Assessment method of depressive disorder level based on graph attention network

Shengfu Lu[1,2,3,4], *Jiaming* Kang[1], *Jinyu* Zhang[1], and *Mi* Li[1,2,3,4,*]

[1]Department of Automation, Faculty of Information Technology, Beijing University of Technology, Beijing 100124
[2]Beijing International Collaboration Base on Brain Informatics and Wisdom Services,Beijing 100124, China
[3]Engineering Research Center of Intelligent Perception and Autonomous Control, Ministry of Education, Beijing 100124
[4]Engineering Research Center of Digital Community, Ministry of Education, Beijing 100124

Abstract. This paper presents an approach to predict the depression self-rating scale of Patient Health Questions-9 (PHQ-9) values from pupil-diameter data based on the graph attention network (GAT). The pupil diameter signal was derived from the eye information collected synchronously while the subjects were viewing the virtual reality emotional scene, and then the scores of PHQ-9 depression self-rating scale were collected for depression level. The chebyshev distance based GAT (Chebyshev-GAT) was constructed by extracting pupil-diameter change rate, emotional bandwidth, information entropy and energy, and their statistical distribution. The results show that, the error (MAE and SMRE)of the prediction results using Chebyshev-GAT is smaller then the traditional regression prediction model.

1 Introduction

Depression is one of the major mental disorders, which is characterized by affective disorder, and which core symptom is that emotional injury leads to low mood depression. Although there are currently effective treatments, early detection of depression can prevent it from developing into depression.

At present, a great deal of research has been done on depression assessment methods based on machine learning, besides traditional machine learning methods (such as SVM) [1] , deep learning methods have also been done on depression assessment[2-14].

However, the current depression assessment methods are mainly based on video and audio signals from the natural interactive interviews [14-17], and lack of research on depression assessment methods using physiological signals.

In this study, the change of pupil diameter induced by emotional video was used as the basic physiological signal, and the associated classification features, including emotional bandwidth, information entropy and information energy, were extracted, depression was assessed using graph convolutional neural network with attention.

[*] Corresponding author: limi@bjut.edu.cn

2 Proposed method

2.1 Data acquisition

To reduce the effect of light, which changes the size of the pupil, a helmet-mounted eye tracker was developed by the team, the hardware acquisition device uses high definition CMOS camera and 850 nm infrared light source to illuminate, and real-time acquisition of eye images at 25 FPS to get 1280*720 pixels images.

The acquired eye images are preprocessed (including dilation, erosion and smoothing), then the pupil diameter is obtained by pupil detection and pupil fitting.

2.2 Feature construction

In this study, the pupil diameter signal was used to calculate 14 indicators of different emotions, which including that: positive and negative emotional bandwidth; positive and negative information entropy and its change rate; positive information entropy based emotional bandwidth and negative information entropy based emotional bandwidth; the positive and negative information energy and its change rate; the positive information energy based emotional bandwidth and the negative information energy based emotional bandwidth; and the rate of positive and negative pupil diameter changes.

Firstly, the mean value of pupil diameter ($\overline{T}_C, \overline{T}_P\ and\ \overline{T}_N$), the information entropy of calm emotion(C_{entropy}), and the energy of calm emotion (C_{energy}) were calculated.

$$\overline{T}_C = \frac{1}{m}\sum_{t=1}^{m} T_C(t), \overline{T}_P = \frac{1}{m}\sum_{t=1}^{m} T_P(t)\ , \overline{T}_N = \frac{1}{m}\sum_{t=1}^{m} T_N(t) \tag{1}$$

$$C_{\text{entropy}} = -\frac{1}{m}\sum_{t=1}^{m} T_C(t)log(T_C(t))\ , C_{\text{energy}} = \sum_{t=1}^{m} T_C(t)^2 \tag{2}$$

Positive (P_{BW}) and Negative emotional bandwidth(N_{BW}):

$$P_{BW} = \overline{T}_P - \overline{T}_C, N_{BW} = \overline{T}_N - \overline{T}_C \tag{3}$$

Positive entropy$\left(P_{\text{entropy}}\right)$ and its rate of change ($P_{\text{entropy}}\%$):

$$P_{\text{entropy}} = -\frac{1}{m}\sum_{t=1}^{m} T_P(t)log(T_P(t)) \tag{4}$$

$$P_{\text{entropy}}\% = (P_{\text{entropy}} - C_{\text{entropy}})/C_{\text{entropy}} * 100 \tag{5}$$

Positive entropy$\left(P_{\text{entropy}}\right)$ based emotional bandwidth ($P_{\text{entropy−BW}}$):

$$P_{\text{entropy−BW}} = P_{\text{entropy}} - C_{\text{entropy}} \tag{6}$$

Negative entropy(N_{entropy}) and its rate of change ($N_{\text{entropy}}\%$):

$$N_{\text{entropy}} = -\frac{1}{m}\sum_{t=1}^{m} T_N(t)log(T_N(t)) \tag{7}$$

$$V_{\text{entropy}}\% = (N_{\text{entropy}} - C_{\text{entropy}})/C_{\text{entropy}} * 100 \tag{8}$$

Negative entropy(N_{entropy}) based emotional bandwidth($N_{\text{entropy−BW}}$):

$$N_{\text{entropy-BW}} = N_{\text{entropy}} - C_{\text{entropy}} \tag{9}$$

Positive energy (P_{energy}) and its rate of change($P_{\text{energy}}\%$):

$$P_{\text{energy}} = \sum_{t=1}^{m} T_P(t)^2 \tag{10}$$

$$P_{\text{energy}}\% = (P_{\text{energy}} - C_{\text{energy}})/C_{\text{energy}} * 100 \tag{11}$$

Negative energy$\left(N_{\text{energy}}\right)$ and its rate of change($N_{\text{energy}}\%$):

$$N_{\text{energy}} = \sum_{t=1}^{m} T_N(t)^2 \tag{12}$$

$$N_{\text{energy}}\% = (N_{\text{energy}} - C_{\text{energy}})/C_{\text{energy}} * 100 \tag{13}$$

The rate of positive and negative pupil diameter changes ($\Delta T_P\%$, $\Delta T_N\%$):

$$\Delta T_P\% = \frac{\bar{T}_P - \bar{T}_C}{\bar{T}_C} * 100 \tag{14}$$

$$\Delta T_N\% = \frac{\bar{T}_N - \bar{T}_C}{\bar{T}_C} * 100 \tag{15}$$

On this basis, 9 kinds of features are extracted according to the statistical distribution, which are: Mean, median, upper quartile, lower quartile; deviation trend: maximum, minimum, Standard Deviation; distribution pattern: skewness coefficient, Kurtosis Coefficient. Finally, 136 features were extracted from each subject to assess the level of depression.

2.3 Chebyshev-GAT

Graph attention network (GAT) combines graph convolutional neural network (GCN) and attention mechanisms. When GAT aggregates node information, each neighbor node is assigned a different weight (attation sroce). GAT often uses multi-heads attention mechanism, each head update parameters separately, then several attention head take the average to get the node expression.

In this study, there is no graph network data, and there is no explicit relationship between the samples. To assess depression levels with GAT, construct a graph network:

Nodes use samples or subjects, so 217 samples make up 217 nodes of GCN.

Chebyshev distance is used to calculate the similarity ($Sim_c(X_v, X_w)$) between nodes X_v and X_w:

$$Sim_c(X_v, X_w) = \lim_{k\to\infty}\left(\sum_{i=1}^{m} |x_{vi} - x_{wi}|^k\right)^{1/k} \tag{16}$$

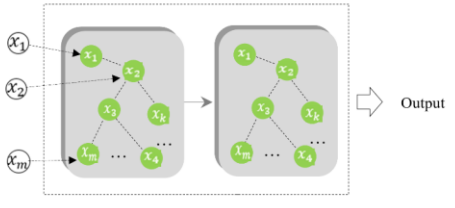

Fig. 1. GCN network structure.

When $Sim(X_v, X_w) \geq 0.5$, the similarity between sample X_v and X_w is considered, there are edges between nodes, and when $Sim(X_v, X_w) < 0.5$, there are no edges between sample X_v and X_w. Chebyshev-GAT as follows:

3 Model training strategy and prediction results

3.1 Training strategy

In this study, 217 samples were grouped according to the ratio of training set: verification set: Test set = 8:1:1, i.e. 173 in training set, 22 in verification set and 22 in test set.

The learning rate (lr), number of hidden layers (H) and number of attention heads (A) of the three important hyper-parameters being tested is:lr=0.001, H=2, A=6。The regularization coefficient of L2 is 5e-4 and the number of iteration is 1000.

The loss function is calculated using mean square error (MSE) :

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{17}$$

The mean absolute error (MAE) and root mean square error (RMSE) were used to evaluate the prediction results:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{18}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{19}$$

$\hat{y}$ is the predicted value of the sample label, and y_i is the true value of the sample label.

3.2 Prediction results

Table 1 is a comparison with traditional machine learning methods. The results show that compared with traditional machine learning methods, the method presented in this study has the best performance in depression level assessment, and the model error is smaller.

Table 1. Comparison of prediction results with traditional machine learning regression methods.

Method	MAE	RMSE
SVM-R	4.33	6.34
KELM-R	3.92	5.89
RF-R	3.27	4.55
Our proposed method	**2.66**	**3.97**

Note: SVM(support vector machines), KELM(kernel extreme learning machine), RF(random forest), R(regression)

Table 2. Comparison with other researchs.

Paper	MAE	RMSE	Prediction score
Valstar et al.[16]	6.12	6.97	PHQ-8
Yang et al. [18](females)	3.26	3.97	PHQ-8
Yang et al. [18](males)	3.19	4.29	PHQ-8
Williamson et al.[19]	5.33	6.45	PHQ-8
Haque et al.[2]	3.67		PHQ-8
Our proposed method	**2.66**	**3.97**	**PHQ-9**

4 Conclusion

We have proposed graph attention network based deep learning method for small-scale data sets is superior to the traditional machine learning method for small-scale data sets in depression assessment, at the same time, compared with the existing methods based on behavioral data (such as expression, speech and language) , this method has achieved a better results. In addition, the research method has the advantages of strong pertinence, simple data acquisition and low calculation, which is of great practical value in mental state examination.

This work is supported by the National Natural Science Foundation of China (61602017), the National Basic Research Programme of China (2014CB744600), 'Rixin Scientist' Foundation of Beijing University of Technology (2017-RX(1)-03), the Beijing Natural Science Foundation (4164080), the Beijing Outstanding Talent Training Foundation (2014000020124G039), the National Natural Science Foundation of China (61420106005), the International Science & Technology Cooperation Program of China (2013DFA32180).

References

1. Pampouchidou, P. Simos, K. Marias, F. Meriaudeau, F. Yang, M. Pediaditis, and M. Tsiknakis, "Automatic assessment of depression based on visual cues: A systematic review" IEEE Trans. Affect. Comput., 1–27 (2017).
2. Haque, Albert, et al. "Measuring depression symptom severity from spoken language and 3D facial expressions." arXiv preprint arXiv:1811.08592 (2018).
3. W. C. De Melo, E. Granger, M. B. Lopez, "Encoding temporal information for automatic depression recognition from facial analysis. " *In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1080-1084 (2020).
4. W. C. De Melo, E. Granger, A. Hadid, "Depression detection based on deep distribution learning." *In 2019 IEEE International Conference on Image Processing (ICIP)*, 4544-4548 (2019).
5. A. Jan, H. Meng, Y. F. B. A. Gaus, and F. Zhang, "Artificial intelligent system for automatic depression level analysis through visual and vocal expressions," IEEE Trans. Cogn. Develop. Syst. **10**, 668–680 (2018).
6. Y. Zhu, Y. Shang, Z. Shao, and G. Guo, "Automated depression diagnosis based on deep networks to encode facial appearance and dynamics,"IEEE Trans. Affect. Comput. **9**, 578–584 (2018).
7. W. C. de Melo, E. Granger, and A. Hadid, "Combining global and local convolutional 3d networks for detecting depression from facial expressions," In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition* (2019).
8. M. A. Jazaery and G. Guo, "Video-based depression level analysis by encoding deep spatiotemporal features," IEEE Trans. Affect. Comput., 1–8 (2018).
9. X. Zhou, K. Jin, Y. Shang, and G. Guo, "Visually interpretable representation learning for depression recognition from facial images," IEEE Trans. Affect. Comput. **11**, 3, 542-552 (2018).
10. X. Zhou, P. Huang, H. Liu, S. Niu. Learning content-adaptive feature pooling for facial depression recognition in videos. Electronics Letters, **55,** 11, 648–650 (2019).

11. X. Zhou, K. Jin, Y. Shang, et al.: ‘Visually interpretable representation learning for depression recognition from facial images’, Trans. Affect. Comput. **11**, 3, 542-552 (2018).
12. M. Muzammel, H. Salam, Y. Hoffmann, M. Chetouani, A. Othmani. AudVowelConsNet: A phoneme-level based deep CNN architecture for clinical depression diagnosis. Machine Learning with Applications **2**, 100005 (2020).
13. L. He, C. Cao. Automated depression analysis using convolutional neural networks from speech. J. Biomed. Inform. **83**, 103-111 (2018).
14. M. Valstar, et al., “AVEC 2013: The continuous audio/visual emotion and depression recognition challenge,” *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*., 3–10 (2013).
15. M. Valstar, et al., “AVEC 2014: 3D dimensional affect and depression recognition challenge,” *Proceedings of the 4th international workshop on audio/visual emotion challenge*, 3–10 (2014).
16. M. Valstar, J. Gratch, B. Schuller, et al. Avec 2016: Depression, mood, and emotion recognition workshop and challenge. *Proceedings of the 6th international workshop on audio/visual emotion challenge* (2016).
17. F. Ringeval, et al., AVEC 2017 - Real-life Depression, and Affect Recognition Workshop and Challenge, *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge* (2017).
18. L. Yang, D. Jiang, L. He, E. Pei, M. C. Oveneke, and H. Sahli, “Decision tree based depression classification from audio video and language information,” *Proceedings of the 6th international workshop on audio/visual emotion challenge.*, 89–96 (2016).
19. J. R. Williamson, et al., “Detecting depression using vocal, facial and semantic communication cues,” *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*, 11–18 (2016).