Multi-Branch Network for Cross-Subject EEG-based Emotion Recognition

Guang Lin LINDANDAN@HDU.EDU.CN
Li Zhu ZHULI@HDU.EDU.CN
Bin Ren MUMUS@HDU.EDU.CN
Yiteng Hu HYT18041513@HDU.EDU.CN
Jianhai Zhang* JHZHANG@HDU.EDU.CN
Hangzhou Dianzi University, Hangzhou, China

Editors: Vineeth N Balasubramanian and Ivor Tsang

Abstract

In recent years, electrocardiogram (EEG)-based emotion recognition has received increasing attention in affective computing. Since the individual differences of EEG signals are large, most models are trained for specific subjects, and the generalization is poor when applied to new subjects. In this paper, we propose a Multi-Branch Network (MBN) model to solve this problem. According to the characteristics of the cross-subject data, different branch networks are designed to separate the background features and task features of the EEG signals for classification to have better model performance. Besides, there is no new-subject data needed during model training. In order to avoid the negative improvement caused by samples with significant differences to model training, a tiny amount of new-subject data is used to filter the training samples to improve the model performance further. Before training the model, the samples with significant differences from the new subject were deleted by comparing the background features between the subjects. The experimental results show that compared with Single-Branch Network (SBN) model, the accuracy of the MBN model is improved by 20.89% on the SEED dataset. Furthermore, compared with other common methods, the proposed method uses less new-subject data, which improves its practicability in practical application.

Keywords: Multi-Branch Network, Electroencephalogram (EEG), Emotion Recognition, Cross-Subject

1. Introduction

Emotion plays a vital role in the interaction between people. Humans expect machines to communicate with us in human-machine interaction (HMI) (Cowie et al. (2001); Fiorini et al. (2020)) according to our emotions, among which emotion recognition is the crucial issue. As a typical artificial intelligence task, emotion recognition has a broad development prospect in HMI, emotional computing, psychology, and clinical medicine. It provides a possible way for computers to understand human emotions. Emotions can usually be identified through information like facial expressions, gestures, speech (Yan et al. (2016); Zhang et al. (2019)), electrocardiogram (ECG), and electroencephalogram (EEG) (Chen et al. (2015); Zheng et al. (2017); Hsu et al. (2017)). Because of its high accuracy, low cost, and ease of operation, EEG signals have been widely used in emotion recognition. At the same time, human beings will deliberately hide their emotions in some scenes. According to facial expressions, gestures, voice, and other external expressions, a person's inner thoughts and emotions cannot be determined, while the EEG signals are hard to hide.

In the last decade, there has been increasing study on the cognition of affective states within the brain-computer interface (BCI) community (Lotte et al. (2007); Aricò et al. (2020)). The ideal affective brain-computer interface (aBCI) (Garcia-Molina et al. (2013); Goshvarpour and Goshvarpour (2019)) can detect the emotional states felt by the user through spontaneous EEG signals without explicit user input and respond to different emotional states accordingly. The EEG signals can record very subtle emotional changes at the high temporal resolution, but the EEG signals are time asymmetric and unstable. Second, the signal-to-noise ratio (SNR) of EEG is shallow, and it is difficult to eliminate the influence of noise. In addition, due to the significant differences in EEG signals between different subjects or different states, the performance of existing models on the new-subject data is lacking (Jayaram et al. (2016); Zheng and Lu (2016)). The EEG patterns vary among subjects, so it is necessary to train the classifier for a specific subject in the presence of multiple subjects (Blankertz et al. (2007)). Training a specific model for each new subject is an inefficient approach to resolving differences, requiring the collection of labeled datasets and retraining of the model (Shen and Lin (2019)). At present, cross-subject emotion recognition based on EEG signals is still a challenge.

In traditional machine learning algorithms, Zheng et al. (2015) used transfer component analysis (TCA) and kernel principle component analysis (KPCA) for subspace projection. These two algorithms projected the data with different distributions to the same feature space and reduced the distribution difference between the two domains. On this basis, Zheng and Lu (2016) carried out a further study and proposed transductive parameter transfer (TPT). They used the extracted features to select the corresponding classifier and learned the regression function between the data distribution and the classifier parameters, which further improved the model performance. In the work of Chai et al. (2017), a new adaptive subspace feature matching (ASFM) strategy was proposed, which considered both the marginal distribution and the conditional distribution of data. Finally, Li et al. (2019) combined the advantages of most of the above algorithms and proposed multisource transfer learning (MSTL), which reached the state-of-the-art (SOTA) performance in traditional machine learning.

In the deep learning algorithm, Li et al. (2018) used deep adaptation network (DAN) to minimize the MK-MMD distance (Long et al. (2015)) on the depth features of different domain data in neural networks so that the feature distribution of the data was similar. Luo et al. (2018) mapped the source domain (existing-subject) data and target domain (new-subject) data to common feature space and used an adversarial network (Ganin and Lempitsky (2015)) to make the distribution of the two more similar. Wang et al. (2018a) and Zhong et al. (2020) both used graph neural network to better extract features and used the adversarial network to solve individual differences. In the work of Du et al. (2020), they proposed ATDD-LSTM model, and its classification accuracy reached the current SOTA performance. The ATDD-LSTM model also used the adversarial network for domain adaptation.

Compared with traditional machine learning algorithms, deep learning algorithms are easier to implement and generally have higher accuracy. However, the existing deep learning algorithms rarely consider the conditional probability distribution of data in domain adaptation, which leads to negative transfer (Chai et al. (2017)). When the difference between the two subjects is too significant, even if the marginal distribution of the two domains is minimized, the classification accuracy is still poor. On the other hand, the above algorithms used transfer learning to eliminate the differences between different subjects, which required the use of all unlabeled target domain data. In practice, it takes a long time to collect data (Li et al. (2019)), so these algorithms still have limitations. In order to solve individual differences and improve the usability of the algorithm in practice, this paper has made the following contributions:

- A novel Multi-Branch Network (MBN) model was designed to solve individual differences without collecting new-subject data. The MBN model can effectively extract background features and task features of different subjects from the original EEG signals, thus better deal with different EEG signals in the cross-subject task.
- A subject similarity matrix was proposed by calculating the distance of the background feature maps. The matrix was used to select samples with higher similarity to the new subject to avoid the negative improvement caused by samples with significant differences during model training.

2. Methods

Fig. 1 shows the framework of the Multi-Branch Network model for EEG-based emotion recognition. It includes three parts: feature organization, Multi-Branch Network model, and sample selection.

2.1. Feature organization

The physiological characteristics of EEG signals vary with frequency, and the emotional responses of different encephalic regions are different (Zheng and Lu (2015); Wang et al. (2018b)). Considering these two factors, this paper uses a 3D structure (Yang et al. (2018); Shen et al. (2020)) to organize the EEG signals (Fig. 1(a)). First, perform band-pass filtering on the EEG signals of all channels, and divide the signal of each channel into delta (1-4 Hz), theta (4-8 Hz), alpha (8-14Hz), beta (14-31 Hz), and gamma (31-50 Hz). Then extract differential entropy (DE) feature (Duan et al. (2013)) in five frequency bands for the t s window size samples. Finally, we organize the DE feature of each frequency band into a 2D map. Therefore, each segment is a 3D structure $X_n \in \mathbb{R}^{h \times w \times d}$, $n = 1, 2, \ldots, N$, where N is the number of samples, h and w are the width and height of the 2D map, $d = 5$ is the number of frequency bands. More details are as follows.

According to the work of Wang et al. (2014), we use $t = 1$ s as the time window for data segmentation and obtain a series of slices with $L \times C$ ($L = t \times W$), where C is the number of channels, L is the data length, and W is the sampling frequency. The DE feature is extracted per second from five frequency bands, and the data structure is changed from $L \times C$ to $1 \times C \times 5$, which is a matrix of $C \times 5$. Differential entropy is the generalized form

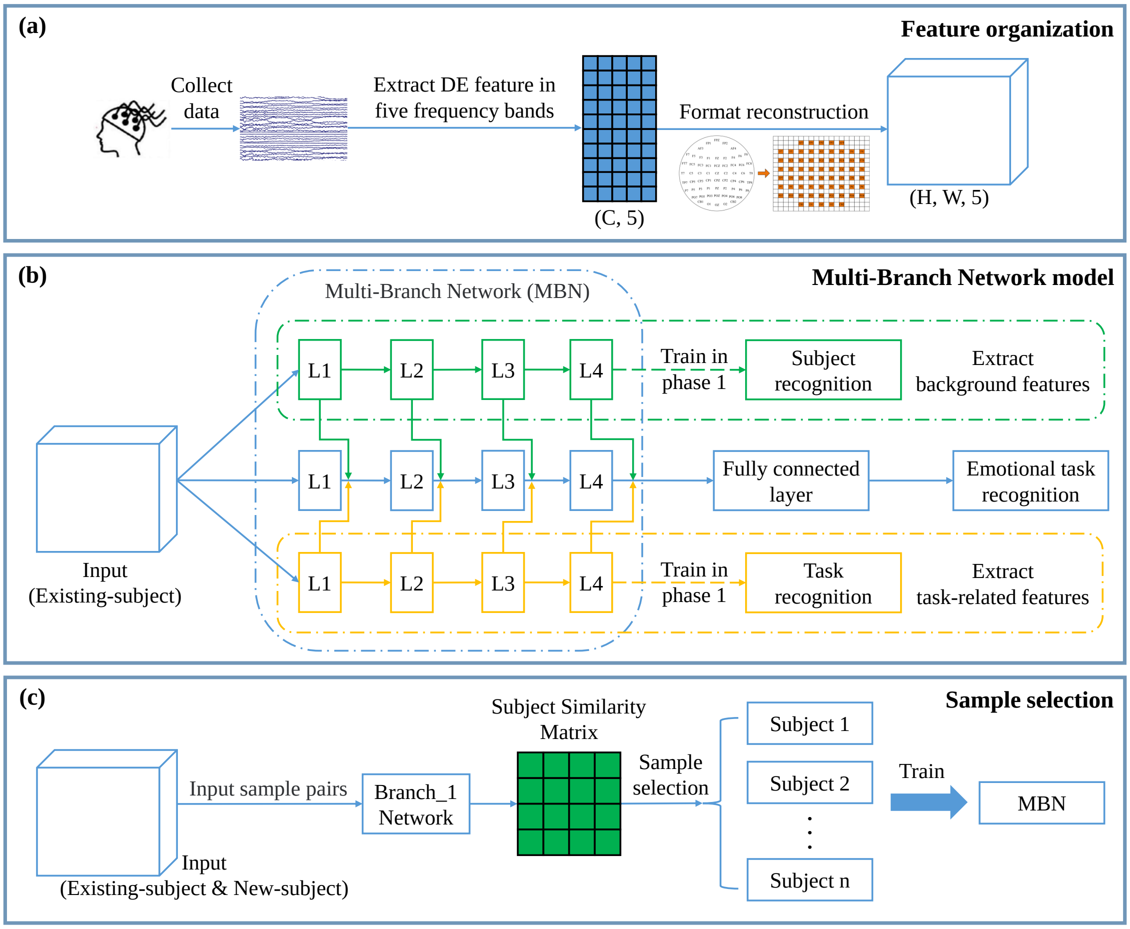


Figure 1: An overview of the proposed EEG-based emotion recognition framework using the Multi-Branch Network model.

of Shannon entropy on continuous variables:

$$DE = -\int_a^b p(x)\log(p(x))dx, \tag{1}$$

where $p(x)$ represents the probability density function of continuous information, and $[a, b]$ represents the interval of information value. For a section of EEG that approximately obeys the Gaussian distribution $N\left(\mu, \sigma^2\right)$, the DE feature is calculated as below:

$$DE = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\frac{(x-\mu)^2}{2\sigma^2} \log\frac{1}{\sqrt{2\pi\sigma^2}} \exp\frac{(x-\mu)^2}{2\sigma^2} = \frac{1}{2}\log 2\pi e\sigma^2. \tag{2}$$

In order to better preserve the spatial features of the EEG signals, according to the spatial distribution of the electrodes (Fig. 1(a)), we convert the C-dimensional DE feature vector on each frequency band into a 2D matrix. Finally, the 3D feature map of $h \times w \times 5$ is obtained.

2.2. Multi-Branch Network model

Starting from the characteristics of the cross-subject EEG signals (consisting of background information and task information), a Multi-Branch Network (MBN) model is used to extract features of background and task, respectively. By inputting multi-feature data (Fig. 1(b)), the deep learning model can learn better. In our method, the training process is divided into two phases: Training two branch networks to extract two features; training the backbone network for classification.

2.2.1. The branch network

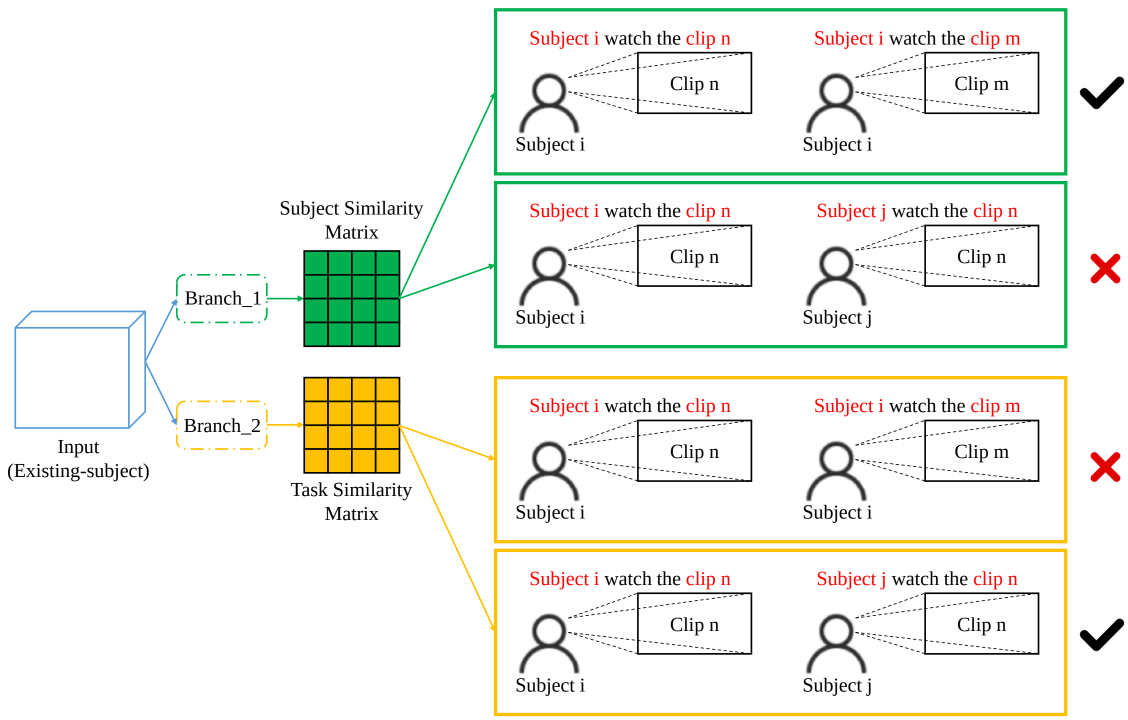


Figure 2: The schematic diagram of phase 1. Teach branch networks to learn background features and task features through opposite task labels. Where i and j represent different subjects, m and n represent clips of different categories.

As shown in Fig. 2, phase 1 uses two opposite tasks to extract different features. The same data have different labels in different branches. In the $Branch_1$ network, select $K-2$ subjects and organize their data samples into sample pairs as training dataset, where K is the number of all subjects in the dataset. Each sample pair consists of two random samples from the dataset. When the two samples are from the same subject, the label is 1; otherwise, it is 0. The data of the remaining 2 subjects are organized into sample pairs in the same way as test dataset. In the sample pair, there are data of different subjects watching the same kind of clips (labeled 0) and data of the same subject watching different types of film clips (labeled 1). After this constraint, the features extracted by the model are more personal-related background features rather than task features. Similarly, after the constraint of the $Branch_2$ network, the extracted features are more related to the task. The constraints for training the branch network are as follows:

$$X_1' = B\left(X_1\right), \tag{3}$$

$$X_2' = B\left(X_2\right), \tag{4}$$

$$d\left(X_1', X_2'\right) = \sqrt{|X_1' - X_2'|^2}, \tag{5}$$

$$\text{label } = \begin{cases} 1, & d < 0.2 \\ 0, & \text{else } , \end{cases} \tag{6}$$

where X_1 and X_2 are the sample pairs input to the branch network $B(\cdot)$, and output the feature maps X_1' and X_2'. Finally, the Euclidean distance d of the two feature maps is calculated as the similarity degree between the two samples. When the Euclidean distance is less than 0.2, the label is predicted to be 1; otherwise, the label is predicted to be 0. In the branch network training, the contrastive loss (Hadsell et al. (2006)) is used as the loss function. The network structures of $Branch_1$ network and $Branch_2$ network are the same. The branch network structure is shown in Fig. 3.

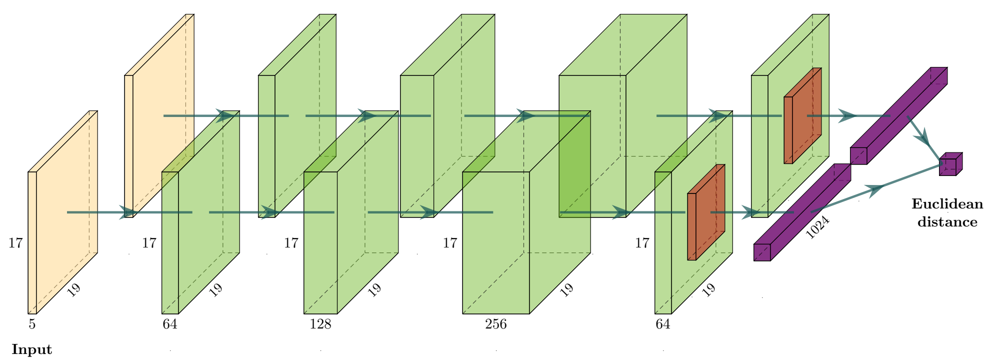


Figure 3: The framework for training one of the branch networks.

In each branch network, there are four convolutional layers, one max pooling layer, and one fully connected layer. The parameters of the four convolutional layers are: 64 feature maps with filter size of 5 × 5; 128 feature maps with filter size of 4 × 4; 256 feature maps with filter size of 4 × 4; 64 feature maps with filter size of 1 × 1. For all convolutional layers, zero-padding and rectified linear units (ReLU) activation function are applied. The pool size of the max pooling layer is 2 × 2, and the stride is 2. Finally, the units in the fully connected layer are 1024, and the result of the fully connected layer is output as a vector. After the training is completed, the parameters are saved for the call of the MBN model.

2.2.2. The backbone network

In the second part of the training phase, a single sample is input into the branch networks to extract different features and fuse them with the feature map of the backbone network. The branch networks are trained in phase 1, and their convolutional layers can extract background features and task features. The backbone network is designed for the final EEG-based emotion recognition. In the MBN model, the feature maps of the same layer are spliced and input into the next layer of the backbone network. Finally, after one max pooling operation, the feature map is input into the fully connected layers for emotional classification. We consider the MBN model defined as:

$$Y = F\left(Concat\left(X, X', X''\right)\right) = F\left(Concat(X, B_1(X), B_2(X))\right), \tag{7}$$

Where $F(\cdot)$, $B_1(\cdot)$, and $B_2(\cdot)$ represent the output of the backbone network and the two branch networks in the convolutional layer, and $Concat(\cdot)$ represents the operation to concatenate a list of inputs.

During the training, the parameters of the branch networks are fixed. The MBN model adds the background features and task features of the EEG signals while inputting the original data. The model can extract the different information between different signals to perform better in the cross-subject task. The structure of the MBN model is shown in Fig. 4.

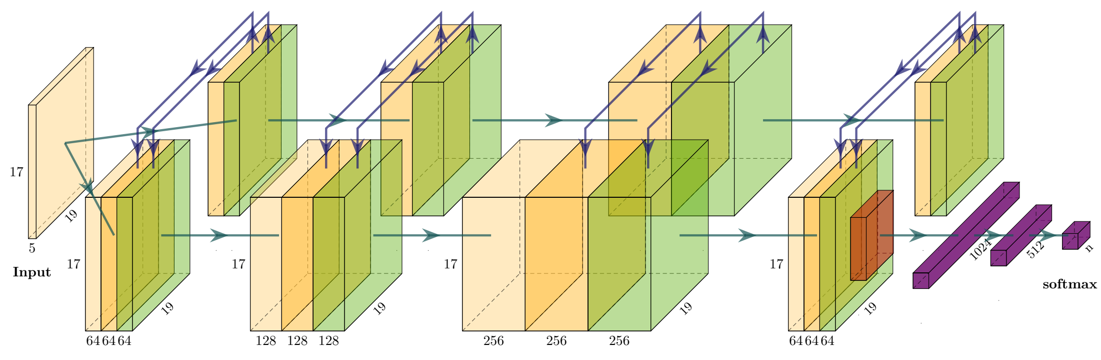


Figure 4: The framework for training the Multi-Branch Network.

In the backbone network, there are also four convolutional layers and one max pooling layer, and its hyperparameters are the same as that of the branch network so that the three networks can splice the feature maps output by the convolutional layer. After the max pooling layer, two dropout layers and three fully connected layers are added. The rate of the two dropout layers is both 0.5, which can effectively improve the generalization ability and enhance the anti-noise ability of the model. The units in the three fully connected layers are 1024, 512, and n, respectively, where n is the number of emotional categories. In training, the cross-entropy loss is used as the loss function.

At the same time, to better compare the performance improvement brought by the MBN model, a network called Single-Branch Network (SBN) is constructed using the backbone network structure. We consider the SBN model defined as:

$$Y = F\left(X\right). \tag{8}$$

2.3. Sample selection

It is a considerable challenge to train a model suitable for the cross-subject task without using the new-subject data. When the difference between training subjects and the new subjects is too significant, it will be difficult to adjust the model to make it have a good performance on the new subjects. Therefore, before the training of the MBN model, a tiny amount of new-subject samples without labels are used to select the training samples.

In the study of applying transfer learning, n types of emotion data need to be collected. In the actual data collection process, the generation and conversion of emotions require a period of initiation time. Therefore, it takes a relatively long time to collect data on multiple emotions. In this paper, since the training task of the $Branch_1$ network is irrelevant to the emotional information, only partial data of one film is needed to be collected instead of multiple films, which further shortens the time required for data collection.

The overall process of sample selection is shown in Fig. 1(c). For each subject in the test dataset, select the top T samples, and form a sample pair with the top T samples of other subjects, to obtain a total of $(K-1)\times T$ sample pairs. After all sample pairs are input into the corresponding $Branch_1$ network model and mapped to feature space, the Euclidean distance between the feature vectors is calculated. It can be seen in Fig. 6(a). Finally, the similarity of T sample pairs is averaged as the similarity between the subjects. According to the similarity of all subjects, the subject similarity matrix $M \in \mathbb{R}^{K\times K}$ is obtained.

$$M_{ij} = \frac{1}{T}\sum_{t=1}^{T} d\left(X_i^t, X_j^t\right), \tag{9}$$

where $i, j = 1, 2, \ldots, K$.

From the similarity matrix M, the subjects with higher similarity to the new subjects are selected to train the MBN model.

3. Experiment

3.1. Dataset

This paper conducted experiments based on the SEED dataset (Zheng and Lu (2015)) which BCMI Lab collected. Through careful selection of film clips, different types of emotions can be stimulated, including positive, negative, and neutral. Finally, 15 Chinese film clips were selected from the material library (6 films) as the stimuli used in the experiment, and every 5 clips correspond to a kind of emotion. Fifteen healthy subjects participated in the EEG signals collection experiment. In each experiment, the above 15 clips were played to stimulate the corresponding emotions, and the 62-channels' ESI NeuroScan System was used to record the EEG signals with a sampling frequency of 1000 Hz. After each clip, subjects were asked to complete a questionnaire immediately to report their emotional response. Each subject conducted the same experiment 3 times in different periods, so there were 45 experiments in total. In order to reduce the storage space and the amount of calculation, after removing some basic noise from the data, the data was down-sampled to 200 Hz.

3.2. Experimental setup

In the feature organization, each subject's data on the SEED was sliced into 3394 slices with $L \times C$ ($L = T \times W$), where $L = 1 \times 200$, $C = 62$. After DE feature extraction and format reconstruction, the 3D feature map of $H \times W \times 5$ was obtained, where $H = 17$, $W = 19$.

The experiments were carried out on single-subject dataset and cross-subject dataset. In the experiment with the single-subject dataset, we applied fivefold cross-validation on each subject and selected the experimental data with the highest accuracy in each subject for subsequent experiments to reduce the calculation time. In the experiment with the cross-subject dataset, the data of 13 subjects was used as the training dataset, and the data of the remaining 2 subjects was used as the test dataset.

The training was done using an NVIDIA TITAN Xp Founders Edition graphics card with CUDA 10.0 and cuDNN v7.6, in TensorFlow v1.15.0 and Keras v2.3.1 (Abadi et al. (2016); Géron (2019)).

3.3. Results and discussion

This section compares and analyzes the performance of the proposed MBN model under different conditions, including single-subject task, cross-subject task, and sample selection.

3.3.1. Single-subject task

In the single-subject task, the difference between the training dataset and the test dataset is small, so the trained model has a good performance on the test dataset.

From Fig. 5(a), we can see the accuracy of the SBN model and MBN model in each subject's three-category emotion recognition. In the MBN model, the accuracy of all subjects exceeded 90%, and the average accuracy reached 95.71%, which was 2.22% higher than that of the SBN model. At the same time, the results on all the subjects show that the MBN model has different degrees of improvement compared with the SBN model, indicating that the extraction of multiple features is beneficial to the EEG-based emotion recognition. The accuracy under the single-subject task will be used as a baseline for comparison with subsequent experiments.

3.3.2. Cross-subject task

In the training branch network section, the network model in Fig. 3 is used to train the model to perform consistency (subject consistency and task consistency) judgments on the data. Then, different information is extracted by training the branch networks through completely opposed tasks (Fig. 2).

The experimental results are as follows: In the subject consistency judgment, the average precision of the model reaches 98.02%, which is an exciting result. In the task consistency judgment, the average precision of the model is only 68.42%. We can see that the background features are obviously different in different subjects and can be easily recognized. On the other hand, the signal-to-noise ratio of the EEG signals is meager, and the performance of the model that uses data containing background noise is poor. Therefore, background features need to be considered in the cross-subject task. The experimental results show that the $Branch_1$ network can effectively extract the background features.

In the training backbone network section, the input and output of the convolutional layer changed from the original $Y = F(X)$ to $Y = F(Concat(X, X', X''))$. These two types of functions represent the SBN model and the MBN model, respectively. Fig. 5(b) shows the test results of the two models on the cross-subject task.

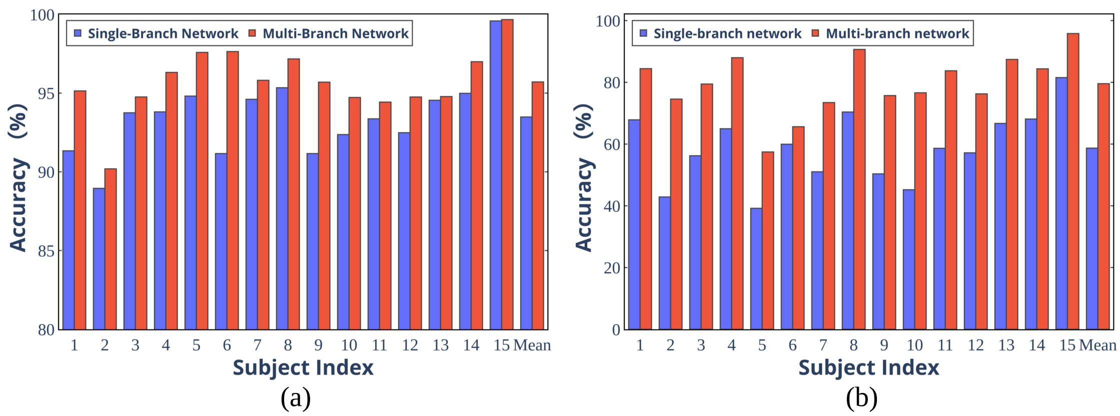


Figure 5: (a) The accuracy under the single-subject task. (b) The accuracy under the cross-subject task.

From Fig. 5, no matter which kind of model is, its classification accuracy is lower in the cross-subject task, which is caused by individual differences. Compared with the single-subject task, in the cross-subject task, the accuracy of the SBN model decreased by 34.81%, and that of the MBN model only decreased by 16.14%.

In the cross-subject task, although the SBN model performed well on the single-subject task with an average accuracy of 93.49%, it performed poorly on the new-subject data with an average accuracy of only 58.68%. The average accuracy of the MBN model reached 79.57%, which was 20.89% higher than the SBN model. Meanwhile, the MBN model is more stable. Except for subjects #5 and #6, the classification accuracy of the other subjects exceeds 70%. For subjects #4, #8, #13, and #15, the average accuracy is over 85%, and the highest accuracy is 95.80%. It shows that adding background features and task features to train the model can effectively capture the differences among subjects and improve the model performance.

Our work is a new attempt to solve the problem of individual differences without using the new-subject data. While considering the task features, the influence of background features on the cross-subject emotion recognition task was considered, and some valuable features were automatically learned during the training process using the deep learning model.

Table 1: Means and standard deviations of multiple methods on the SEED dataset under the cross-subject task

Algorithm	SBN	TCA	TPT	MSTL	DAN	ATDD-LSTM	MBN
Mean (%)	58.68	71.80	76.31	88.92	83.81	90.92	79.57
Std. (%)	11.21	13.99	15.89	10.35	8.56	1.05	9.53
Improve (%)	-	13.12	17.63	30.24	25.13	32.24	20.89

3.3.3. Comparison of other methods

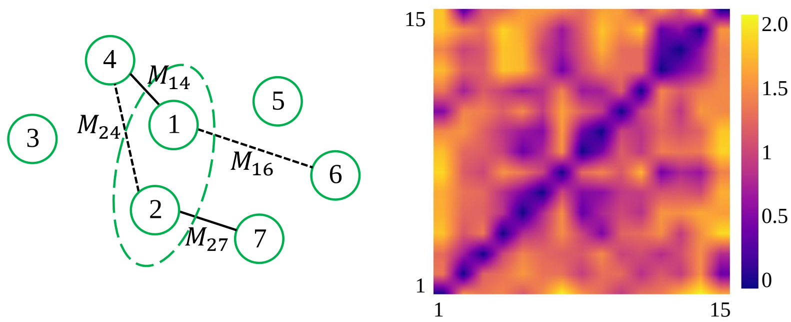
For SEED, we compared the performance of various methods mentioned in section 1 with the MBN model. The average accuracy and standard deviation (STD) represent the final performance of the model, and all results were obtained on the data of 15 subjects. As shown in Table 1, the SBN model is compared as the baseline. The five algorithms used for comparison are all based on transfer learning, using all or part of the unlabeled new-subject data to adjust the model. TCA, TPT, and MSTL are traditional machine learning algorithms. MSTL taken negative transfer into account and trained the model by selecting specific subject data, which achieved SOTA performance in traditional machine learning. DAN and ATDD-LSTM are deep learning algorithms, all of which used domain adaptation to train the model. Under the ATDD-LSTM algorithm, the cross-subject classification accuracy reached 90.92%.

The performance of the MBN model is better than two common traditional machine learning algorithms, but there is still a certain gap with the current SOTA performance. Our work is of great significance for the practical BCI application. To our knowledge, the MBN model is the first algorithm that does not use the new-subject data and extracts background features to reduce individual differences. Compared with the other methods, the proposed method can shorten the time required for data collection in practice.

3.3.4. Sample selection

It is a considerable challenge to train a model suitable for the cross-subject task without using the new-subject data. Although the MBN model has achieved good accuracy, it is still not satisfactory. In this session, we use the subject similarity matrix M calculated by collecting a tiny amount of data to select samples to improve the classification accuracy further.

First, select the top 60 new-subject data samples of one clip, and combine these 60 samples with the top 60 samples of other subjects to form sample pairs, resulting in a total of $14 \times 60 = 840$ sample pairs. After obtaining the sample pairs, they are input into the $Branch_1$ network to calculate the similarity between the samples, and the similarity of every 60 sample pairs is averaged as the similarity between the two subjects. Fig. 6(a) shows the distribution of different samples mapped to the feature space, and the similarity is determined according to M_{ij}. Finally, the heat map of the subject similarity matrix M is shown in Fig. 6(b).



(a) (b)

Figure 6: (a) The distribution of samples in the same feature space. The sample in the dotted circle is from new subjects. The dashed and solid lines represent the sample pairs with the largest distance (the most significant difference) and the smallest distance (the highest similarity). (b) The heat map of the subject similarity matrix.

In Fig. 6(b), the similarity between each subject (each row) and other subjects is obtained from the corresponding $Branch_1$ network model, so the resulting matrix is not a completely symmetrical matrix. According to the above experiments, the $Branch_1$ network has excellent performance. Although the results are output by 15 different models, the matrix still maintains a high symmetry, which further confirms the credibility of successfully extracting background features. Finally, based on the similarity matrix, two different experiments are carried out, and the results obtained are shown in Fig. 7.

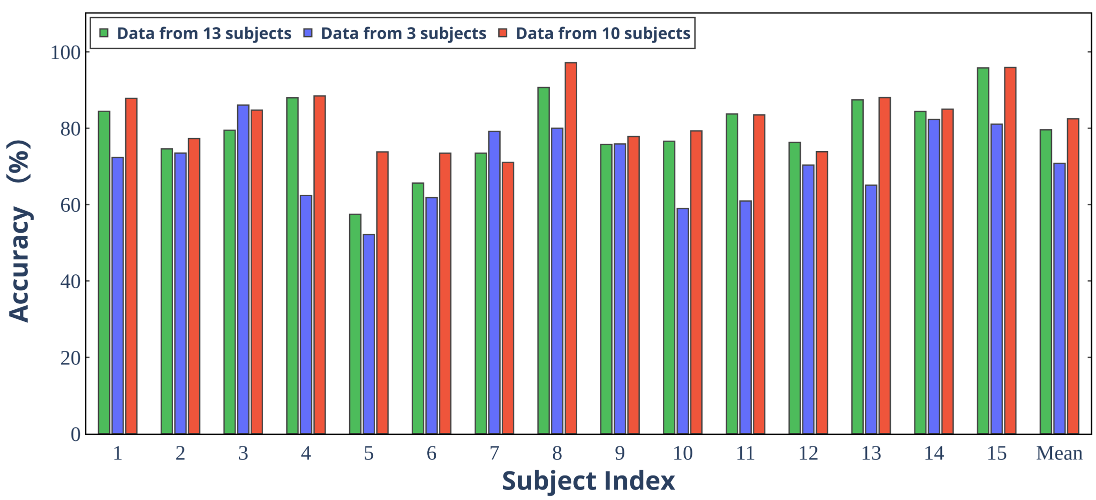


Figure 7: The accuracy of training the model with different amounts of subject data.

When using the data of 3 existing subjects who most similar to the new subject to train the model, the accuracy is further improved on subjects #3, #7, and #9. It indicates that the similarity between the subjects is high, and even when less data is used for training, better performance can be obtained. However, for subjects #1, #4, #8, #10, #11, #13, and #15, the model's accuracy dropped significantly, with a decline rate of more than 10%. The average accuracy of 15 subjects was 70.79%, which does not exceed that of the MBN model without sample selection, indicating that too little data is terrible for model training. In limited data samples, using more data can improve the model's generalization ability, and better model performance can be obtained on new-subject data. However, while increasing the amount of data, some samples with too significant differences may be added, which will cause the model to perform worse.

In consideration of the above situation, the data of the 3 subjects, which were significantly different from the new subjects, were removed, and the remaining data were used to train the model. In most subjects, the classification accuracy is improved. For subjects #7, #11, and #12, the accuracy is dropped slightly, with an average drop of 1.70%. For subjects #3, #5, #6, and #8, the performance of the model is significantly improved, and the classification accuracy of the model is all increased by more than 5%. Finally, among the 15 subjects, the average accuracy of the MBN model reached 82.47%, which was 2.90% higher than that of the MBN model without sample selection, confirming the effectiveness of selecting sample by this method for training.

In this session of the experiment, we used the $Branch_1$ network to calculate the subject similarity matrix and selected samples for training to avoid the negative improvement of the model caused by the excessively different subject data. In the end, this method improved the model performance while shortening the time required for model training.

4. Conclusions

Emotion recognition is an essential field of artificial intelligence, and its vast potential promotes the development of society. This paper provides a new framework based on Multi-Branch Network (MBN) for EEG-based emotion recognition, which promotes the analysis of cross-subject EEG signals. The MBN model extracts the background features and task features to jointly train the backbone network to achieve high classification accuracy in the cross-subject task. The average accuracy of the MBN model reached 79.57% on the SEED dataset, which is 20.89% higher than that of the SBN model. At the same time, the MBN model does not use the new-subject data in the training process, which improves the availability in practice. When the difference between some subjects in the training dataset and the new subjects is too large, it will lead to poor performance of the model on the new subjects. To this end, the similarity between the subjects is calculated by the background features, and a tiny amount of new-subject data is used to filter the training samples to improve the model performance further. After removing the data of the 3 most different subjects, the average accuracy of the MBN model reached 82.47%. Compared with other transfer learning methods, the proposed method uses less new-subject data, which shortens the time required for data collection in practice. At the same time, it is feasible to use deep learning to extract background features and calculate the similarity among subjects, which can be used for other studies.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61633010, U20B2074) and Key Research and Development Project of Zhejiang Province (2020C04009) and National Key Research Development Project (2017YFE0116800), Fundamental Research Funds for the Provincial Universities of Zhejiang (GK209907299001-008), Health Bureau of Zhejiang Province (2019PY054), and was also supported by Laboratory of Brain Machine Collaborative Intelligence of Zhejiang Province (2020E10010).

References

Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, pages 265–283, 2016.

Pietro Aricò, Nicolina Sciaraffa, and Fabio Babiloni. Brain–computer interfaces: Toward a daily life employment, 2020.

Benjamin Blankertz, Guido Dornhege, Matthias Krauledat, Klaus-Robert Müller, and Gabriel Curio. The non-invasive berlin brain–computer interface: fast acquisition of effective performance in untrained subjects. *NeuroImage*, 37(2):539–550, 2007.

Xin Chai, Qisong Wang, Yongping Zhao, Yongqiang Li, Dan Liu, Xin Liu, and Ou Bai. A fast, efficient domain adaptation technique for cross-domain electroencephalography (eeg)-based emotion recognition. *Sensors*, 17(5):1014, 2017.

Xuhai Chen, Zhihui Pan, Ping Wang, Lijie Zhang, and Jiajin Yuan. Eeg oscillations reflect task effects for the change detection in vocal emotion. *Cognitive Neurodynamics*, 9(3): 351–358, 2015.

Roddy Cowie, Ellen Douglas-Cowie, Nicolas Tsapatsoulis, George Votsis, Stefanos Kollias, Winfried Fellenz, and John G Taylor. Emotion recognition in human-computer interaction. *IEEE Signal processing magazine*, 18(1):32–80, 2001.

Xiaobing Du, Cuixia Ma, Guanhua Zhang, Jinyao Li, Yu-Kun Lai, Guozhen Zhao, Xiaoming Deng, Yong-Jin Liu, and Hongan Wang. An efficient lstm network for emotion recognition from multichannel eeg signals. *IEEE Transactions on Affective Computing*, 2020.

Ruo-Nan Duan, Jia-Yi Zhu, and Bao-Liang Lu. Differential entropy feature for eeg-based emotion classification. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 81–84. IEEE, 2013.

Laura Fiorini, Gianmaria Mancioppi, Francesco Semeraro, Hamido Fujita, and Filippo Cavallo. Unsupervised emotional state classification through physiological parameters for social robotics applications. *Knowledge-Based Systems*, 190:105217, 2020.

Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International conference on machine learning*, pages 1180–1189. PMLR, 2015.

Gary Garcia-Molina, Tsvetomira Tsoneva, and Anton Nijholt. Emotional brain–computer interfaces. *International journal of autonomous and adaptive communications systems*, 6 (1):9–25, 2013.

Aurélien Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, 2019.

Atefeh Goshvarpour and Ateke Goshvarpour. Eeg spectral powers and source localization in depressing, sad, and fun music videos focusing on gender differences. *Cognitive neurodynamics*, 13(2):161–173, 2019.

Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1735–1742. IEEE, 2006.

Yu-Liang Hsu, Jeen-Shing Wang, Wei-Chun Chiang, and Chien-Han Hung. Automatic ecg-based emotion recognition in music listening. *IEEE Transactions on Affective Computing*, 11(1):85–99, 2017.

Vinay Jayaram, Morteza Alamgir, Yasemin Altun, Bernhard Scholkopf, and Moritz Grosse-Wentrup. Transfer learning in brain-computer interfaces. *IEEE Computational Intelligence Magazine*, 11(1):20–31, 2016.

He Li, Yi-Ming Jin, Wei-Long Zheng, and Bao-Liang Lu. Cross-subject emotion recognition using deep adaptation networks. In *International conference on neural information processing*, pages 403–413. Springer, 2018.

Jinpeng Li, Shuang Qiu, Yuan-Yuan Shen, Cheng-Lin Liu, and Huiguang He. Multisource transfer learning for cross-subject eeg emotion recognition. *IEEE transactions on cybernetics*, 50(7):3281–3293, 2019.

Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *International conference on machine learning*, pages 97–105. PMLR, 2015.

Fabien Lotte, Marco Congedo, Anatole Lécuyer, Fabrice Lamarche, and Bruno Arnaldi. A review of classification algorithms for eeg-based brain–computer interfaces. *Journal of neural engineering*, 4(2):R1, 2007.

Yun Luo, Si-Yang Zhang, Wei-Long Zheng, and Bao-Liang Lu. Wgan domain adaptation for eeg-based emotion recognition. In *International Conference on Neural Information Processing*, pages 275–286. Springer, 2018.

Fangyao Shen, Guojun Dai, Guang Lin, Jianhai Zhang, Wanzeng Kong, and Hong Zeng. Eeg-based emotion recognition using 4d convolutional recurrent neural network. *Cognitive Neurodynamics*, 14(6):815–828, 2020.

Yi-Wei Shen and Yuan-Pin Lin. Challenge for affective brain-computer interfaces: Non-stationary spatio-spectral eeg oscillations of emotional responses. *Frontiers in human neuroscience*, 13:366, 2019.

Xiao-Wei Wang, Dan Nie, and Bao-Liang Lu. Emotional state classification from eeg data using machine learning approach. *Neurocomputing*, 129:94–106, 2014.

Xue-han Wang, Tong Zhang, Xiang-min Xu, Long Chen, Xiao-fen Xing, and CL Philip Chen. Eeg emotion recognition using dynamical graph convolutional neural networks and broad learning system. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1240–1244. IEEE, 2018a.

Yi Wang, Zhiyi Huang, Brendan McCane, and Phoebe Neo. Emotionet: A 3-d convolutional neural network for eeg-based emotion recognition. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE, 2018b.

Jingjie Yan, Wenming Zheng, Qinyu Xu, Guanming Lu, Haibo Li, and Bei Wang. Sparse kernel reduced-rank regression for bimodal emotion recognition from facial expression and speech. *IEEE Transactions on Multimedia*, 18(7):1319–1329, 2016.

Yilong Yang, Qingfeng Wu, Yazhen Fu, and Xiaowei Chen. Continuous convolutional neural network with 3d input for eeg-based emotion recognition. In *International Conference on Neural Information Processing*, pages 433–443. Springer, 2018.

Zixing Zhang, Bingwen Wu, and Björn Schuller. Attention-augmented end-to-end multi-task learning for emotion prediction from speech. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6705–6709. IEEE, 2019.

Wei-Long Zheng and Bao-Liang Lu. Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development*, 7(3):162–175, 2015.

Wei-Long Zheng and Bao-Liang Lu. Personalizing eeg-based affective models with transfer learning. In *Proceedings of the twenty-fifth international joint conference on artificial intelligence*, pages 2732–2738, 2016.

Wei-Long Zheng, Yong-Qi Zhang, Jia-Yi Zhu, and Bao-Liang Lu. Transfer components between subjects for eeg-based emotion recognition. In *2015 international conference on affective computing and intelligent interaction (ACII)*, pages 917–922. IEEE, 2015.

Wei-Long Zheng, Jia-Yi Zhu, and Bao-Liang Lu. Identifying stable patterns over time for emotion recognition from eeg. *IEEE Transactions on Affective Computing*, 10(3):417–429, 2017.

Peixiang Zhong, Di Wang, and Chunyan Miao. Eeg-based emotion recognition using regularized graph neural networks. *IEEE Transactions on Affective Computing*, 2020.