Emotion recognition model based on CLSTM and channel attention mechanism

Yuxia Chen[1], *Dan* Wang[1,*], *Xiaoxi* Wang[2]

[1]Faculty of Information Technology, Beijing University of Technology, Beijing, China
[2]State Grid Management College, Beijing, China

Abstract. In this paper, we propose an emotion recognition model based on convolutional neural network (CNN), long short term memory (LSTM) and channel attention mechanism, aiming at the low classification accuracy of machine learning methods and the uneven spatial distribution of electroencephalogram (EEG) electrodes. This model can effectively integrate the frequency, space and time information of EEG signals, and improve the accuracy of emotion recognition by adding channel attention mechanism after the last convolutional layer of the model. Firstly, construct a 4-dimensional structure representing EEG signals. Then, a CLSTM model structure combining CNN and LSTM is designed. CNN is used to extract frequency and spatial information from 4-dimensional input, and LSTM is used to extract time information. Finally, the channel attention module is added after the last convolutional layer of CLSTM model structure to allocate the weight of different electrodes. In this paper, an emotion recognition model based on CLSTM and channel attention mechanism was proposed from the perspective of integrating the frequency, space and time 3-dimensional information of EEG signals. The average classification accuracy of the model on SEED public data set reached 93.36%, which was significantly improved over the existing CNN and LSTM emotion recognition models.

Keywords: Emotion recognition, EEG, CNN, LSTM, Attention mechanism.

1 Introduction

In recent years, with the development of artificial intelligence, more and more deep learning methods have been applied to process physiological signals such as EEG, EMG and ECG. EEG is favoured by many experts and scholars because of its rich information in time domain, frequency domain and space domain. However, it is still a challenge to integrate EEG information from different domains in order to obtain a better emotion recognition model [1]. At present, most scholars' research mainly focuses on two aspects: the one is to find new features that can better represent EEG signals; the other is to design a network model more suitable for EEG emotion recognition. In view of the above problems,

* Corresponding author: wangdan@bjut.edu.cn

domestic and foreign scholars have made a lot of attempts.

Liao [2] proposed a feature that integrates shallow and deep emotional features to obtain higher emotional characterization, and the recognition accuracy of this feature in DEAP data set is 15.10% and 0.94% higher than that of shallow and deep emotional features alone. Yang [3] proposed a 3-dimensional feature representation method, which retained the frequency and spatial information of EEG signals, and adopted a continuous convolutional network in the DEAP data set to achieve average recognition rates of 90.24% and 89.45% in the validity and arousal dimensions. In literature [4], wavelet transform was used to obtain 2-dimensional time-frequency maps of EEG signals, and adaptive CNN model was used in DEAP data set to achieve average recognition rates of 76.56% and 80.46% in terms of potency and arousal.

At present, most methods using deep learning for EEG emotion recognition do not simultaneously consider the time, frequency and space information of EEG signals. In order to solve the above problem and the uneven distribution of spatial information of EEG electrodes, this paper proposes an emotion recognition model ACLSTM based on CNN-LSTM and channel attention mechanism. This model can effectively integrate the frequency, space and time information of EEG signals. In addition, the introduction of attention mechanism can assign weight to different channels to enhance the key emotional information in EEG signals.

2 Methods

This model integrates CNN and LSTM. It mainly includes three parts, namely constructing 4-dimensional feature structure, model based on CLSTM and channel attention mechanism, and classifier. Firstly, the original EEG signal was transformed from 2-dimensional structure to 4-dimensional structure integrating frequency and spatial information. Then, CNN was used to extract frequency and spatial information from 4-dimensional inputs. After the last convolutional layer of CNN network, channel attention mechanism was added to allocate weights of different electrodes adaptively. LSTM was used to extract time information of EEG signals from the output of CNN. Finally, the output of the last node of the LSTM is fed into the SoftMax classifier for triage.

2.1 4-dimensional structural features

Since the EEG signal acquisition devices are placed in different positions of the cerebral cortex, the positional relationship between these electrodes contains intrinsic information, and making full use of this information can improve the performance of emotion recognition. Therefore, In order to maintain the spatial information of the electrode position, the position information of the 62 channels is mapped to a 2-dimensional map. The height of the 2-dimensional map is 8 and the width is 9. The mapping relationship is shown in Figure 1. Use 0 to indicate that the channel is not used. The values in the matrix represent the corresponding channel names. The 2-dimensional plot of 8*9 is chosen because for each electrode, the eigenvalues are closely surrounded by the values of adjacent electrodes, and too tight features may lead to information leakage in convolutional or pooling.

With the above mapping relationship, the 4-dimensional differential entropy feature of the original EEG signal is constructed. First, in order to increase the number of samples, the original EEG signal is divided into non-overlapping 1S-length signal segments. Then, the signal segments are divided into 4 frequency bands by the Butterworth filter, and the differential entropy features are calculated, and then converted into a 2-dimensional feature maps with spatial information and then spliced vertically. At this time, an 8*9*4 3-dimensional feature matrix with differential entropy features is obtained. Finally, since

each EEG signal is divided into non-overlapping N segments of 1S- length, the 3-dimensional feature matrix of each signal segment is obtained, and finally a 4-dimensional feature matrix of N*8*9*4 is obtained.

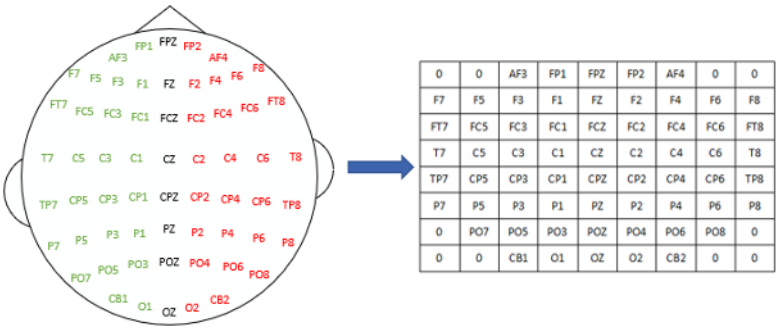

Fig. 1. 62 channels mapping 2D matrix.

2.2 CLSTM model

The 4-dimensional structure obtained by the above method is used to extract frequency and spatial information through CNN. Compared with the traditional CNN model structure, in order to reduce information loss, we only add a pooling layer after the last convolutional layer, and then the output of the pooling layer is flattened and input to the fully connected layer. Then LSTM is used to extract time information from THE output of CNN and classify the input into three categories. The CLSTM model of this scheme is shown in Figure 2. The model was used to triage the differential entropy features of EEG extraction from SEED data set, and the average accuracy was 87.36%.

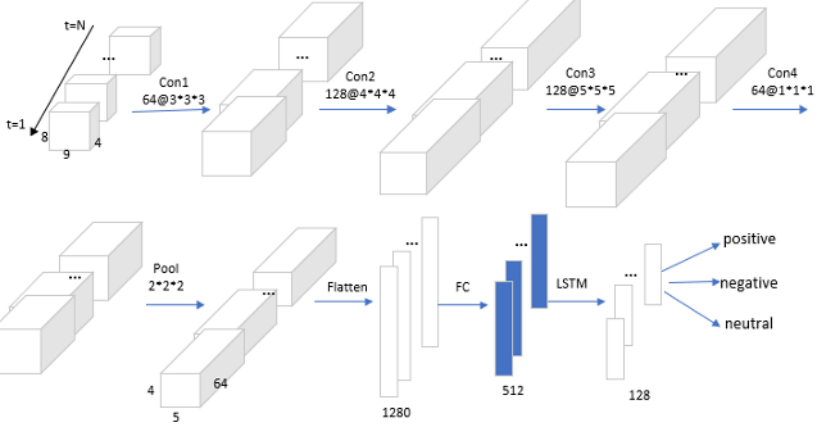

Fig. 2. CLSTM model structure diagram.

2.3 ACLSTM emotion recognition model based on channel attention mechanism

Attention mechanism plays an important role in human perception. An important characteristic of human visual system is that people will use a series of glimpses of parts in a scene and selectively focus on the prominent parts. In addition, EEG in different brain regions has different ability to express emotion. In recent years, attention mechanism has been gradually introduced into EEG emotion recognition to explore the emotion-related

characteristics of EEG in time, space and frequency [5]. Therefore, this chapter proposes an ACLSTM emotion recognition model based on channel attention mechanism, adding channel attention mechanism to the last convolutional layer of CLSTM model, making weighted adjustment of its output signal, and making better use of the channels significantly related to emotion in EEG signals. The attention mechanism module used in this chapter is shown in Figure 3.

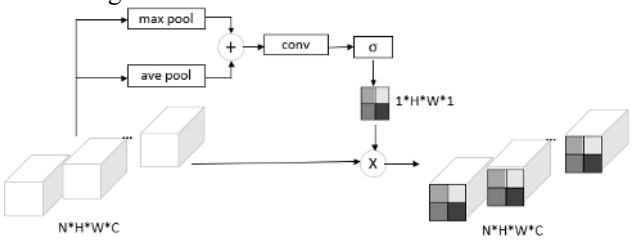

Fig. 3. Channel attention mechanism module.

As shown in Figure 3, for the output feature $X^{'}$ of the last convolutional layer of the CLSTM model, max pool and average pool are performed respectively, and the two are spliced to obtain a new feature $S \in 1*H*W*2$. Finally, 4-dimensional convolutional and sigmoid activation function are successively applied to generate channel attention force $F \in 1*H*W*1$ for S. The process can be expressed as follows:

$$S_{\max,(h,w)} = \max_{n}(\max_{c}(X^{'}{}_{h,w}))$$

$$S_{ave,(h,w)} = \frac{1}{N*C}\sum_{n=1}^{N}\sum_{c=1}^{C}(X^{'}{}_{h,w})$$

$$S = cat(S_{\max,(h,w)}, S_{ave,(h,w)})$$

$$F(X^{'}) = sigmoid(Conv2(S))$$

Where, $X^{'}{}_{h,w} \in N*C$ represents the values of all channels corresponding to row h and column w of the input signal $X^{'}$.

Finally, the obtained H*W channel attention weight coefficient is applied to feature $X^{'}$, and finally the output feature Y of channel attention mechanism is obtained, that is, the output signal after the input signal of the last convolutional layer passes through the channel attention mechanism.

3 Experiment

3.1 SEED data set

The SEED data set was provided by the laboratory of Shanghai Jiao Tong University. Chinese movie clips with a duration of about 4 minutes were used to induce three emotions of positive, neutral and negative valences. The EEG data of 15 subjects were collected using a 62-lead Neuroscan system at a sampling rate of 1000 Hz. Each subject did 3 experiments at different times, and each experiment needed to watch 15 movie clips, for

feedback, participants were asked to complete a questionnaire immediately after viewing each clip to report their emotional response to each clip.

3.2 Experimental setup

In order to increase the number of samples, the original EEG signal is divided into non-overlapping signal segments. The length of the EEG signal segment will determine how much emotional information it contains. Therefore, we take 0.5S, 1S, 1.5S, 2S and 3S respectively. The length of the signal segment is compared in the CLSTM model. Table 1 shows the classification accuracy and variance under different length signal segments. Two pieces of information can be seen from the table. First, the performance obtained by intercepting the EEG signal segment with a length of 1S is the best; It can make full use of EEG signal information to classify emotions, and is not affected by the length of the signal segment.

Table 1. Classification performance of signal segments of different lengths.

T(S)	Accuracy (%)	Standard
0.5	86.48	2.34
1	**87.36**	2.57
1.5	87.01	2.37
2	86.96	2.51
3	86.83	3.57

In order to verify that our model is effective, the following comparative experiments are now performed to compare with several commonly used deep learning methods:

Table 2. Performance comparison of different models.

Number	Method	Information	Accuracy (%)	Standard
1	CCNN	frequency+space	88.60	2.60
2	EmotionNet	time+space	73.40	3.13
3	CNN	frequency+time	73.11	4.81
4	CLSTM	frequency+time+space	87.36	2.57
5	ACLSTM	frequency+time+space	**93.36**	2.25

CCNN [3]: The model is a continuous CNN model consisting of 4 convolutional layers and a fully connected layer. The method uses the extracted 3-dimensional EEG structure of four frequency bands θ,α,β and γ as input, and integrates the frequency and spatial information of the EEG signal at the same time.

EmotionNet [6]: The model is a continuous CNN model consisting of 4 convolutional layers and a fully connected layer. The method uses the extracted 3-dimensional EEG structure of four frequency bands θ,α,β and γ as input, and integrates the frequency and spatial information of the EEG signal at the same time.

CNN: This is a CNN model designed in this paper, which contains three convolutional layers and two fully connected layers. Each convolutional layer has a BatchNorm2d layer, an activation function layer, and an average pooling layer. The model only considers the frequency and time information of EEG signals.

CLSTM: This is the network structure model proposed in this paper, which is composed of CNN and LSTM. The system can well integrate the frequency, space and time information of EEG signals.

ACLSTM: Based on the CLSTM model, this model adds a channel attention mechanism module to assign different weights to different channels, and effectively utilizes the frequency, space and time information of EEG signals.

For the first two methods, the experiments are reproduced according to the parameters given in the original paper. The third method is a CNN model designed in this paper for comparative experiments. The last two methods are proposed in this paper. Table 2 below shows the average accuracy and variance for each method. It can be seen that the ACLSTM proposed in this paper achieves the highest classification accuracy.

3.3 Results

This section compares the classification performance of different length signal segments and the classification performance of different models by setting up comparative experiments. The following conclusions can be drawn, EEG signal segments of different lengths will determine how much emotional information they contain, but the performance difference is not large, which shows that our CLSTM model can make full use of EEG signal information to classify emotions, and it is not trusted. The effect of segment length. Furthermore, our proposed CLSTM and ACLSTM models achieve good classification performance compared with several other deep learning models, which proves that our model is also effective.

4 Conclusion

This paper proposes an ACLSTM model based on CNN-LSTM and channel attention mechanism, which can simultaneously consider the frequency, temporal and spatial information of EEG signals, and achieves good classification accuracy on the SEED public data set. The key points of this model lie in two parts: first, the 4-dimensional features of EEG signals are converted into 4-dimensional features with spatial information, which can effectively utilize the key information between different channels of EEG signals; second, the channel attention mechanism module is introduced to The EEG signals of different channels are assigned different weights to enhance the key emotional information in the EEG signals and suppress the useless information that interferes with the recognition effect.

References

1. Esposito R, Bortoletto M, Miniussi C. Integrating TMS, EEG, and MRI as an approach for studying brain connectivity [J]. Neuroscientist, 2020, 26(5-6): 471-486.
2. Liao L H C. Research and application of emotion recognition method based on physiological signal [D]. University of Electronic Science and Technology of China, 2020.
3. Yang Y L, Wu Q, Fu Y, et al. Continuous convolutional neural network with 3d input for EEG-based emotion recognition[C]. Proceedings of the International Conference on Neural Information Processing. Berlin: Springer, 2018: 433-443.
4. Kwon Y H, Shin S B, Kim S D. Electroencephalography based fusion two-dimensional (2D)-convolutional neural network model for emotion recognition system [J]. Sensors, 2018, 18(5): 1383.
5. Yang X L, Lin S Z. Method for multi-band image feature-level fusion based on the attention mechanism [J]. Journal of Xidian University, 2020, 47(01):120-127.
6. Wang Y, Huang ZY, McCane B, Neo P. EmotioNet: a 3-D convolutional neural network for EEG-based emotion recognition. In: 2018 international joint conference on neural networks.