EEG-Based Emotion Recognition with Similarity Learning Network

Yixin Wang[1,2], Shuang Qiu[1], Jinpeng Li[1,2], Xuelin Ma[1,2], Zhiyue Liang[4], Hui Li[4], Huiguang He[1,2,3,*]

***Abstract*— Emotion recognition is an important field of research in Affective Computing (AC), and the EEG signal is one of useful signals in detecting and evaluating emotion. With the development of the deep learning, the neural network is widely used in constructing the EEG-based emotion recognition model. In this paper, we propose an effective similarity learning network, on the basis of a bidirectional long short term memory (BLSTM) network. The pairwise constrain loss will help to learn a more discriminative embedding feature space, combined with the traditional supervised classification loss function. The experiment result demonstrates that the pairwise constrain loss can significantly improve the emotion classification performance. In addition, our method outperforms the state-of-the-art emotion classification approaches in the benchmark EEG emotion dataset–SEED dataset, which get a mean accuracy of 94.62%.**

I. INTRODUCTION

Recently, Affective Computing (AC) has been attracting more and more attention, which helps to make connection between the human and the computer by developing computational systems that can recognize and react to human emotions [1]. The ultimate purpose of affective computing is to make the artifact machines more ’sympathetic’ in the human machine interaction [2]. Emotion recognition is an important part of affective computing, which is the process of identifying human emotion.

Various measures have been used to recognize emotion states. On the one hand, several non-physiological signals, for example, facial expression [3], body gesture [4] and voice signal [5], are widely used for emotion recognition. On the other hand, they are physiological signals [6], including electroencephalography (EEG), electromyogram (EMG) and so forth. Because of the high accuracy, high temporal resolution, and quite objective assessment, EEG has shown a greater potential in evaluating emotion states. Many psychophysiology studies that human emotions can be reflected by EEG signal. There are five bands that always be mentioned: Delta, Theta, Alpha, Beta, and Gamma [7]. And then we can extract EEG features from each band to distinguish the emotional processes. However, EEG signal is non-stationary, it may be caused by variations of user’s physiological activities, including different stimulus, increasing fatigue and varying electrode impedances, etc. Thus the problem of classification for EEG-based emotion recognition is still seen as a challenge.

There are many feature extraction techniques to characterize the EEG signals, including the differential entropy (DE) [8] feature, the power spectral density (PSD) [9] feature and so on. Nowadays, numerous classifier have been implemented for EEG-based emotion recognition, such as the Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Linear Regression (LR). For example, Zheng [10] proposed selecting 12 channel electrodes features in SVM, which provided 86.65% on average. Zheng [11] proposed a discriminative Graph regularized Extreme Learning Machine (GELM) which obtained a mean accuracy of 91.07%. Furthermore, deep learning becomes leader in the field of the machine learning recently [12], there are many related methods being used in the EEG-based emotion recognition task, including deep belief networks (DBNs), convolutional neural networks (CNNs) [13], recurrent neural networks (RNNs) [14] and so on. For instance, Li [15] organized DE features extracted from 62 channels as 2-D maps to train the hierarchical convolutional neural network (HCNN) and achieved 88.2% at the Gamma wave band. Zhang [16] used a unified deep network framework called spatial-temporal RNN (STRNN) with DE features, which can get the accuracy of 89.5%.

Typical BLSTM models can utilize long-range context for the current prediction. In our paper, we take account into that the EEG signals contain the temporal information, and take advantage of the Bidirectional Long Short-Term Memory (BLSTM) [17] framework to process it. We also innovatively import the pairwise constrain loss [18] which is a useful metric learning approach to change the distribution

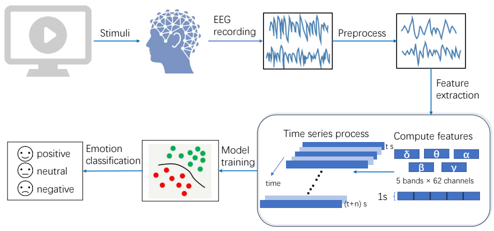

Fig. 1. The flow chart of EEG emotion classification with similarity learning network.

*This work was supported in part by the National Natural Science Foundation of China under Grant 81701785, 91520202, in part by the Strategic Priority Research Program of CAS, Grant No.XDB32040000, in part by the CAS Scientific Equipment Development Project under Grant YJKYYQ20170050, in part by the Beijing Municipal Science and Technology Commission under Grant Z181100008918010 and in part by the Youth Innovation Promotion Association CAS.

[1] Research Center for Brain-inspired Intelligence, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Science, Beijing, China

[2] University of Chinese Academy of Sciences, Beijing, China

[3] Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Science, Beijing, China

[4] Department of Educational technology, Capital Normal University, Beijing, China

*Corresponding author `huiguang.he@ia.ac.cn`

978-1-5386-1311-5/19/$31.00 ©2019 IEEE

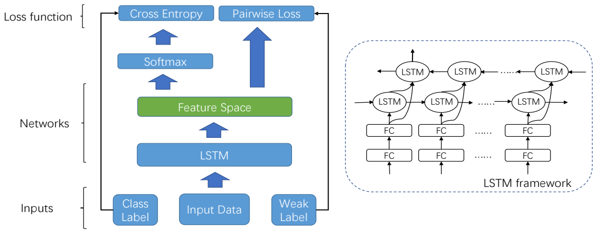

Fig. 2. The framework of our similarity learning network.

of the EEG signal. The basic idea of pairwise loss is to minimize the distance in a feature space for similar pairs, and maximize the distance for dissimilar pairs with the use of weak labels [19]. The loss will help to learn a more discriminative embedding feature space, combined with the traditional supervised classification loss function. The experiment results demonstrates that the pairwise constrain loss can significantly improve the emotion classification performance, and our method outperforms the state-of-the-art emotion classification approaches based on EEG.

II. THE PROPOSED METHOD

A. LSTM

Recurrent Neural Networks (RNNs) are neural networks adapted for sequence data $(x_1, ..., x_T)$. At each time step $t \in \{1, ..., T\}$, the hidden state vector h_t is updated by the equation $h_t = \sigma(Wx_t + Uh_{t-1})$ in which x_t is the input at time t. W is the weight matrix from inputs to the hidden-state vector and U is the weight matrix on the hidden-state vector from the previous time step h_{t-1}. In this equation and below the logistic function is denoted by $\sigma(x) = (1 + e^{-x})^{-1}$.

The Long Short-Term Memory (LSTM) variant of RNNs in particular has success in tasks related to EEG signals analysis. A LSTM is parametrized by weight matrices from the input and the previous state for each of the gates, in addition to the memory cell. We use the standard formulation of LSTMs with the logistic function (σ) on the gates and the hyperbolic tangent ($\tanh$) on the activations. In the equations (1) below, $\circ$ denotes the Hadamard (elementwise) product.

$$\begin{aligned}
i_t &= \sigma(W_i x_t + U_i h_{t-1}) \\
f_t &= \sigma(W_f x_t + U_f h_{t-1}) \\
o_t &= \sigma(W_o x_t + U_o h_{t-1}) \\
\tilde{c}_t &= \tanh(W_c x_t + U_c h_{t-1}) \\
c_t &= i_t \circ \tilde{c}_t + f_t \circ \tilde{c}_{t-1} \\
i_t &= o_t \circ \tanh(c_t)
\end{aligned} \tag{1}$$

Bidirectional RNNs [20] incorporate both future and past context by running the reverse of the input through a separate RNN. The output of the combined model at each time step is simply the concatenation of the outputs from the forward and backward networks.

B. Pairwise-constrain loss

The proposed network contains two layers of Bidirectional LSTM nodes. The activations at last timestep of the final BLSTM layer are picked to produce the output. The whole framework is shown in Fig 2.

Let $f_W(x_1)$ and $f_W(x_2)$ be the projections of x_1 and x_2 in the embedding space computed by the network function f_W, there we choose BLSTM as framework to utilize the time series information. We define the energy of the model E_W to be the Euclidean distance between the embeddings of x_1 and x_2:

$$E_W(x_1, x_2) = \|f_W(x_1) - f_W(x_2)\|^2 \tag{2}$$

To turn the distance into a cost, we define that if x_1 and x_2 come from a similar pair, the cost will be plain Euclidean distance; otherwise, it will be the hinge loss (still using the Euclidean distance). The weak labels indicate the pairwise relationship, which is converted from the class label. If a pair has the same class label, then it is a similar pair, otherwise it is dissimilar.

$$\begin{aligned}
loss_{pair} = {} & I_s(x_1, x_2) E_w(x_1, x_2) + \\
& I_{ds}(x_1, x_2) max(0, c - E_W(x_1, x_2))
\end{aligned} \tag{3}$$

Function I_s in (3) will be equal to one when (x_1, x_2) is a similar pair, while $I_{d}s$ works in reverse manner. The only hyper-parameter in the loss is the c in (3).

III. MATERIALS

A. Experiment settings

In our paper, we use the public dataset – the SJTU emotion EEG dataset (SEED) [10]. There are three kinds of emotion states in our stimulating film clips, including positive, neutral and negative. The number of film clips is 15, and the duration of each film clip is about 4 minutes. The order of presentation is arranged so that two film clips targeting the same emotion are not shown consecutively.

For the feedback, participants are told to report their emotional reactions to each film clip by completing the questionnaire immediately after watching each clip. EEG signals of 15 subjects were recorded while they were watching the emotional film clips. EEG signals are recorded by an ESI NeuroScan system at a sampling rate of 1000 Hz from 62-channel electrode cap according to the international 10-20 system.

B. Data preprocessed and Feature extraction

The data was first downsampled to 200Hz. A bandpass frequency filter from 0-75Hz was applied. We extracted the EEG segments corresponding to the duration of each movie. The EEG data were visually checked and the recordings seriously contaminated by electromyography (EMG) and Electrooculography (EOG) were removed manually from the dataset. EOG was simultaneously recorded in the experiments to vanish the blink artifacts from the recordings.

According to five frequency bands: delta (1- 3Hz); theta (4-7Hz); alpha (8-13Hz); beta (14-30Hz); and gamma (31-50Hz), we compute the traditional PSD features using Short Time Fourier Transform (STFT) with a 1s long window and

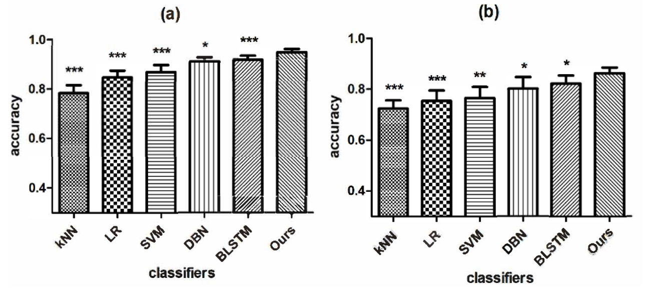

Fig. 3. The results of 5 compared classifiers and our purposed method with the feature of (a) DE and (b) PSD respectively. (* : p<0.05, ** : p<0.01, *** : p<0.001.)

TABLE I

COMPARISONS ON EEG-BASED EMOTION DATASET SEED

Method	Feature	Frequency bands	Channels number	Accuracy (%)
SVM [10]	PSD	$\delta, \theta, \alpha, \beta, \gamma$	62	59.60
	DE	$\delta, \theta, \alpha, \beta, \gamma$	62	83.99
	PSD	$\delta, \theta, \alpha, \beta, \gamma$	12	62.92
	DE	$\delta, \theta, \alpha, \beta, \gamma$	12	86.65
DBN [10]	PSD	$\delta, \theta, \alpha, \beta, \gamma$	62	61.90
	DE	$\delta, \theta, \alpha, \beta, \gamma$	62	86.08
HCNN [15]	DE	γ	62	88.20
STRNN [16]	DE	$\delta, \theta, \alpha, \beta, \gamma$	62	89.51
BDAE [21]	DE eye movement	$\delta, \theta, \alpha, \beta, \gamma$	62	91.01
Ours	PSD	$\delta, \theta, \alpha, \beta, \gamma$	62	**86.27**
	DE	$\delta, \theta, \alpha, \beta, \gamma$	62	**94.96**

no overlapping Hanning window. The differential entropy feature [8] is defined as follows.

$$h(X) = -\int_{\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} exp\frac{(x-\mu)^2}{2\sigma^2} log\frac{1}{\sqrt{2\pi\sigma^2}} \\ exp\frac{(x-\mu)^2}{2\sigma^2} dx = \frac{1}{2} log 2\pi\sigma^2 \tag{4}$$

where X submits the Gauss distribution $N(\mu, \sigma^2)$, x is a variable, and π and e are constants. In each frequency band, DE is equivalent to the logarithmic power spectral density for a fixed length EEG sequence. Since each frequency band signal has 62 channels, we extracted differential entropy features with 310 dimensions for a 1 s sample. After this process, we use a no-overlapping slicing window of 10s to temporally scan the sequences, and get the 10*310 dimensions time series sample.

C. Classifiers training details

For every data file, the data from the subjects watching the first 9 movie clips are used as training samples and the rest 6 movie clips are used as test samples. The model which doesn't contain time property will use the original 310-dimension sample, and BLSTM-based models include the 10s information in one sample so that the shape of a sample is 10*310.

- For kNN, we use k=5 for baseline in comparison with other classifiers.
- For LR, we employ L2-regularized LR and search the regularization parameter in $[1.5 : 10]$ with a step of 0.5.
- For SVM, we use the linear kernel SVM and search the parameter space $2^{[-10:10]}$ with a step of one for C to find the optimal value.
- For deep neural networks, we construct a DBN with two hidden layers. We search the optimal numbers of neurons in the first and the second hidden layers with step of 50 in the ranges of $[200 : 500]$ and $[150 : 500]$, respectively.
- We construct a BLSTM with two hidden layers, and search the L2 regularization parameter and the learning rate in the range of $10^{[-5:-1]}$ and $10^{[-3:-4]}$ with a step of one. As for our purposed method, we add one parameter—-the weight of the pairwise constrain loss, in the range of $10^{[-3:-1]}$.

IV. RESULTS AND DISCUSSION

In this section, we design series of contrast experiments and present the results of our approaches on the SEED dataset.

A. Classification performance

We first compare two kinds of features: DE and PSD. We can clearly see that the DE features have a higher accuracy and lower standard deviation than the traditional PSD features, implying that the DE feature are more descriptive feature than the PSD feature in emotion recognition research. We also show the classification results of kNN, LR, SVM, DBN, BLSTM and our pairwise-constrain loss method for the mean accuracy of 15 subjects in Fig 3.

Accuracy means and standard deviation using DE feature in percentage(%) of kNN, LR, SVM, DBN, BLSTM and the pairwise loss BLSTM are 78.35/12.46, 84.84/10.58, 86.99/10.66, 91.27/6.36, 92.0/5.98, 94.96/4.86, respectively. It shows that DBN, BLSTM and our purposed method perform better than the traditional shallow model. The results of PSD feature can also verify our implication, which are respectively 72.38/12.42, 75.38/15.81, 76.52/16.91, 80.32/16.87, 82.22/12.30 and 86.27/8.60.

In order to evaluate the performance of our proposed pairwise constrain loss, we compared our proposed method with the standard BLSTM without the pairwise constrain loss. BLSTM with the pairwise constrain loss significantly outperforms the framework without it. (DE feature: 92.05/5.98 vs 94.96/4.86 p:0.004 and PSD feature 83.51/12.30 vs 86.27/8.60 p:0.012). We also compared our result with those of various existed algorithms which is also used the SEED dataset, shown in Table 1, The pairwise constrain BLSTM model with the DE feature outperforms better than other methods with the best accuracy of 94.96%.

B. Feature Visualization

For simplicity, we select one typical subject's data to exhibit our loss's performance. We provide the visualizations of two models on the feature space layer and softmax layer.

T-Distributed Stochastic Neighbor Embedding (t-SNE)[22] is a technique for dimensionality reduction that is used for the visualization of high-dimensional datasets. We choose t-SNE to visualize the high-dimensional features. In Fig 4, in the input space, the original data can not be easily discriminated by linear classifiers. We put the continuous 10s together to get the time series and input it in our framework. In the two models, the features of different emotion states become separable. To investigate the effect of the pairwise constrain loss to the BLSTM model, we compare the first row (without this loss) with the second row (with this loss) in Fig 4. We found that using the pairwise loss could make the distance between different categories farther, and help to learn a more discriminative embedding feature space.

We can also find additional phenomena in the 2-D projection of the representations. First of all, the positive emotion is easy to separate. In each of two rows, we find the positive emotion is more easily to be separated from the neutral emotion and the negative emotion. Secondly, the neutral emotion and the negative emotion are more similar to each other, it may imply that the 'no emotion' state in our cognition pattern can be more similar to the negative state. What's more, the distance between the green dots and blue dots is farther visibly when using the pairwise constrain. It shows that the loss has a useful effect on the two states which hard to discriminate.

V. Conclusions

In this paper, we proposed the BLSTM with the pairwise constrain loss to classify the three states of emotion, our method outperforms the state-of-the-art emotion classification approaches in the benchmark EEG emotion dataset–SEED dataset , which get a mean accuracy of 94.62%.

The visualizations of the embedding feature show the feature evolution along the cascaded layers, where the representations become linearly separable. It shows that pairwise constrain loss can help to learn a more discriminative embedded feature space, combined with the traditional supervised classification loss function.

As future work, we will focus on the following issues that we have not covered in this paper. We want to go deeper with the performance of the pairwise constrain loss when parameters change. Besides, more experiments are needed in order to study the stability of the network.

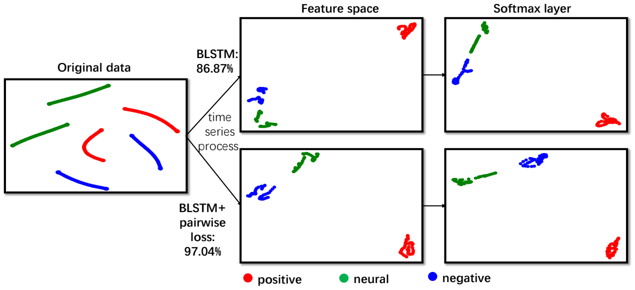

Fig. 4. The visualization of the LSTM network with or without the pairwise constrain loss. Only testing data are shown and we choose the DE features to get better performance.

References

[1] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," *IEEE transactions on affective computing*, vol. 3, no. 2, pp. 211–223, 2012.

[2] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, "Emotion recognition in human-computer interaction," *IEEE Signal processing magazine*, vol. 18, no. 1, pp. 32–80, 2001.

[3] K. Anderson and P. W. McOwan, "A real-time automated system for the recognition of human facial expressions," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 36, no. 1, pp. 96–105, 2006.

[4] H. Gunes and M. Piccardi, "Bi-modal emotion recognition from expressive face and body gestures," *Journal of Network and Computer Applications*, vol. 30, no. 4, pp. 1334–1345, 2007.

[5] J. Ang, R. Dhillon, A. Krupski, E. Shriberg, and A. Stolcke, "Prosody-based automatic detection of annoyance and frustration in human-computer dialog," in *Seventh International Conference on Spoken Language Processing*, 2002.

[6] Z. Khalili and M. Moradi, "Emotion detection using brain and peripheral signals," in *Biomedical Engineering Conference, 2008. CIBEC 2008. Cairo International*, pp. 1–4, IEEE, 2008.

[7] G. L. Ahern and G. E. Schwartz, "Differential lateralization for positive and negative emotion in the human brain: Eeg spectral analysis," *Neuropsychologia*, vol. 23, no. 6, pp. 745–755, 1985.

[8] L.-C. Shi, Y.-Y. Jiao, and B.-L. Lu, "Differential entropy feature for eeg-based vigilance estimation," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pp. 6627–6630, IEEE, 2013.

[9] C. A. Frantzidis, C. Bratsas, C. L. Papadelis, E. Konstantinidis, C. Pappas, and P. D. Bamidis, "Toward emotion aware computing: an integrated approach using multichannel neurophysiological recordings and affective visual stimuli," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 3, pp. 589–597, 2010.

[10] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.

[11] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, "Identifying stable patterns over time for emotion recognition from eeg," *IEEE Transactions on Affective Computing*, 2017.

[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[14] R. J. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural computation*, vol. 1, no. 2, pp. 270–280, 1989.

[15] J. Li, Z. Zhang, and H. He, "Hierarchical convolutional neural networks for eeg-based emotion recognition," *Cognitive Computation*, pp. 1–13, 2017.

[16] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, "Spatial-temporal recurrent neural network for emotion recognition," *IEEE transactions on cybernetics*, no. 99, pp. 1–9, 2018.

[17] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[18] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *null*, pp. 1735–1742, IEEE, 2006.

[19] Y.-C. Hsu and Z. Kira, "Neural network-based clustering using pairwise constraints," *arXiv preprint arXiv:1511.06321*, 2015.

[20] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *International Conference on Machine Learning*, pp. 1310–1318, 2013.

[21] W. Liu, W.-L. Zheng, and B.-L. Lu, "Multimodal emotion recognition using multimodal deep learning," *arXiv preprint arXiv:1602.08225*.

[22] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.