Emotion Recognition of Single-electrode EEG based on Multi-feature Combination in Time-frequency Domain

Xiang Cao, Kunyuan Zhao and Dan Xu*
School of Information Science and Engineering, Yunnan University, Kunming, Yunnan, 650500, China
*Corresponding author’s e-mail: danxu@ynu.edu.cn

Abstract. Most researches on EEG emotion recognition are based on multi-electrode EEG signal acquisition equipment. There are few studies on single-electrode EEG equipment and the classification accuracy obtained is not ideal. In order to further improve the classification accuracy, the paper proposes an EEG emotion recognition method based on the combination of multiple features in the time-frequency domain. Firstly, the wavelet transform is used to decompose the EEG signals under the three types of emotion labels and extracts the time-frequency information of four frequency bands, including alpha, low beta, high beta, and gamma, and then uses different sizes of sliding time windows to calculate separately their statistical characteristics in the time-frequency domain. Finally, the Long Short-Term Memory (LSTM) network with time-frequency domain characteristics is used to extract the sequence information of the deep features and combine the output results of the softmax classifier. The results of experiments show that the average accuracy of the proposed method is 93.09% and 98.36% on emotion food and emotion state data set, respectively. Compared with traditional machine learning methods and other deep learning methods, the time-frequency LSTM model named TF-LSTM in this paper has better EEG generalization ability and classification performance, and also provides a new feasible scheme for emotion recognition research based on single-electrode EEG signal.

1. Introduction
Human emotions not only include physical and psychological reactions, but also embody people's own needs and subjective attitudes. They are a combination of feelings, thoughts and behaviours[1]. Recent studies indicated that the production or activity of human emotions is highly correlated with the activities of the cerebral cortex, which provides an important basis for the study of emotion classification through electroencephalograph (EEG). And with the continuous development of brain science, neuroscience and cognitive science, emotion classification and recognition based on EEG have become a hot topic in emotion studies, which is of great significance for the development of human-computer interaction, brain-computer interface, and artificial intelligence systems[2].

In recent years, EEG acquisition equipment has developed rapidly, the wireless head-mounted EEG acquisition devices that do not need to be coated with conductive glue have emerged gradually. Compared with professional multi-electrode medical acquisition equipment, this type of new equipment has become the primary choice for studying EEG emotion recognition due to portability and flexibility. However, the most important research content in the emotion recognition based on EEG is how to extract the emotional features better in the EEG data with time series characteristics, and how to design a classification model to accurately identify user's emotional state.
Therefore, in this paper, we propose a new single-electrode EEG emotion recognition method based on the combination of multiple features in the time-frequency domain. First, we decompose the single-channel EEG signal into four frequency bands (Alpha, Low Beta, High Beta and Gamma), and extract the time-frequency domain features that is highly correlated with the emotional dimension as input to the classifier. And then the feature sequence is formed in the time direction, and each statical feature represents a type of information of feature sequence. Finally, the long short-term memory network (LSTM) is used to extract the sequence information of the deep features, and the final emotion prediction is output through the softmax classifier.

2. Related work
Feature extraction and emotion recognition play an important role in the field of brain-computer interface (BCI). In the past, most researchers focused on traditional machine learning methods for emotion extraction. Murugappan et al.[3] used Discrete Wavelet Transform (DWT) to decompose the EEG signal into three frequency bands, and combined with the K-Nearest Neighbor (KNN), the accuracy reached 83.26%. Bhardwaj et al.[4] used Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) to identify seven emotions, with the average accuracy of 74.13% and 66.50%. In recent years, Tripathi et al.[5] used Convolutional Neural Networks (CNN) to classify emotions, the accuracy reached 75.58% and 73.28%. Li et al.[6] used Stacked Autoencoding (SAE) and Recurrent Neural Network (RNN) to recognize mixed physiological signals including EEG, and the accuracy reached 79.26%. But the above methods ignore that EEG itself is a highly non-stationary time series signal, and time series information is essential to improve the performance of feature extraction and subsequent classification.

In order to effectively extract timing information of EEG and obtain better classification results, we take advantage of the characteristics of LSTM based on the combination of multiple features in the time-frequency domain and build a new model named TF-LSTM to realize single-electrode EEG recognition.

3. Methods and materials
The research flowchart of this paper is divided into two parts (see Figure 1): the first is the feature combination of time-frequency domain. First, Wavelet Transform (WT) is used to decompose four frequency bands. Then use different sizes of time sliding windows to convert them in each time window. Each band is calculated for the mean, standard deviation, maximum, minimum and other characteristic attributes to obtain emotion food dataset. The emotion food data set is used to represent the time domain and frequency domain features of EEG; the second is emotion recognition model: TF-LSTM, the features we extracted are applied as the input of the LSTM, and finally passed a fully connected layer (FC) and a softmax layer to output model prediction results.

3.1. Time-frequency domain feature combination
In previous work, Zhao et al.[7] quantitatively analysed the characteristics of each frequency band by calculating the characteristic data of the four frequency bands in the time domain and frequency domain: the average value of emotion in each frequency band (except High Beta) is lower than neutral
and negative emotion. The biggest difference in brain wave amplitude is in the Low Beta. The standard deviation of the negative emotion band values (except Alpha) is larger than the neutral emotion and positive emotion.

According to the differences of emotions in four frequency bands, we use one second sliding time window to divide the EEG signal frequency bands of the three groups of emotion tags (neutral, positive and negative) into multiple signal segments overlapping by 50%, and then extract from each EEG frequency band characteristics. In order to obtain more EEG feature information as much as possible. The 1s time window is subdivided into 0.5s time window and 0.25s time window, and the time-frequency domain statistical characteristics between adjacent windows are extracted respectively. Finally, the features extracted in each frequency band are connected along the time direction to form an EEG feature sequence, which is used as the input of the classifier to output the classification result. The statistical characteristics of each time sliding window of the emotion Food dataset are: mean values, standard deviation values, maximum, minimum and the difference between them.

3.2. Emotion recognition model

3.2.1. Long Short-Term Memory

Recurrent Neural Networks (RNN) is widely used in time series signal processing in these years, but it has the problem of gradient explosion or gradient disappearance during the training process, which makes RNN unable to effectively use early historical information[8]. To solve this problem, Hochreite et al. introduced memory units into the internal structure of RNNs, and proposed LSTM to make it has better performance in processing and predicting important events with long intervals and delays in time series, and is more suitable for expressing the time domain information in EEG signals[9].

On the basis of RNN, LSTM introduces memory units and adds the forget gate, the input gate and output gate, which can control unit status of the memory and the degree of forgotten information of the previous moment and current moment, so it can storage long-term memory[10]. The structure of LSTM memory unit is shown in Figure 2. The general idea of LSTM is as follows:

a) Forget gate: it determines how much unit state \( C_{t-1} \) of previous moment is retained to unit state \( C_t \) of current moment. The equation is as follows:

\[
f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f)
\]

In equation (1), \( X_t \) is the input vector at time \( t \); \( h_{t-1} \) is the hidden state at the previous time; \( \sigma \) is the activation function; \( W_f \) represents weight matrix; \( b_f \) is an applied bias.

b) Input gate: it determines how much the input vector \( X_t \) at the current moment is retained to the current unit state \( C_t \). The equations of the decision vector \( I_t \) and the candidate information \( \bar{C}_t \) are:

\[
I_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i)
\]

\[
\bar{C}_t = \tanh(W_c \cdot [h_{t-1}, X_t] + b_c)
\]

\[
C_t = C_{t-1} \times f_t + C_t \times I_t
\]

where \( W_i \) and \( W_c \) are the input weight matrix; \( b_i \) and \( b_c \) represents the corresponding bias.

c) Output gate: it determines how much the cell state \( C_t \) at the current moment is input to the hidden state \( h_t \). The equations for the decision vector \( O_t \) and the hidden state \( h_t \) are:

\[
O_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o)
\]

\[
h_t = \tanh(C_t) \times O_t
\]
Where $W_o$ and $b_o$ are the weight matrix and bias, $\times$ represents the multiplication operation among the elements.

In this paper, TF-LSTM has two layers of LSTM, the number of hidden layer memory units is 256 and 128. All input vectors are subjected to max pooling operation, and then through a fully connected layer and a softmax layer to output classification results. Emotion recognition model of TF-LSTM is shown in Figure 3.

4. Experiments and result analysis

4.1. EEG data acquisition

Ten adults (5 males, 5 females, 23~26 years old, average age 24.7 years) participated in this study from the volunteers who participated in the previous emotional survey. Among them, 2 subjects were eliminated due to excessive noise and artifacts (such as touching the forehead or eyes), and construct the emotion food dataset.

In the experiment, three groups of emotionally labelled food pictures were established for each participant, namely neutral, positive and negative. Each participant is required to perform three sets of tests. The experimental platform (E-prime 2.0) will play pictures on the screen. In the first 10s, 5 unrelated test images for relaxation will appear on the screen. After the formal experiment beginning, 15 pictures of food will be played in sequence until the end as one trial. Figure 4 shows the food picture playback flowchart and Figure 5 shows some pictures of the food image data set.

4.2. EEG preprocessing and frequency band extraction

In the experiment, we use a wireless EEG headset (Mind Wave) with a single dry electrode to recode EEG data. The device collects the signal of the Fp1 electrode of the frontal lobe and the sampling frequency is 512 Hz. This equipment is different from the wet electrode acquisition method commonly used in traditional brain electrical signal measurement experiments. The subjects do not need to apply conductive gel, it has the characteristics of low contact impedance, convenience and comfort, and can achieve long-term brain electrical signal acquisition and testing.
We use the Stationary Wavelet Transform (SWT) to denoise raw EEG, and then adopts Wavelet Transform (WT) to decompose the preprocessed EEG signal and extract four frequency bands: Alpha (8~ 16 Hz), Low Beta (16~24 Hz), High Beta (24~32Hz) and Gamma (32~40 Hz).

4.3 Experimental environment and settings
The experimental platform used in this paper is NVIDIA RTX 1060, the system disk is 512GB NVME solid state hard disk, and the storage disk is 6GB. The experiment was implemented by Python language programming, the algorithm in this paper was realized by combining Tensorflow2.0.0 deep learning framework. The loss function of the emotion recognition model uses the cross-entropy to judge parameter optimization. It mainly describes the distance between the probability of actual output and the probability of expected output. The cross-entropy loss function equation is:

\[
\text{loss} = - \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n} y_{ij} \log(p(x_{ij}))
\]  

In equation (7), \(m\) represents the number of samples, \(n\) represents the number of different categories to which the samples belong, \(y_{ij}\) represents the category \(j\) to which the sample \(i\) belongs, and \(p(x_{ij})\) represents the predicted probability that the sample \(i\) belongs to the category \(j\).

The experimental parameter optimizer in this paper uses Adam optimizer[11]. In model training, the learning rate is 0.00015, the epoch is 500 and the batch size is 64.

4.4 Experiment results
In order to verify the effectiveness of TF-LSTM emotion recognition model, four quantitative model evaluation criteria are used to evaluate the performance of the model. They are classification accuracy, Precision, Recall and F1-score[12].

|       | S1   | S2   | S3   | S4   | S5   | S6   | S7   | S8   | Average |
|-------|------|------|------|------|------|------|------|------|---------|
| LDA   | 0.6138 | 0.5357 | 0.3962 | 0.5360 | 0.4081 | 0.6245 | 0.6302 | 0.4595 | 0.5255  |
| SVM   | 0.5041 | 0.6190 | 0.4764 | 0.5631 | 0.4332 | 0.7424 | 0.6679 | 0.4903 | 0.5621  |
| KNN   | 0.6057 | 0.5714 | 0.6651 | 0.6351 | 0.6171 | 0.7729 | 0.7170 | 0.5907 | 0.6469  |
| NU-SVM| 0.6626 | 0.7262 | 0.6557 | 0.6164 | 0.6196 | 0.7729 | 0.7208 | 0.6641 | 0.6798  |
| RF    | 0.9024 | 0.9048 | 0.8774 | 0.8789 | 0.9043 | 0.9327 | 0.9170 | 0.8764 | 0.8992  |
| TF-LSTM | 0.9146 | 0.9114 | 0.9387 | 0.9414 | 0.9326 | 0.9476 | 0.9245 | 0.9369 | 0.9309  |

Table 1 shows the comparison of experimental results of 5 traditional machine learning methods and TF-LSTM on emotion food dataset and 8 subjects(S1-S8). It can be obviously seen that the classification accuracy and performance of Random Forest (RF) is significantly higher than Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Nuclear Support Vector Machine (NU-SVM), with the average accuracy of 89.92%. Due to the non-steady state characteristics of EEG signals and the differences between subjects, the classification accuracy of RF on subject S3, subject S4, and subject S8 are only 87.87%, 87.89% and 87.64%, indicating that the method has weak ability to extract EEG features.

As shown in Figure 6, we can see the Precision, Recall and F1-score of subject S3, subject S4, and subject S8 perform poorly compared with other subjects. The accuracy of these criteria is less than 84%, and there is an imbalance between subjects, which indicates that EEG generalization ability and representation ability of random forest are insufficient.
As shown in Figure 7. It can be seen that the classification performance of TF-LSTM network model on the subject S3 is better than random forest, and the accuracy of the three evaluation criteria of the model all exceed 90%. In addition, the average Precision, average Recall and average F1-score increased by 4.62%, 4.43% and 4.48%, respectively. Although there are individual differences between subjects, the evaluation indicators are all maintained between 0.88-0.94. Subject S3, subject S4, and subject S8 have all performances more stable than RF, indicating that TF-LSTM has better generalization ability.

In order to further verify the effectiveness of TF-LSTM, we apply it into Emotion State dataset and combine with the four algorithms of Bird et al.: Deeply evolved multilayer perceptron (Devo MLP), single-layer LSTM, AB (Devo-MLP) and AB (LSTM) for comparison[13]. Among them, AB (Devo-MLP) and AB (LSTM) are using Adaptive Boosted algorithms to change the data distribution to prevent over-fitting in MLP and single-layer LSTM. The emotion state dataset was based on whether a person was feeling positive, neutral and negative emotions. 6 minutes for each state were recorded from two adults (1 male and 1 female aged 21 ± 1), producing a total of 36 minutes of brainwave activity data. The experimental results are shown in Table 2.

| Method         | Accuracy |
|---------------|----------|
| Devo MLP       | 96.11    |
| LSTM           | 96.86    |
| AB(Devo MLP)   | 96.23    |
| AB(LSTM)       | 97.06    |
| TF-LSTM        | **98.36**|

Despite Bird et al. performed well in the overall recognition results with 4 methods, and alleviated the over-fitting to a certain extent after adopting the adaptive enhancement method, but the classification effect still depends on the choice of classifiers. Consequently, the accuracy of the improvement is not obvious, indicating that the method does not consider the important influence of deep information such as the time-frequency domain characteristics of the EEG signal. In this paper, we consider the different combination of EEG features in time and frequency domain, combined with the powerful feature extraction ability of LSTM for time series information, and build a new model named TF-LSTM. The recognition accuracy of three kinds of emotions reaches 98.36%, which indicates that the method has better performance in EEG classification ability and EEG representation ability.
5. Conclusion
Aiming at the problems of low accuracy of EEG emotion recognition and poor generalization ability, we propose a new EEG emotion recognition model called TF-LSTM. The method adopts different time sliding windows to extract EEG signals, and combine the time-frequency domain statistical characteristics of the four frequency bands with the LSTM to obtain EEG sequence features for emotion recognition. The results of experiments show that the method has an obvious improvement in classification accuracy, precision, recall and F1-score. At the same time, through similar research and comparison experiments, it can be concluded that by combining EEG features in the time and frequency domains, LSTM can automatically learn the time and frequency domain feature that has the highest correlation with the emotional dimension, and the classification model has better generalization capabilities and robustness.

Acknowledgments
The authors thank the editor and reviewers for their valuable comments which help to improve this paper. This work was partly supported by the Yunling Scholars Program of Yunnan Ten Thousand Talents Program (YNWR-YLXZ-2018-022), partly supported by the "Double First Class" Construction Joint Fund Project of Yunnan University (No. 2019FY003012), and the National Natural Science Foundation Funding from the Science Foundation Project (No. 61761046, No. 61540062).

References
[1] Chen J X, Zheng R, Jia X Y, et al. (2019) EEG emotion classification based on co-space mode and wavelet packet decomposition. Computer Engineering and Applications, 055(001):149-153.
[2] Jiang JF, Zeng Y, Lin Z M, et al. (2016) Review of Emotion Evaluation Research Based on EEG Signal[J]. Journal of Information Engineering University, 17(006):686-693.
[3] Murugappan M, Ramachandran N, Sazali Y. (2010) Classification of human emotion from EEG using discrete wavelet transform. Journal of Biomedical Science & Engineering, 03(04):390-396.
[4] Bhardwaj A, Gupta A, Jain P, et al. (2015) Classification of human emotions from EEG signals using SVM and LDA Classifiers. In: 2nd International Conference on Signal Processing and Integrated Networks. Noida.pp:180-185.
[5] Tripathi S, Acharya S, Sharma R D, et al. (2017) Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset. In: Twenty- Ninth IAAI Conference, pp: 4746-4752.
[6] Li Y J, Huang J J, Wang H Y, et al. (2017) Research on multimodal physiological signal fusion and emotion recognition based on SAE and LSTM RNN. Journal on Communications, 38(012):109-120.
[7] Zhao K, Xu D. (2019) Food Image-Induced Discrete Emotion Recognition Using a Single-Channel Scalp-EEG Recording.2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI).
[8] Zhou J, Yang G Y, Xu T. (2019) Multi-class motor imaging EEG classification based on spatial frequency and time series information. Chinese Journal of Medical Physics, 036(001):81-87.
[9] Hochreiter S, Schmidhuber J. Long Short-Term Memory. (1997) Neural Computation, 9(8):1735-1780.
[10] Gers F A, Schmidhuber, Jürgen and Cummins F. (2000) Learning to Forget: Continual Prediction with LSTM. Neural Computation, pp.2451-2471
[11] LeCun Y, Bengio Y and Hinton G. (2015). Deep learning. Nature 521, 436–444.
[12] Li C, Tang Y, Li C. (2011) A GERT-based analytical method for remanufacturing process routing.
[13] Bird J J, Faria, D R, Manso, L J, Ekárt A and Buckingham CD. (2019) A Deep Evolutionary Approach to Bioinspired Classifier Optimisation for Brain-Machine Interaction. Complexity.