Emotion recognition based on multi-channel EEG signals

Huiping Shi¹, Hong Xie¹* and Mengran Wu²
¹,² Shanghai Maritime University, Shanghai 201306, China
*Corresponding author’s e-mail: hongxie@shmtu.edu.cn

Abstract. Emotion recognition is a key technology of human-computer emotional interaction, which plays an important role in various fields and has attracted the attention of many researchers. However, the issue of interactivity and correlation between multi-channel EEG signals has not attracted much attention. For this reason, an EEG signal emotion recognition method based on 2DCNN-BiGRU and attention mechanism is tentatively proposed. This method firstly forms a two-dimensional matrix according to the electrode position, and then takes the pre-processed two-dimensional feature matrix as input, in the two-dimensional convolutional neural network (2DCNN) and the bidirectional gated recurrent unit (BiGRU) with the attention mechanism layer Extract spatial features and time domain features in, and finally classify by softmax function. The experimental results show that the average classification accuracy of this model are 93.66% and 94.32% in the valence and arousal, respectively.

1. Introduction
Emotions are people's reactions to external events and affect people's bodies. When exposed to external things, they will make emotional reactions, including positive emotional reactions and negative emotional reactions [1]. Emotions play a vital role in human-computer interaction, medicine, military and other fields. Studies have shown that in past experiments, emotion recognition is usually based on brain electrical signals, sounds or facial expressions. Among them, the emotion recognition method based on EEG signals is more accurate. Therefore, emotion recognition methods based on EEG signals are favored by more and more scholars. The typical emotion recognition method is to use traditional machine learning models, which need to manually extract the characteristic parameters of the signal through different analysis methods, and then use different classifiers to classify the signal. In previous studies, deep learning has also made some progress in the emotional classification of EEG signals, but there is still a lot of room for development[2-4]. Therefore, this article tentatively proposes a multi-channel EEG signal emotion recognition method based on 2DCNN-BiGRU and attention mechanism. This method inputs two-dimensional features into the model for emotion recognition, and the two-dimensional input greatly preserves the spatial characteristics of the EEG signal and the correlation between electrodes. Experiments show that the classification accuracy of this model is significantly higher than other models.

2. Data source and preprocessing

2.1. Data source
The experimental research data used a public DEAP data set[5]. This article uses the EEG signals of the first 32 channels. Each subject performed 40 trials. After each trial, the subject was asked to complete the self-assessment manikin (SAM) potency, arousal, superiority, and liking. Each self-
assessment dimension ranges from 1 to 9. Russell's valence arousal scale model[6] is widely used in sentiment analysis due to its simplicity. Figure 1 is a two-dimensional emotional space, and the emotional state is described in a two-dimensional space.

![Two-dimensional emotional space](image1)

Figure 1. Two-dimensional emotional space.

2.2. Data preprocessing and construction of two-dimensional matrix

This experiment uses preprocessed EEG data (down-sampling from 512HZ to 128HZ, filtering with a bandpass filter of 4.0 ~ 45.0HZ, removing ocular artifacts, etc.). Differential entropy (DE) feature extraction is performed on the EEG signal. If the signal obeys a different distribution, DE will be solved by a different method[4]. Assuming that the EEG signal X obeys Gaussian distribution, the formula is as follows:

\[
DE = -\int_{-\infty}^{\infty} p(x) \log[p(x)] \, dx
\]

(1)

\[
p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}
\]

(2)

\[
DE = \frac{1}{2} \log(2\pi e\sigma^2)
\]

(3)

Where DE stands for differential entropy.

The electrode distribution position according to the international 10-20 system is mapped into a 9×9 two-dimensional matrix, but since this data set collects 32 channels of EEG signals, in order to maintain the integrity of the spatial information without affecting the characteristics, Fill the extracted features to the position of the used channel, and use 0 to represent the unused channel. The international 10-20 system and its mapping matrix are shown in Figure 2.

![International 10-20 lead system and its mapping matrix](image2)

Figure 2. International 10-20 lead system and its mapping matrix.
3. Formatting the text

The model used in this article takes a two-dimensional matrix as the input of a two-dimensional convolutional neural network, extracts spatial features through the two-dimensional convolutional neural network, and then inputs the result into a two-way gated recurrent unit to extract temporal features, and finally extracts. The obtained features are input into the self-attention mechanism layer, and attention information is extracted according to the importance of the sample and classified by the softmax function.

3.1. 2DCNN module

CNN[7] is a feedforward neural network with a two-dimensional convolutional layer and a two-dimensional pooling layer, which was originally developed in computer vision. A typical convolutional neural network includes a convolutional layer, a pooling layer and a fully connected layer. The network framework of 2DCNN is shown in Figure 3.

![2DCNN network framework](image)

Taking the two-dimensional feature matrix as the input of 2DCNN, the formula is:

\[
\hat{s}_{i,j} = \sum_{m} \sum_{n} f(i + m, j + n) w(m, n)
\]

Where \( W \) is the convolution kernel, \((m, n)\) is the size of the convolution kernel \( W \), \( f(c) \) is the input matrix, and \((i, j)\) is the matrix coordinates. After each convolution operation, batch normalization (BN) is performed on the feature data of each layer, and the RELU activation function is added to make the model have nonlinear feature conversion capabilities. The RELU function is expressed as

\[
\text{RELU}(x) = \max(x, 0) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}
\]

Where \( \text{max} \) is the maximum function and \( x \) is the input.

In this model, the corrected linear unit (ReLU) function is used for all convolutional layers. Using the ReLU function can reduce the amount of calculations and quickly converge, effectively alleviating the problem of gradient disappearance, and the ReLU function will also make some neurons output 0, providing the sparse ability of the neural network and reducing the dependence between parameters. It also reduces overfitting. Table 1 shows the main parameters in the 2DCNN model.

| Layer | Filters | Size  | Stride |
|-------|---------|-------|--------|
| Cov1  | 64      | (5,5) | 1      |
| Cov2  | 128     | (4,4) | 1      |
| Cov3  | 256     | (4,4) | 1      |
| Cov4  | 64      | (1,1) | 1      |
| pool  | 64      | (2,2) | 1      |
3.2. BiGRU module

GRU[8] is an improvement of LSTM. GRU not only inherits the control principle of LSTM, but also simplifies the neuron structure and reduces the complexity of the model. Compared with LSTM, GRU has smaller data processing capabilities, fewer parameters, faster calculation speed, stronger convergence, and can better solve the problem of gradient disappearance. The GRU consists of two gates, namely the update gate $z_t$ and the reset gate $r_t$. The GRU uses these two gates to process timing information. The calculation formula is as follows:

$$ z_t = \sigma(w_r x_t \oplus U_r h_{t-1}) $$

(6)

$$ r_t = \sigma(w_z x_t \oplus U_z h_{t-1}) $$

(7)

$$ h_t = \tanh\left[w_h x_t \oplus U_h(r_t h_{t-1})\right] $$

(8)

$$ h_t = (1-z_t)h_{t-1} \oplus z_t h_t $$

(9)

Among them, $w_r$, $w_z$, $U_r$, $U_z$ and $U_h$ are the weight matrices of GRU, $\overrightarrow{h_t}$ is the candidate activation unit, $h_t$ is the hidden unit at time $t$, $h_{t-1}$ is the hidden unit at time $t-1$, and $\sigma$ is the activation function. $X_t$ is the input of GRU, $\oplus$ is the multiplication element, and $\oplus$ is the addition element. In this paper, the BiGRU network is composed of forward GRU, backward GRU, and forward and backward output state connection layers. As shown in Figure 4, it mainly includes an input layer, a hidden layer and an output layer.

![Figure 4. BiGRU network structure.](image)

3.3. Attention mechanism

The self-attention mechanism is also a kind of attention mechanism [9]. Since the development of the self-attention mechanism, it has attracted widespread attention and proved its amazing ability through applications in many fields. In addition, since it does not require recursion and convolution modules, its training efficiency is greatly improved and training time is reduced. In this paper, a self-attention mechanism [10] layer is added after BiGRU to connect different positions of a single input sequence to calculate the representation of the same sequence. To obtain these representations, each input must be multiplied by a set of keyword weights (indicated by $K$), a set of query weights (indicated by $Q$), and a set of value weights (indicated by $V$). Then, pay attention to learning a function to map the query $Q$ to a series of key-value pairs $(K, V)$, as shown below:

$$ \text{Attention} V = Q K^T V $$

(10)

Attention is essentially to assign a weight coefficient to each element in the sequence, which can also be understood as soft addressing. If each element is stored, the degree of interest can calculate the similarity between $Q$ and $K$. The similarity calculated by $Q$ and $K$ reflects the importance of the extracted V value, which is the weight, and then the weighted sum is added to obtain the degree of interest value. The special feature of the self-attention mechanism in the K, Q, V model is that $Q=K=V$:
\[
\text{Attention}(Q,K,V) = \text{soft max}\left(\frac{QK^T}{\sqrt{d_K}}\right)V
\]

(11)

4. Simulation results and analysis

Aiming at the research on emotion recognition of EEG signals, this paper proposes an emotion recognition model based on multi-channel EEG signals 2DCNN-BiGRU and attention mechanism. In this experiment, in order to obtain more credible results, 80% of the data set is used as the training set and 20% as the test set. The classification effect of the model proposed in this article is evaluated using different indicators, including overall accuracy (accuracy, ACC), precision (PR), specificity (SP), recall (recall, RE), F1 score (F1 score, F1). The specific calculation method of each index[11] is as follows:

\[
\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN}
\]

(12)

\[
\text{PR} = \frac{TP}{TP + FP}
\]

(13)

\[
\text{SP} = \frac{TN}{TN + FP}
\]

(14)

\[
\text{RE} = \frac{TP}{TP + FN}
\]

(15)

\[
\text{F1} = 2 \times \frac{\text{PR} \times \text{RE}}{\text{PR} + \text{RE}}
\]

(16)

Where TP means the number of positive classes predicted as positive classes, TN means the number of negative classes predicted as negative classes, FP means the number of negative classes predicted as positive classes, and FN means the number of positive classes predicted as negative. Since this article is studying a multi-classification problem, the category to be evaluated is regarded as the positive category, and the other categories are regarded as the inverse category to calculate each index.

This paper proposes 2DCNN-BiGRU and attention mechanism models, which have achieved certain results in various indicators such as accuracy and precision. In order to better show the experimental results, Figures 5 show the accuracy of emotion recognition and classification of 32 subjects in the valence dimension and in the arousal dimension. The accuracy rates are mostly between 90% and 98%. Fluctuate up and down from time to time. Since the training and testing samples are randomly selected in each experiment in this article, the accuracy of individual subjects may fluctuate widely.

![Figure 5. The classification accuracy of 32 subjects in the valence and arousal dimension.](image)

Table 2 shows the average classification accuracy, precision, specificity, recall rate, and F1 score of this model in terms of potency and arousal. From Table 3, it can be seen that the average classification
accuracy has reached more than 90%, and it has achieved good results. The effect has a certain reference value.

Table 2. Experimental results evaluation index.

| Evaluation index | Valence | Arousal |
|------------------|---------|---------|
| ACC              | 93.66%  | 94.32%  |
| PR               | 92.61%  | 92.38%  |
| SP               | 93.36%  | 93.21%  |
| RE               | 94.19%  | 94.14%  |
| F1               | 94.52%  | 94.84%  |

The experimental results are compared with other documents. As shown in Table 3, Huifang Yao[12] and others used the multi-scale window deep forest method to extract features through multiscale window multi-grain scanning, and use cascaded forest to classify features. The average accuracy rate reached 84.90%; Chen Tian[13] and others used wavelet packet decomposition and Hilbert-Huang transform to extract EEG signal features, combined with convolutional neural network, recurrent neural network and support vector machine. For classification, the average accuracy rate reached 86.22%; Liu Changyuan[14] et al. used an improved multi-scale attention residual network for feature extraction and classification, and the average accuracy rate reached 85.20%; YongPeng[15] et al. Using the popular regularized extreme learning machine method to classify the extracted differential entropy features, the average accuracy rate reaches 81.01%.

Table 3. Compare with other documents.

| References | Model   | average accuracy |
|------------|---------|------------------|
| References[12] | MSWDF   | 84.90%           |
| References[13] | C-R-SVM | 86.22%           |
| References[14] | MAResnet| 85.20%           |
| References[15] | MRELM   | 81.01%           |
| This paper        | 2DCNN-BiGRU | 93.99%           |

5. Conclusion
This article uses the multi-channel EEG signal research to greatly retain the spatial characteristics and correlation between the electrodes. The 2DCNN-BiGRU model is used to extract the spatiotemporal features of EEG signals, and the attention mechanism layer extracts important attention information. Experiments show that the average classification accuracy of the model in the dimensions of arousal and valence is 93.99% and 94.32%, respectively. Compared with the traditional model, the accuracy has been significantly improved. In future research, more features can be extracted to further improve accuracy.

Acknowledgments
Project Fund: National Natural Science Foundation of China (61550110252);

References
[1] Ante Topic, Mladen Russo, Emotion recognition based on EEG feature maps through deep learning network, Engineering Science and Technology, an International Journal, 2021, ISSN 2215-0986.
[2] Chen Jing-xia, Zheng Ru, Jia Xiao-yun, Zhang Peng-wei. EEG Emotion Classification Based on Common Spatial Patterns and Wavelet Packet Decomposition [J]. Computer Engineering and Applications, 2019, 55 (01): 149-153.
[3] Zhu Yong-sheng, Zhong Qing-hua, Cai Dong-li, Liao Jin-xiang. Fast sample entropy electroencephalogram emotion analysis of double sliding coarse granulation [J]. Application Research of Computers, 2021, 38(01): 149-153.
[4] Huiping Jiang, Rui Jiao, Zequn Wang, Ting Zhang, Licheng Wu, "Construction and Analysis of Emotion Computing Model Based on LSTM", Complexity, vol. 2021, Article ID 8897105, 12 pages, 2021.
[5] Koelstra S, Muhl C, Soleymani M, et al. DEAP: A Database for Emotion Analysis Using Physiological Signals [J]. IEEE Transactions on Affective Computing, 2012, 3(1): 18-31.
[6] Russell, J. A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology,
39(6), 1161–1178.

[7] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," in Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.

[8] XIAO Ping-wan, WANG Zi-niu, GAO Jian-ling. Multi-label text classification based on BiGRU-Capsule[J]. Intelligent Computer and Applications, 2021, 11(05): 193-197.

[9] WANG Xiao-ru, ZHANG Heng. Relation Network Based on Attention Mechanism and Graph Convolution for Few-shot Learning[J/OL]. Computer Engineering and Applications: 1-8 [2021-06-28].

[10] S. Sun, Y. Tang, Z. Dai and F. Zhou, "Self-Attention Network for Session-Based Recommendation With Streaming Data Input," in IEEE Access, vol. 7, pp. 110499-110509, 2019.

[11] A LSTM based deep learning network for recognizing emotions using wireless brainwave driven system, Expert Systems with Applications, Volume 173, 2021, 114516, ISSN 0957-4174.

[12] H. Yao, H. He, S. Wang and Z. Xie, "EEG-based Emotion Recognition Using Multiscale Window Deep Forest," 2019 IEEE Symposium Series on Computational Intelligence (SSCI), 2019, pp. 381-386.

[13] CHEN Tian, CHEN Zhan-gang, YUAN Xiao-hui, JU Si-hang, REN Fu-ji. Emotion Recognition Method Based on Instantaneous Energy of Electroencephalography [J]. Computer Engineering, 2019, 45(04): 196-204.

[14] Liu Chang-yuan, Sun Yu-han, Li Wen-qiang, Lan Chao-feng. Research on EEGemotion recognition based on MAResnet [J]. Chinese Journal of Scientific Instrument, 2020, 41(07): 235-242.

[15] Peng Y, Zhu JY, Zheng WL, Lu BL. EEG-based emotion recognition with manifold regularized extreme learning machine. Annu Int Conf IEEE Eng Med Biol Soc. 2014;2014:974-7.