.W.: A meta-analysis of the magnitude of biased attention in depression. Depression Anxiety **27**, 1135–1142 (2010)
7. Lu, S.: Attentional bias scores in patients with depression and effects of age: a controlled, eye-tracking study. J. Int. Med. Res. **45**, 1518–1527 (2017)
8. Li, X., Cao, T., Sun, S., Hu, B., Ratcliffe, M.: Classification study on eye movement data: towards a new approach in depression detection. In: 2016 IEEE Congress on Evolutionary Computation (CEC), pp. 1227–1232. IEEE (2016)
9. Alghowinem, S., Goecke, R., Wagner, M., Parker, G., Breakspear, M.: Eye movement analysis for depression detection. In: 2013 20th IEEE International Conference on Image Processing (ICIP), pp. 4220–4224. IEEE (2013)
10. Alghowinem, S., et al.: Multimodal depression detection: fusion analysis of paralinguistic, head pose and eye gaze behaviors. IEEE Trans. Affect. Comput., 1 (2017)
11. Zhu, Y., Shang, Y., Shao, Z., Guo, G.: Automated depression diagnosis based on deep networks to encode facial appearance and dynamics. IEEE Trans. Affect. Comput., 1 (2017)
12. Girard, J.M., Cohn, J.F., Mahoor, M.H., Mavadati, S.M., Hammal, Z., Rosenwald, D.P.: Nonverbal social withdrawal in depression: evidence from manual and automatic analysis. Image Vis. Comput. **32**, 641–647 (2014)
13. Tottenham, N., et al.: The NimStim set of facial expressions: judgments from untrained research participants. Psychiatry Res. **168**, 242–249 (2009)
14. Han, J., Pei, J., Kamber, M.: Data Mining: Concepts and Techniques. Elsevier, New York (2011)
15. Zhang, X.-L., Liu, P.: A new delay jitter smoothing algorithm based on Pareto distribution in cyber-physical systems. Wirel. Netw. **21**, 1913–1923 (2015)
16. Zhao, H., Min, F., Zhu, W.: Cost-sensitive feature selection of numeric data with measurement errors. J. Appl. Math. **2013**, 1–13 (2013)
17. Cai, H., Zhang, X., Zhang, Y., Wang, Z., Hu, B.: A case-based reasoning model for depression based on three-electrode EEG data. IEEE Trans. Affect. Comput. 1 (2018)
18. Mogg, K., Bradley, B.P.: Attentional bias in generalized anxiety disorder versus depressive disorder. Cogn. Ther. Res. **29**, 29–45 (2005)