Multi feature fusion EEG emotion recognition

1st Guo Guodong
Medical Instrumentation College
University of Shanghai
for Science and Technology
Shanghai
156896284@qq.com

2nd Gao Yahan*
Medical Instrumentation College
Shanghai University of
Medicine Health Sciences
Shanghai
gaoyh@sumhs.edu.cn

2021 7th International Conference on Big Data and Information Analytics (BigDIA) | 978-1-6654-2466-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/BIGDIA53151.2021.9619674

***Abstract*—In recent years, the research on emotion recognition of EEG signals has attracted much attention. It is an important task to realize the advanced stage of artificial intelligence. How to realize real-time and efficient human-computer interaction has become an important direction of EEG signal research. This study aims to improve the accuracy of emotion recognition of EEG signals, and proposes a binary classification and emotion EEG recognition method based on feature fusion is carried out after multi feature extraction to improve the recognition rate. For the preprocessed EEG signals, the eigenvalues extracted from time-frequency, spatial domain, nonlinear dynamics and convolution neural network are used as the initial eigenvectors, and the dimension is reduced by principal component analysis. Finally, the long short-term memory neural network is used for classification. The emotion recognition experiment was carried out on the EEG emotion data set deap. The accuracy of two classification in pleasure and arousal was 84.42% and 85.61% respectively. The recognition rate is higher than that under single feature and other combined features. The experimental results show that compared with single feature extraction, multi feature fusion has better characteristics of emotional EEG signals, and high classification accuracy can be achieved by using the long short-term memory neural network. The performance of the emotion recognition method of EEG signals proposed in this paper is better than other methods based on traditional artificial design features and SVM or DBM, It is verified that the method proposed in this paper is feasible.**

Index Terms*—*Multi feature fusion, EEG, Feature extraction, LSTM, Emotion recognition

I. Introduction

Emotion is an internal subjective experience, which is people's physiological and psychological response to external factors or self stimulation [1]. Emotion recognition is of great significance in the fields of medical treatment, education, criminal trial and so on. The key to realize emotion recognition of EEG signals is to extract the emotion related features of signals and optimize the recognition of classifiers. Candar [2] et al. Extracted wavelet coefficients, wavelet entropy and different frequency range index was extracted from the DEAP data set as EEG features, and SVM classifier was used to classify and recognize the Valence and Arousal.The average accuracy was 76.80% and 74.30% respectively. Liu [3] et al. Extracted EEG signal features on DEAP data set, fused high-dimensional features by using maximum correlation, minimum redundancy and principal component analysis, used SVM for classification, and carried out classification and recognition on valence and arousal, obtaining 72.4% and 76.1% accuracy respectively. Tripathi [4] et al. Extracted statistical features such as median, mean, variance and kurtosis from DEAP EEG data set, and used Convolutional Neural Networks (CNN) for emotion classification and recognition, with the highest accuracy of 81.40% and 73.36% respectively. Chao Hao [5] et al. Extracted the initial features of emotion in DEAP EEG data set, expressed the initial features in high-level abstraction through DBN, and then used RBM to realize emotion classification. The accuracy of emotion recognition valence and arousal emotion models reached 84.42% respectively 85.61%.

In order to further improve the emotion recognition accuracy of EEG based on dimensional emotion model, and avoid the insufficient extracted EEG eigenvalues and excessive dependence on manual feature extraction, this paper proposes to extract the time-frequency domain features of EEG signals through Hilbert Huang transform, select fuzzy entropy as nonlinear dynamic features, and combine the time-frequency domain the nonlinear eigenvector and the eigenvalue extracted by the convolution neural network are used as the initial eigenvector, the dimension is reduced by principal component analysis, and then sent to the long short-term memory neural network for classification to obtain the simulation accuracy of the algorithm, as shown in Fig.1.

Foundation:Scientific research and innovation project of Shanghai Municipal Commission of Education (14YZ168)

Natural Science Foundation of Shanghai (21ZR1428300),

Shanghai Science and Technology Innovation Action Plan Biomedicine Infrastructure Program (20S31905100).

II. Experimental data

A. DEAP data set and preprocessing

DEAP data set [6] recorded electroencephalogram (EEG) and peripheral physiological signals of 32 subjects, and each person watched 40 1-minute music video clips. In a single experiment, the EEG signals of subjects were collected for 63s. The first 3S was the baseline signal (not stimulated), and the last 60s was the EEG signal when watching audio. In order to reduce the impact of individual differences on emotion recognition, the relevant eigenvalues of the baseline signal

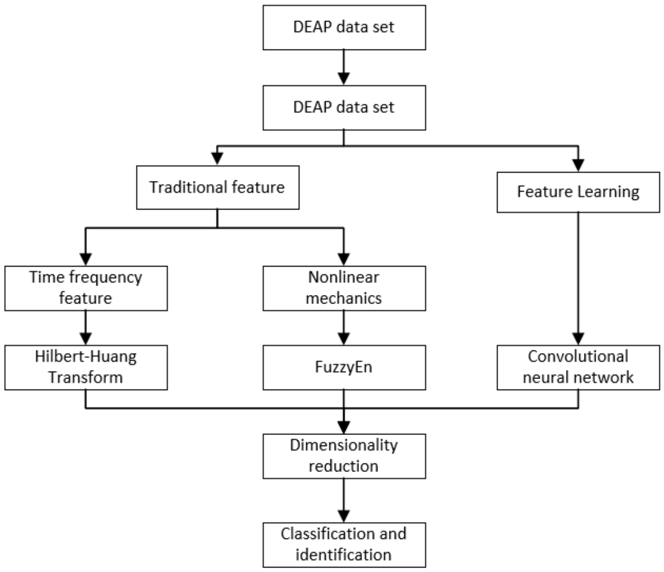

Fig. 1. Algorithm flow chart.

were successively subtracted from the emotion induction data of 60 seconds. The data format is 40×40×8064, including 40 music clips, 32 lead EEG signals and 8 other physiological signals. Each data segment contains 8064 data points. The baseline EEG signal is $40 \times 32 \times 384$, the emotional induction data was $40 \times 32 \times 7680$The locations of 32 EEG channels collected in DEAP data set are selected according to the international 10-20 system, as shown in Fig.2.

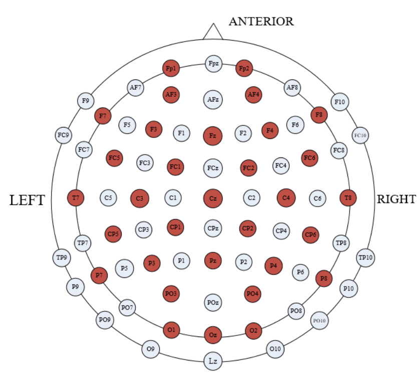

Fig. 2. 10-20 system.

B. Emotional label

The label of DEAP data set is that after watching each video, participants score music video clips by $1 \sim 9$ points in three dimensions: valence, arousal and dominance. The two-dimensional model of V-A dimension emotion composed of arousal and valence[7] is used as an index to analyze the specific emotional situation, describe the positive and negative and strength of emotion, generate emotion labels, and study the emotion recognition model. The second category label is to divide a single dimension, set 5 points as the threshold, and divide a single dimension into two labels: high and low. The specific division is shown in Table 1.

TABLE I
THE LABEL OF BINARY CLASSIFICATION

Category	Emotional label	Quantity
Type I	Hight Valence	944
	Low Valence	756
Type II	Hight Arousal	964
	Low Arousal	736

III. FEATURE EXTRACTION

High quality feature extraction is the basis of emotion recognition of EEG signals. This paper, Hilbert Huang transform (HHT) is used to extract the time-frequency domain features of EEG signals, fuzzy entropy is selected as the nonlinear dynamic features, and combined with one-dimensional convolutional neural network to automatically extract the emotional features of EEG signals for multi feature fusion.

A. Hilbert-Huang Transform

When extracting time-frequency features, HHT is used. Firstly, the preprocessed EEG signal is decomposed into several Inherent Mode Functions (IMF) through EmpiricalMode Decomposition (EMD) HHT mainly decomposes the preprocessed EEG signal into several IMF through EMD, and then Hilbert Transform the IMF to obtain the corresponding energy spectrum and marginal spectrum.

- The maximum and minimum of the original signal $x(t)$ are found by cubic spline interpolation, make the envelope curve of the maximum and minimum values, and calculate its average value

$$m_t\left(t\right) = \frac{U\left(t\right) + L\left(t\right)}{2} \tag{1}$$

- Whererepresents the curve function of amplitude varying with time, and the difference between the original signal $x(t)$ and $m(t)$ is taken as the first component of IMF

$$h_1 = x\left(t\right) - m\left(t\right) \tag{2}$$

- Repeat operation
- Hilbert Transform each IMF component to obtain the corresponding Hilbert spectrum. Finally, the Hilbert spectrum of all IMF components is superimposed to obtain the Hilbert spectrum of the original signal [11]. Hilbert Transform is:

B. Fuzzy Entropy

In the two algorithms of Approximate Entropy of eigenvalues of nonlinear dynamics and Sample Entropy, a threshold is added in the evaluation: if it is greater than the threshold, it will be chaotic, and if it is less than, it will not be chaotic. Fuzzy Entropy [12] proposed an improved algorithm for the discontinuous entropy of EEG signals, adding the idea of fuzzy to make the evaluation more scientific. Using fuzzy entropy as nonlinear dynamics, on the premise of having the

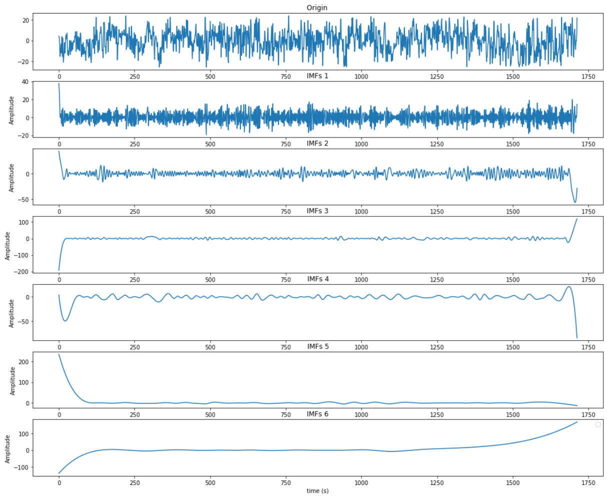

Fig. 3. IMF of EEG signals.

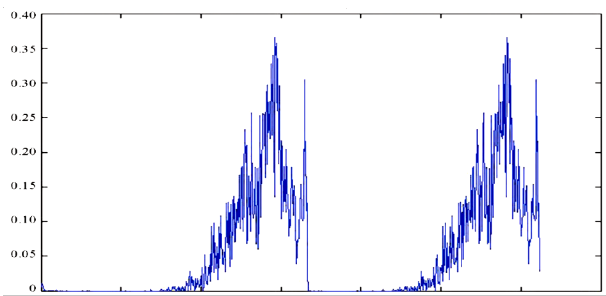

Fig. 4. HHT Marginal Spectrum.

advantages of sample entropy, it can effectively reduce the dependence on the length of time series signal, and has good continuity and robustness.Processing steps of fuzzy entropy on the preprocessed time series $\{x(i), i = 1, 2, ..., n\}$:

- Given time series $\{x(i), i = 1, 2, ..., n\}$,the dimension of phase space is defined as $m\,(m < N - 2)$ and rreconstructed phase space

$$X\,(i) = [u(i), u(i+1), ..., u(i+m-1)] - u_0\,(i) \quad (3)$$

- Introducing fuzzy relation functionand calculate:

$$F_{ij}^{m} = exp\left[-ln(2) \times (\frac{d_{ij}^{m}}{r})^2\right] \quad (4)$$

- Obtain $C_i^m(r)$ and definition:

$$\varnothing^m(r) = \frac{1}{N-u+1}\sum_{i=1}^{N-u+1} C_i^m(r) \quad (5)$$

- Fuzzy Entropy can be expressed as :

$$FuzzyEn\,(t) = ln\varnothing^m(r) - ln\varnothing^{m+1}(r) \quad (6)$$

C. Single channel Emotion Feature Extraction Based on CNN

After preprocessing the collected EEG signals, one-dimensional CNN [13] is used to extract emotional features for different channels. The structure of one-dimensional CNN is shown in Fig.5, including 3 convolution layers, 2 maximum pooling layers and 1 Average pooling layer. The convolution layer convolutes the local perception area in the previous layer, and each convolution layer uses 1×5, and the step size of the convolution kernel is set to 1. After the convolution layer,Rectified Linear Unit (ReLU) activation function [14] is used to add nonlinear factors, so that the output of some neurons in the trained network is 0, with moderate sparsity, accelerate the convergence of the network, reduce the mutual dependence of parameters, avoid the over fitting problem of the model, and improve the generalization ability of the model. The output of the average pooling layer is characterized by the high-level characteristics of a single channel.

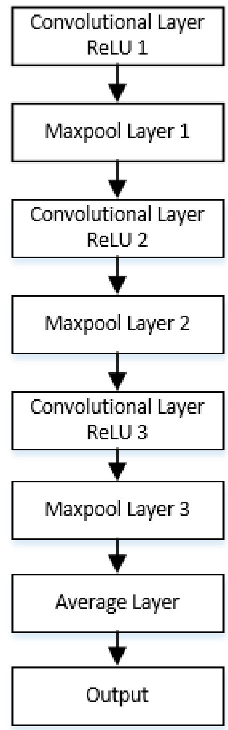

Fig. 5. Structure diagram of one-dimensional CNN.

D. Feature fusion

The EEG signal in this paper integrates the time-frequency domain, nonlinear features and the eigenvalues extracted by convolution neural network as the initial eigenvectors. The obtained eigenvectors are merged and the input eigenvectors $F' = \{F1, F2, F3\}$ are constructed. Because different features have different meanings, and high-dimensional feature vectors are easy to cause "dimension disaster" [15], which is not conducive to subsequent classification and recognition. Therefore, it is necessary to process the input feature vectors by dimensionality reduction to eliminate the feature vectors with small contribution to classification. Set the feature vector contribution rate to 85%, and reduce the dimension of the feature vector PCA [16] to obtain a 16 dimensional new feature vector.

IV. LASSIFICATION AND IDENTIFICATION

LSTM is an improved time cycle neural network model [17], with special gate and memory structure, as shown in the figure below. LSTM network is composed of sequence input layer and LSTM layer. The sequence input layer can input sequence or time series data into the network. LSTM layer can learn the long-term dependence between time steps of sequence data, and can better process and predict the signals with long interval and delay in time seriess, and solves the problem of gradient disappearance of RNN. A typical LSTM cell structure [18] is shown in Fig.6.

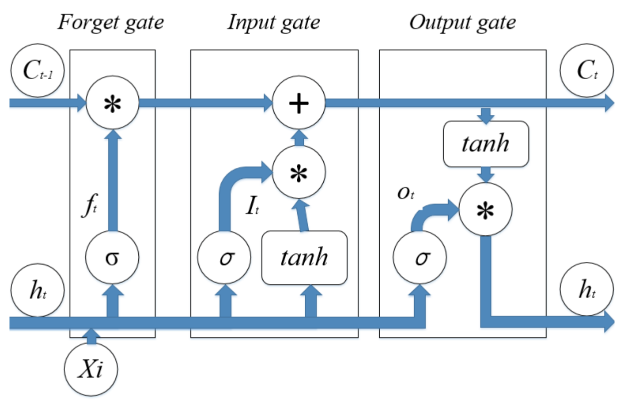

Fig. 6. LSTM unit structure diagram.

The feature sequences of EEG signals are connected along time and still belong to time series. Therefore, for each feature sequence, the dynamic information of different features can still be learned by using LSTM units respectively.

In the model, the classifier is composed of LSTM units and fully connected neural network. The first layer is composed of multiple LSTM units with the same structure, in which each LSTM unit is only responsible for processing one feature sequence, so that each LSTM unit can learn dynamic information from one feature sequence. The corresponding feature sequence is represented by $L(n, m, t)$, Represents the value of the m-th characteristic sequence of the n-th channel at time t. The second layer is the fully connected neural network (Relu as the activation function), and its input is the linear combination of the outputs of all LSTM units in the last time step, The function of this layer is to combine different feature dynamic information learned from each LSTM unit. Finally, the output of this layer is input to the soft max layer to predict the corresponding emotional state.

V. EXPERIMENTAL RESULTS AND ANALYSIS

In this study, HHT is used to extract the time-frequency domain features of EEG signals, fuzzy entropy is selected as the nonlinear dynamic features, and combined with one-dimensional convolutional neural network to automatically extract the emotional features of EEG signals for multi feature fusion. The LSTM is used for classification. The test recorded the classification accuracy of each subject in the two dimensions of valence and arousal.

A. Data processing

The experiment was conducted separately for each tester, and there were 32 groups of preprocessed data in each time period in the database. Each group of data contains 8064 data points. EEG signals include baseline signals and emotion induction data. If the above 32 groups of preprocessed data are directly tested as 32 samples, the trained model will be unstable because the number of samples is too small. Therefore, new training samples and test samples are generated based on the 32 groups of preprocessed data. The first 24 groups of preprocessed data in each time period in the database are processed as training samples. The last 8 groups of preprocessed data were processed as test samples.

B. Experimental results

Through many experiments, the highest accuracy of the model is compared with other advanced research results. Each experiment adopts the same evaluation criteria. The highest accuracy of various models is shown in Table 2[2-7]. The LSTM emotion recognition model based on multi feature fusion proposed in this paper has greatly improved the accuracy.The accuracy of two classification in valence and arousal reached 84.42% and 85.61% respectively.

TABLE II
ACCURACY OF EACH EMOTION RECOGNITION MODEL

Model	accuracy(Valence)	accuracy(Arousal)
Candar	76.80%	74.30%
Liu et al	72.40%	76.10%
Mert and Akan	72.87%	75.00%
Li	75.54%	75.12%
Tripathi	81.40%	73.36%
Chao Hao	84.42%	85.61%
This paper	84.66%	86.62%

Through comparison, it can be seen that the model has good results in the calculation accuracy of emotion recognition. The structure of the model designed in this paper is relatively complex, and there are many parameters, which leads to high operation cost. The next step is to optimize the lead by using the attention mechanism, and adopt the new model of model migration to reduce the amount of parameter training to improve the learning efficiency. In the subsequent experiments, we can further study and analyze the impact of increasing the number of convolution layers to mine deeper channel features on the accuracy. In addition, the research on multimodal emotion recognition by combining human physiological signals with external emotion representation carriers such as expression, speech, behavior and posture also has very important practical significance to further improve the accuracy of emotion recognition.

ACKNOWLEDGMENT

The author would like to thank Ms. Gao Yahan. During the time of learning, I have experienced and gained a lot, both in academic research and personal mind. I want to thank the laboratory teachers and students who have helped me along the way.Thank Mr. Gao Yahan for creating a good scientific research environment and hardware resources for us. Mr. Gao Yahan is rigorous, knowledgeable, approachable and realistic. He sacrificed his rest time to communicate with me and help me solve problems.

Thank Jiang Qiang and Zhang Yining for their gentle help. In the stumbling process of scientific research, your company and encouragement have added many beautiful memories to scientific research life. Thanks to all the students who provided help for the data.

REFERENCES

[1] Christian, Mühl, Brendan, Allison, Anton, and Nijholt. "A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges," Brain-Computer Interfaces, vol. 1, pp. 66-84,2014.

[2] Candra, H. and Yuwono, M. and Chai, R. and Nguyen, H. T. and Su, S, "EEG emotion recognition using reduced channel wavelet entropy and average wavelet coefficient features with normal Mutual Information method," International Conference of the IEEE Engineering in Medicine Biology Society. IEEE, 2017.

[3] Liu, W. and Zheng, W. L. and Lu, B. L. "Multimodal Emotion Recognition Using Multimodal Deep Learning," International Conference on Neural Information Processing. Springer, Cham, 2016.

[4] Tripathi, Samarth and Acharya, Shrinivas and Sharma, Ranti Dev and Mittal, Sudhanshi and Bhattacharya, Samit, "Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset". 2017.

[5] Zhao, H. Li, Y. l. Lian, W, "Multi analysis domain feature fusion based on integrated deep learning model in EEG emotion recognition," Control and Decision, pp. 1674-1680, 2020.

[6] Koelstra S, "DEAP: A Database for Emotion Analysis ;Using Physiological Signals," IEEE Transactions on Affective Computing, vol. 3, pp. 18-31, 2012.

[7] MURRAY. I. R, ARNOTT. J. L, "Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion," The Journal of the Acoustical Society of America, 1993, pp. 1097-1108.

[8] Sharabaty. H , Jammes. B , D. Esteve, "EEG Analysis Using HHT: One Step Toward Automatic Drowsiness Scoring," IEEE 22nd International Conference on Advanced Information Networking and Applications. IEEE, 2008.

[9] Ji, Ma, Dong, Zhang, "EEG Signals Feature Extraction Based on DWT and EMD Combined with Approximate Entropy," Brain sciences, vol. 9, pp. 201, 2019.

[10] Kai. F, Qu. J, Yi. C, "Classification of seizure based on the time-frequency image of EEG signals using HHT and SVM," Biomedical Signal Processing Control, vol. 13, pp. 15-22, 2014.

[11] S. Huiyun, F. Yunfa, X. Xin, "Research on EEG pattern recognition based on HHT motor imagination," Acta Automatica Sinica, vol.9, pp. 1686-1692, 2015.

[12] Y. Muhan, C. Feng, "EEG classification based on HCSP and fuzzy entropy," Computer Engineering and Design, vol. 39, pp. 557-562, 2018.

[13] S. Tripathi, S. Acharya, R. D. Sharma, "Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset," Twenty-Ninth IAAI Conference, pp. 4746-4752, 2017.

[14] L. Wenhui, D. Baotian, W. Zesheng, "Short-term Traffic Flow Prediction Based on CNN-SVR Hybrid Deep Learning Model," Transportation System Engineering and Informa-tion, vol. 17, pp. 68-74, 2017.

[15] L. R. Quitadamo ,F. Cavrini, L. Sbernini, "Support vector machines to detect physiological patterns for EEG and EMG-based human-computer interaction: a review," Journal of Neural Engineering, vol. 14, pp. 11001, 2017.

[16] L. Liu, W. Li, H. Jia, "Method of Time Series Similarity Measurement Based on Dynamic Time Warping," Computers, Materials amp, vol. 57, pp. 97-106, 2018:

[17] S. Hochreiter, J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, pp. 1735-1780, 1997.

[18] Y. Yong, S. Xiaosheng, H. Changhua, Z. Jianxun, "A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures," Neural Computation, vol. 31, pp. 1235-1270, 2019.