Expert Systems With Applications 185 (2021) 115581

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Cross-subject EEG emotion classification based on few-label adversarial domain adaption☆

Yingdong Wang [a,b,1], Jiatong Liu [a,2], Qunsheng Ruan [a,2], Shuocheng Wang [a,2], Chen Wang [a,*,2]

[a] School of Informatics, Xiamen University, No. 422, Siming South Road, Xiamen, Fujian, China
[b] Guangzhou Panu Polytechnic, School of Information Engineering, No. 1342, Shiliang Road, Panyu District, Guangzhou, Guangdong, China

ARTICLE INFO

Keywords:
Electroencephalogram (EEG)
Emotion classification
Cross-subject
Few label adversarial domain adaption

ABSTRACT

Emotion classification signal based on the electroencephalogram (EEG) is an important part of big data associated with health. One of the main challenges in this regard is the varying patterns of EEG indifferent subjects. Domain adaptation is an effective method to reduce the data difference between the source domain and the target domain. However, it is an enormous challenge to make a discriminator-based domain adaptation with a small target data and transform the target domain to the source domain. In the present study, a novel method called "few-label adversarial domain adaption" (FLADA) is proposed for cross-subject emotion classification tasks with small EEG data. The proposed method involves three steps: (a) Selecting subjects of the close source domain forming an adapted list. Few labeled target data are tested based on each emotion model of the source subject to get the subject list of the source domain. (b)Training three models based on each selected subject and the target subject. Three loss functions and six groups' dataset are designed to get a domain adaption model for each selected source subject. (c) Distilling all classifiers for classifying the target emotion. In general, the main purpose of the proposed method, which originates from the Meta-learning, is to find a feature representation that is broadly suitable for the target subject and source subject with limited labels. The proposed method can be applied to all deep learning oriented models. In order to evaluate the performance of the proposed method, extensive experiments are carried out on SEED and DEAP datasets, which are public datasets. It is found that with a small amount of target data, the proposed FLADA model outperforms the state-of-art methods in terms of accuracy and AUC-ROC. All codes generated in this article are available at github: https://github.com/heibaipei/FLADA.

1. Introduction

Studies show that the emotion recognition based on the electroencephalogram (EEG) is an essential issue in many fields (Gunes et al., 2011). In psychological researches, emotion recognition provides a quantitative reference for studying emotion-related behaviors. Moreover, in other applications such as brain-computer-interface (BCI), it promotes the human–computer interaction. Investigations show that in these applications, EEG is a more real way of the emotion detection compared with facial emotions (Cowie et al., 2001). In the medical field, the emotion recognition helps the diagnosis and treatment of various mental diseases such as autism spectrum disorders (Kuusikko et al., 2009) and depression (Joshi et al., 2013). Furthermore, the emotion assessment helps physicians and specialists to monitor the recovery process of patients. The long-term emotional recording and analysis helps people better understand their emotional state and make necessary adjustments timely. In E-learning education, the emotion recognition helps teachers to get student feedback simultaneously, which is of great significance to improve online education quality (Spuler et al., 2016).

☆ This work was supported in part by the Key Project of National Key RD Project (No. 2017YFC1703303); Natural Science Foundation of Fujian Province of China (No. 2019J01846, No. 2018J01555, No. 2017J01773); External Cooperation Project of Fujian Province, China (No. 2019I0001); Science and Technology Guiding Project of Fujian Province, China (2019Y0046).

The code (and data) in this article has been certified as Reproducible by Code Ocean: (https://codeocean.com/). More information on the Reproducibility Badge Initiative is available at https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals.

* Corresponding author.
E-mail addresses: wangyingdong@stu.xmu.edu.cn (Y. Wang), liujiatong@stu.xmu.edu.cn (J. Liu), 24320170155333@stu.xmu.edu.cn (Q. Ruan), scwang@stu.xmu.edu.cn (S. Wang), 24320182203277@stu.xmu.edu.cn (C. Wang).
[1] Researcher.
[2] Co-ordinator.

https://doi.org/10.1016/j.eswa.2021.115581
Received 22 February 2021; Received in revised form 6 July 2021; Accepted 7 July 2021
Available online 21 July 2021
0957-4174/

In all applications of the emotion recognition, it is essential to recognize the cross-subject emotions based on the EEG.

Reviewing the literature indicates that the experimental paradigm design (Chen et al., 2020), signal processing (Stojanovic et al., 2020), feature extraction and personal emotion classification model of EEG have been intensively investigated so far (Alarcao & Fonseca, 2019). However, some challenges remain unsolved for cross-subject EEG emotion recognition. It should be indicated that these challenges mainly originate from the individual differences and the non-stationary characteristics of EEG (Lin, 2019; Rodrigues et al., 2019). Accordingly, obtaining the EEG emotion classification model of a new subject is highly demanded. In this regard, the most common method is collecting a large amount of EEG data for new subjects. However, this training process costs at least half a day. Meanwhile, further investigations proved that models trained on small data are more likely to get poor results (Cheng et al., 2020). Only a few investigations have been conducted in this regard. More specifically, Zheng and Lu (2016) studied this issue in a nutshell.

Considering the foregoing shortcoming, it is intended to focus on the domain adaption method with multi-source subjects and limited labeled data for target subjects. It is expected that the proposed method can train the EEG emotion classification model in a short time with a small amount of EEG data so that it can be an appropriate scheme in practical applications. When a new user buys a new device (e.g. an emotional health data detection device), a correction is required prior to use. Therefore, the EEG of different subjects is referred to as an individual domain. In the cross-subject case, the target refers to the new subject with a few data with labels, while the source refers to the existing subjects. It should be indicated that since there are multiple sources, the proposed method is expected to solve problems with several numbers of source subjects and one target subject. On the other hand, considering few labeled data, the internal classification model should be fine-tuned prior to use.

For conducting the domain adaption, conventional methods extract versatile and subject invariant features with either signal processing techniques or deep learning models. In this regard, there are three paradigms as the following:

1. Selecting a large number of samples to extract common features for sentiment classification.
2. Using a domain adaptive method to map the target data to the source data and applying the source classifier for the classification.
3. Generating a personalized feature matrix based on the input data from different people, and using the optimized feature for the classification.

Ang et al. (2008) utilized the filter bank (FB) and common spatial pattern (CSP) for effective feature extraction followed by a fisher linear discriminator (FLD). Moreover, Fazli et al. (2009) adopted multiple classifiers on the top of the CSP and combined them with l1 regularized regressions to achieve improved performance. Moreover, scholars (Lan et al., 2019; Li et al., 2019, 2018) proposed unsupervised domain adaption paradigms and tried to find a feature function to satisfy this important goal through adversarial learning. In this method, the main purpose is to find a domain discriminator, which can effectively distinguish objects between different source domains. In this case, a classifier is trained with the standard cross-entropy loss. Once the discriminator is learned, adversarial learning is applied to update the function of the target feature extractor to confuse the discriminator. In other words, the adversarial training aims to train a feature extractor function to map the target sample to a feature space. Consequently, the discriminator no longer distinguishes the target sample from a source sample and then the results can be obtained from the source classifier. Song et al. (2020) generated a personalized matrix based on the input data and then optimized the feature through multi-dimensional graph convolution. Finally, they classified the area pooling features to achieve

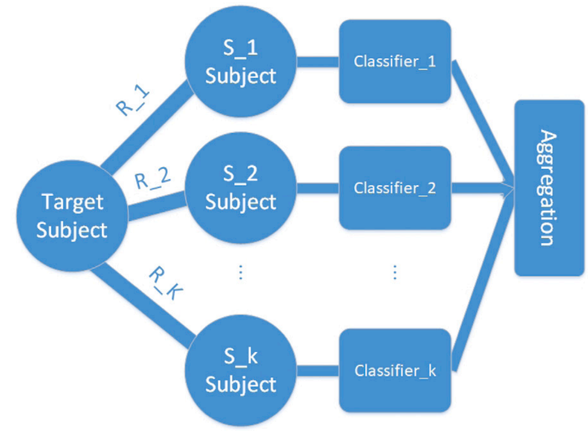

Fig. 1. Multi-subject domain adaption based on the few-label Adversarial Domain Adaption.

the emotion classification. Further investigations showed that although this method is an effective scheme to achieve higher results, it requires training all of the source data until it converges, and then the best transform parameters can be obtained. It should be indicated that this method has high requirements for initialization and learning rate, and requires specialists to adjust the system and achieve the best efficiency.

All of the reviewed methods can be applied only if distributions are represented by a sufficiently large dataset. More specifically, unsupervised domain adaption paradigms are in a position of weakness for small target source data. On the other hand, some unsupervised domain adaptation algorithms such as DAN (Ganin et al., 2016) do not need any target labels and can extract common features through the adversarial network (Li et al., 2019, 2018). These algorithms use the anti-domain adaptive method by the inverse gradient of the neural network. However, it is not possible in multiple subjects to determine which subject comes from which domain. Accordingly, they ensure that the common feature is extracted at the expense of high accuracy. It is worth noting that these methods always train all source subjects (Ma et al., 2019). However, emotional and personal characteristics of EEG are closely correlated, while the target domain may be very different from some data sources in the source domain. Therefore, applying all known subjects as source targets may lead to significant deviations in results.

In order to solve the above-mentioned problems, the idea in Li et al. (2019) is adopted in the present study. This idea is the style transfer mapping method based on selecting close models by testing small calibration data on each source subject model and choosing the source subject with high accuracy. Finally, top-k results are summed to determine the test data label of the new subject. Studies show that this method effectively improves the classification accuracy (Li et al., 2019).

In order to keep high migration currency with a small amount of labeled data inspired by the Meta-learning twin network (Motiian et al., 2017), multiple groups of data are generated to expand available data. In the present study, six groups of datasets are performed. Details are presented in Fig. 3. Since condition possibility of the same emotion has the same labels while different emotion data have different labels, the target model is trained through three loss functions. Finally, the results are obtained by adding the pre-trained model possibility with the same weights to $\frac{1}{k}$.

Fig. 1 presents the multi-subject domain adaption based on the few-label learning. It indicates that top-k subject data is initially selected as the source domain data. Then pairs are generated from the source and domain data to train a shared feature model for each source domain. Finally, these k classifiers are aggregated to obtain the possibility of the final emotion. Contribution of the present study can be summarized in three points as the following:

1. Multiple groups are formed to ensure the same semantics between source domain subjects and tackle the small target data. Obtained results demonstrate that the accuracy of the proposed method outperforms that of state-of-the-art methods.
2. A sharing feature extractor is proposed for the space between the target and source subjects. The proposed extractor significantly reduces the training time and the number of parameters. It is found that the simple addition distillation method effectively improves classification accuracy.
3. The proposed method facilitates the fast acquisition of emotion recognition models for new users, which is of great significance in fast-deployment scenarios. Compared to conventional methods, the proposed method requires fewer labeled samples during the verification process.

The present article can be organized as follows: A comprehensive literature survey is presented in Section 2. In Section 3, the method for cross-subject EEG emotion classification is described based on few-label adversarial domain adaption. The dataset and results are presented in Section 4. Initial parameters is important for result of classification (Cheng et al., 2021; Tao et al., 2020; Wei et al., 2021). In Section 5, the impact of other factors on the results is investigated. Finally, the achievements and contributions of the present study are summarized in Section 6.

2. Related works

2.1. Siamese neural networks

Despite the significant advantages of end-to-end neural network models, they need a large number of samples to achieve reasonable results. On the other hand, when a human sees a kitten or bird for example once or twice, they can be easily distinguished in the next encounter. In this regard, different meta-learning techniques have been proposed to learn new concepts and skills with little data. More specifically, Koch et al. (2015) proposed Siamese neural networks for one-shot image recognition. In this scheme, a Siamese network is initially trained to judge whether the two images belong to the same category or not; then the test data with the support set distance is computed during testing and the image label with the highest similarity is selected as the final label. Although the main purpose of this method and the proposed method are different, this idea is also applied in the present study. Accordingly, six groups of data are formed with respect to labels and domain so that the classifier can simply adapt to both domains. Moreover, the method makes full use of small target samples.

2.2. EEG emotion feature (EEG)

In order to facilitate the EEG emotion classification, Li et al. (2020) proved vital features of differential entropy (DE) as an extension for Shannon entropy to measure the complexity of a continuous random variable. Further investigations show that the DE method outperforms conventional Power Spectral Density(PSD) and EEG emotion recognition methods from accuracy and stability points of view (Laureanti et al., 2013). Consequently, the DE method is adopted in the present study as a feature. Studies show that when the variable X obeys the Gaussian distribution $N(\mu, \delta^2)$, the DE will get the max entropy. Therefore, changing the data into Gaussian distribution, the DE method can be mathematically expressed in the form below:

$$DE = -\int_{-\infty}^{+\infty} p(x) log p(x) dx = \frac{1}{2} log(2\pi e \delta^2) \tag{1}$$

where X is a random variable and $p(x)$ denotes the probability density function of X.

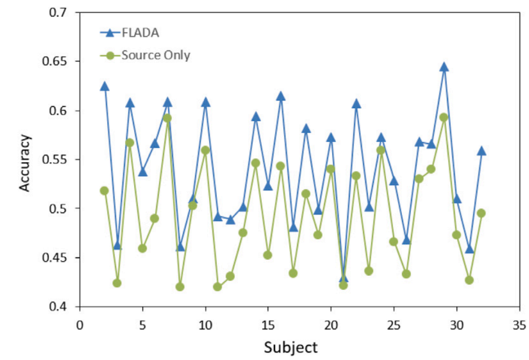

Fig. 2. Results of S01 non-transfer (Source Only) and transfer methods (FLADA) on 31 source domains individual.

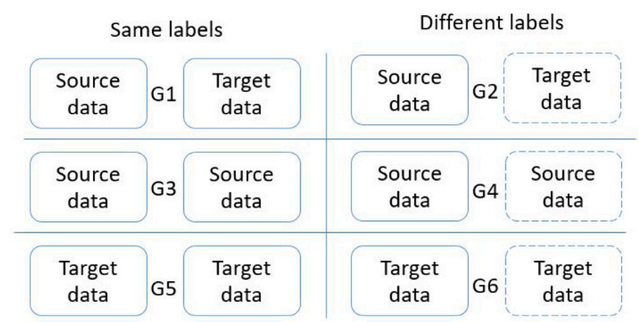

Fig. 3. The content of six groups.

3. Method

Details of the proposed method are presented in Fig. 1. It indicates that the proposed method has four main steps. Firstly, top-K source domains are selected from data for source subjects according to the calibration data test, which results in different source models. Secondly, the calibration data is combined with the selected domain data to generate six groups of data for each domain subject. Based on generated groups, three models, including feature extractor, classifier, and discriminator, are trained. Details of six groups are presented in Fig. 4. Thirdly, six generated groups of data are applied to assure that the same emotions have the same labels. Having common features as a condition possibility, groups of data are trained with cross labels to confuse the boundary between different domains. Finally, results of the top-k model are added to get the final emotion possibility.

3.1. Top-k source domain models

In order to determine the top-k source-domain list, which are closer to the target domain, calibration data are tested by different source subject models. Conventional methods apply different metrics such as maximum mean discrepancy (Gretton et al., 2006), Kullback Leibler divergence (Kullback & Leibler, 1951), Jensen–Shannon divergence (Goodfellow et al., 2014) to calculate the distance between distributions. The effectiveness of these methods has been proved in diverse applications. However, these methods require remarkable calculations and high computational expense in high-dimensional features. It is worth noting that many other indicators have been proposed so far in this regard. For example, Fig. 2 presents the results of S01 in the DEAP dataset transform from the other 31 source subjects and that of the source only (SO). It should be indicated that the source only means that the results from the model are trained on the source subject and then they are tested on the target subject. This is the main idea in non-transfer methods. FLADA is the result of classification by few-label domain

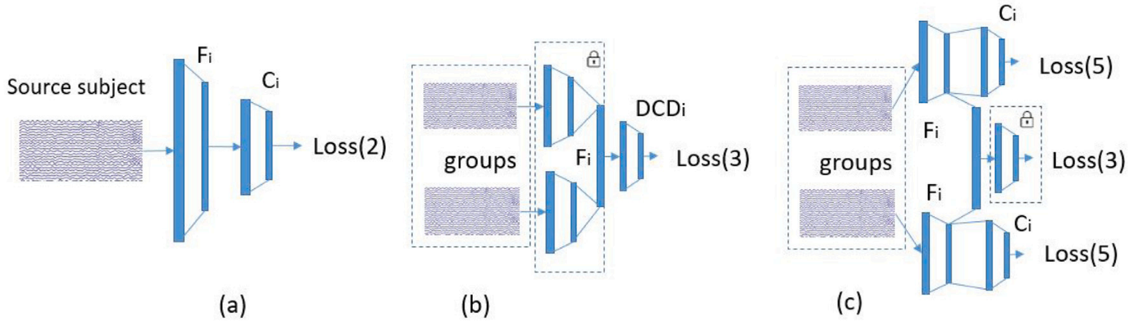

Fig. 4. Few-label adversarial domain adaption structure. The proposed network shares feature extractor ($F_s = F_t = F$) and includes three parts: (a) Establishing the source model of the feature extractor F_i and classifier C_i, (b) With six-group data, freezing the F_i model, training domain-class discriminator DCD_i model (six domains). (c) freezing the DCD_i model and updating F_i parameters with loss (see Eq. (4)).

adaptation. It indicates that there exists a strong correlation between the non-transfer and transfer learning results. Accordingly, the source-subject data of the present study is selected according to that of the non-transfer result. In this case, a source list can be obtained according to the non-transfer score from high to low.

Details of selecting process are as follows: First, all source subject models are obtained. Source subject models consist of two parts, including the feature extractor and the classifier. Moreover, $S_1 \sim S_K$ and $S_i \epsilon S$, indicating that all subjects are selected from known data. Each S has a train dataset made of several pairs $S_k = (x_i^K, y_i^K)_{i=1}^k$. In order to extract more task discriminative features and learn an accurate classifier, the feature extractor F_i and the classifier C_i is considered for each labeled source domain S_i with an unshared weight model. More specifically, in a N-class classification task, F_i and C_i can be optimized by minimizing the cross-entropy loss in the form below:

$$Loss(F_i, C_i) = -E_{(x_i, y_i)}\ p_i \sum l_{n=y_i} log(\sigma(C_i(F_i(x_i)))) \tag{2}$$

where σ and l denotes the softmax function and an indicator function, respectively. Compared with the model that uses common parameters to extract features in each domain, non-shared parameters of the feature extractor have significant feature embedding and accuracy in different domains. After obtaining all source-domain models, labeled data with the limited target are tested on each source subject model and then the top-k of F_i and C_i models are saved according to the score for being used in the next step.

3.2. Consistent information

The source list was prepared in the foregoing sections. In this section, it is intended to select the top-k sources, which are closer to the target, to perform the domain adaptive learning. In this case, the problem transforms into the adaptation of the target subject domain to each source domain as well as obtaining a classifier. For a target domain adapted to one source domain, the proposed method is inspired by the siamese network (Koch et al., 2015) in the meta-learning. It is worth noting that the siamese network produces a pair of samples from the source data to form positive and negative samples. The pair of samples having the same labels belongs to positive samples, while those with different labels produce negative samples. Finally, these paired samples are applied to train a model. The more the similarity of the paired samples, the closer the output of the model to 1. On the other hand, the greater the difference, the closer the output to 0. In other words, a model is established for evaluating the similarity between different samples so that the similarity between the sample in the query and the sample in a few-shot can be calculated. Then it is possible to determine which class the query belongs to. Moreover, pairs of samples are created, but the pair of data come from the target subject and source subject.

Fig. 3 indicates that in the proposed method, six groups of pairs (i.e. G_i, i = 1, 2, 3, 4, 5, 6), including three positive pairs and three negative pairs, are generated. The positive pairs in groups 1, 3 and 5 share the same labels, where group 1 consists of samples from the source subject, while pairs of group 3 contain source subjects and labeled target subjects. Meanwhile, pairs of group 5 come from source subjects. On the other hand, negative pairs in groups 2, 4 and 6 have different labels. The pairs of group 2 consist of samples from the source subject, while pairs of group 4 contain source subjects and target subjects. Meanwhile, pairs of group 6 come from source subjects.

In order to allocate the same amount of members for each group, the number of group pairs is considered the same as the number of labeled target data. In classical domain adaption, many methods can be applied to learn a domain discriminator, which contributes to training a feature extractor and confuse the target and source subjects. Since only semantic information is considered in the present study, a multi-class discriminator or domain-class discriminator (DCD) is applied to introduce the semantic alignment between the source domain and the target domain. Moreover, a multi-class classifier instead of a binary classifier is introduced to determine which sample pair belongs to which group. To this end, Softmax activation is utilized in the last layer to model the DCD with two fully connected layers and then a standard classification cross-entropy loss can be used for training.

$$Loss(DCD_i) = -E[\sum_{i=1}^{6} y_{group_i} log(D_i f(F_i))] \tag{3}$$

where y_{group_i} is the label of $group_i$ and D_i denotes the DCD function. Moreover, f is a function that concatenates results from F_i of the source subject and that of the target subject. In the next step, the distribution of different domains should be considered to guarantee the semantic consistency in the previous period. To this end, F_i is updated to confuse the DCD so that the discriminator can no longer be distinguished between groups 1, 3 and 5, and groups 2, 4 and 6. The loss of discriminator can be expressed as follows:

$$\begin{aligned} Loss(D_i) = -E[y_{G3} log(D(f(F(G1)))) + y_{G3} log(D(f(F(G_5)))) \\ + y_{G4} log(D(f(F(G_2)))) + y_{G4} log(D(f(F(G_6))))] \end{aligned} \tag{4}$$

3.3. The target model

After obtaining six types of discriminators, the next training condition is set as the following: features from the F_i feature extractor help the same-labeled data share consistency information. In this case, the loss function can be rewritten in the form below:

$$\begin{aligned} Loss(All) = -\alpha * Loss(D_i) + E[Loss(C_i(F_i(X_s)), Y_s)] \\ + E[Loss(C_i(F_i(X_t)), Y_t)] \end{aligned} \tag{5}$$

where α balances the classification and confusion. It is worth noting that misclassifying pairs show that DCD_i can no longer distinguish different distributions from the source subject or target subject, while the last two loss make sure that $classifier_i$ could discriminate positive pairs from negative pairs. Considering the low number of labeled target samples, unsupervised domain adaptation only focuses on the confusion of the domain so that it cannot be applied to solve the class separability.

Table 1
The details of the two device.

System	ESI Neuroscan	Bio-semi ActiveTwo
Dataset	SEED	DEAP
Component	SynAmps1 amplifier, E-prime stimulation system, 64-channel Quick-Cap electrode caps.	Neurobehavioral, Active electrodes, AD-box, USB2 Receiver.
Channels	62	32
Link method	Wire	Wire
Conductive paste	Need	None
Sampling frequency	1000 Hz	512 Hz
Preparation time	Long	Short

3.4. Adding all features

$$pred = Sigmoid(\sum_{k}^{i} w_i * F_i) \tag{6}$$

After three consecutive steps, each source subject data produces one feature model. In this step, it is necessary to sum all features together through Eq. (6), where w_i is the weight of each source model. It should be indicated that w_i can be considered as a combination of different data. For example, $w_i = 1/k$ refers to a geometric sequence or arithmetic sequence based on the position in the source list.

3.5. Training process

During the training process, F_i and G_i extractors in the target subject and source subject share the same parameters. Accordingly, since the feature extractor F_i reflects the feature from the source subject and target subject in the same space, then C_i fit both source subject and target subject, respectively.

Algorithm FLADA

Input: The source subject data and the target subject data S_i, the weight of DCD α, the number source data k
Output: The prediction Label of target data
1: Train F_i and C_i for each S_i using (2).
2: Get source list L_k according to the score tested by target labeled data.
3: for i = 1 to k do:
    Form sample G1, G2 from source $subject_k$, target subject.
    Form sample G3, G4 from source $subject_k$, source $subject_k$.
    Form sample G5, G6 from target subject, target subject.
    Train DCD_i w.r.t $F_t = F_s = F_i$ using (3)
    While not convergent **do**
    Update F_i and C_i by minimizing (5)
    Update DCD_i by minimizing (3)
    end while
4: end for
5: prediction $y = f(\sum w_i * F_i)$

Once the feature extractor functions F_i and the classifier C_i are chosen, the following steps should be taken:

1. F_i and C_i are initialized with the source dataset D_s.
2. The source list is obtained by the score.
3. Six groups of pairs should be created with source subject and labeled target subject.
4. Train DCD with six groups of pairs when freezing F_i.
5. The feature extractor F_i and classifier C_i should be updated to confuse DCD and maintain high classification accuracy. During the training, DCD should be frozen.
6. To keep F_i and maintain the reasonable performance of classifier C_i, DCD should be trained again with different batch sizes and learning rates. Then steps (5) and (6) should be repeated until the loss convergences to the preset value.

Details are presented in Algorithm 1 and Fig. 4.

4. Experiments

4.1. Datasets

In order to evaluate the effectiveness of the proposed method, two public databases, including DEAP[3] (Koelstra et al., 2012) and SEED[4] (Duan et al., 2013) are adopted in the present study. It should be indicated that both datasets are widely used in the EEG-based emotion recognition (Tripathi et al., 2017; Yin et al., 2020). In the DEAP database, 32 healthy participants participate (50% of women) aged between 19 and 37 (average age 26.9) in the experiment. Each participant was asked to watch 40 one-minute music videos, while their brain electricity is recorded during the experiment. The published EEG is applied to remove artifacts and finally, 32 channels with a sampling frequency of 128 Hz are obtained. After the experiment, participants score each video from different aspects, including the total Valence, Arousal, Dominance and Liking. Each score is an integer within the range of 1 to 10. For example, the 1.0 of Valence means very sad, and the 10.0 of Valence means very happy. The data recorded by each participant includes 40 EEG data segments and the corresponding labeled EEG data. Each segment includes a 60 s experimental signal and a 3s baseline signal in a relaxed state.

In the SEED dataset, 15 subjects (7 males and 8 females; MEAN: 23.27, STD: 2.37), participate in the experiment. EEG signals are recorded when they watch 15 Chinese four-minute video clips. These videos elicit three types of emotions, including neutrality, positivity and negativity. For each subject, this recording process is repeated three times at different dates, and different intervals vary from there days to one month. Accordingly, three sessions are obtained for each subject, where each session contains 15 trials of EEG signals, 64 channels with a sampling frequency of 200 Hz.

The two devices used in the dataset are different, and the specific parameters are shown in 1

4.2. Experimental design

In both DEAP and SEED datasets, the window-size of 1s is set as the sample data. Alnafjan et al. (2017) indicated that the duration of emotion is about 0.5–4 s, while other scholars reported 1s-segment in this regard. Moreover, five bands, including delta waves (1–4 Hz), theta waves (4–8 Hz), alpha waves (8–15 Hz), beta waves (15–30 Hz) and gamma waves (30–50 Hz) (Rozgic et al., 2013), are applied to compute the DE of both datasets.

In the DEAP database, one session of a subject has 60 s data of 32 channels and there are 40 sessions for each person. Therefore, one subject data could be segmented into 2400 samples with a size of 32×128 through a non-overlapping window. In the present study, the moving average techniques are applied to smooth the DEAP data, while the window size is set to 3. Therefore, the DEAP feature that forms

[3] http://www.eecs.qmul.ac.uk/mmv/datasets/deap/.
[4] http://bcmi.sjtu.edu.cn/~seed/seed.html.

a 160-D vector has 32-channel × five-band. For the label processing, each segment of the signal, which is divided by the sliding window, is labeled with the same label as the whole segment of the signal. The present research is mainly focused on the valence. The valence of the scale is between 1 and 9 and it is divided into two binary classification problems with a threshold of 5, high/low arousal and high/low valence ($low \leq 5$, $high > 5$). Therefore, the final classification can be considered as a binary classification.

For the SEED dataset, the preprocess is similar to that of the DEAP dataset, where each 1-s segment is decomposed with 200 points using the short-time Fourier transform and non-overlapped Hanning window so that one sample size is 200 × 62. Data in an open-access platform includes DE features, PSD features and so on. The DE feature has been smoothed with the linear dynamic system (LDS), which is an effective approach to use of the time dependency of emotional changes and filter out emotion-unrelated EEG components (Shi & Lu, 2010; Zheng et al., 2019). Similar to the DEAP dataset, the label of each segment is divided by the sliding window with the same label as the whole segment of the signal. Therefore, the final classification can be considered as a three-classification task.

In this article, the most simple multilayer perceptron (MLP) is applied to implement the F, C, and DCD in the model. The model structure of the DEAP dataset has five layers. The network is 160-128-64 (feature extractor), 32-2 (classifier), and 32-6 is the DCD network. For the SEED dataset, the model structure has six layers, the network is 310-256-128-64 (feature extractor), and the 32-3 (classifier) and 32-6 is the DCD network. Both models apply the Relu function in each layer to accelerate the convergence and SoftMax function in the last layer of the classifier and DCD network. Moreover, more layers are explored to generate the feature extractor. For example, a 310-256-225-186-124-96-64 as the feature extractor, 32-3 (classifier) model to conduct experiments with the SEED dataset. Obtained results show that although the training time is long, the corresponding accuracy decreases. This issue will be discussed in detail in the discussion section. The learning rate of the DEAP and SEED are 1×10^{-3} and 1×10^{-4}, respectively. Furthermore, the default values for α and batch size are 0.7 and 30, respectively. With the golden feature, the source models convergence quickly, so the train size in the first epoch of the SEED is set to 10, while that in the DEAP is set to 5. The next discriminator epoch in SEED and DEAP are set to 150 and 80, respectively. Moreover, the final epoch in SEED and DEAP are 80 and 50, respectively. All experiments are conducted in Pytorch libraries with an NVIDIA GeForce GTX 1080 GPU.

4.3. DEAP results

4.3.1. Comparing the results with other methods

Table 2 compares results obtained from different methods. It should be indicated that the source-only results are trained by the source subject data, while they are tested by the target subject data. It is observed that the obtained results are at least 10% less than those from other methods. Then DANN, MMD and ADA methods are trained by the zero-shot-adversarial domain adaption, which does not need any labeled target data. They extract the common feature by the Adversarial network, but the corresponding accuracy is relatively low. This may be attributed to too many factors to consider in this method. In D-tf and RA-MDRM methods, the transfer learning methods are used. It should be indicated that the D-tf method is mainly focused on the transform network design, while the other one is focused on the data represented through spatial covariance matrices of the EEG signals. Both methods exploit the recently proposed techniques based on the Riemannian geometry (Caselles et al., 1995) of the manifold of the symmetric positive definite (SPD) matrices, and both methods improve the classification accuracy of EEG emotion. However, these methods cannot make full use of all data completely. The MUPS and proposed method apply the meta-learning so that the MUPS apply the MAML (Finn et al., 2017) as the base method to find the most appropriate model to adopt the new data rapidly. However, the MUPS method requires a significant amount of tasks to learn the model. In all studied approaches, 90 calibration samples are applied for each emotion label and the time of test EEG is set to three minutes. Meanwhile, the same parameters are applied in all algorithms. Obtained results show that the proposed method improves the accuracy by at least 0.8% when the K = 6 and the α = 0.3. The variance is relatively small, and the result is relatively stable.

Table 2
Mean accuracies (ACC) and standard deviations (STD) of DEAP.

Method	Accuracy	ROC-AUC
Source-only	40.2(18.9)	45.6(14.8)
DANN (Radford et al., 2015)	50.5(9.8)	52.5(10.4)
MMD (Sejdinovic et al., 2013)	48.8(12.4)	53.3(8.9)
ADA (Haeusser et al., 2017)	55.9(13.1)	56.7(5.5)
D-tf (Tan et al., 2018)	63.8(0.1)	76.7 (0.1)
RA-MDRM (Zanini et al., 2018)	61.4(0.1)	75.8(0.1)
MUPS (Duan et al., 2020)	67.2(0.06)	78.2(0.05)
Proposed method	68.0 (0.05)	65.5(0.06)

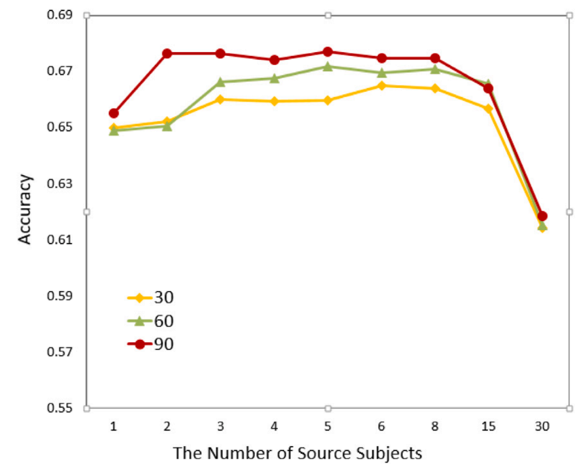

Fig. 5. Mean accuracy achieved with different source numbers.

4.3.2. Source number evaluation

In the multi-source domain adaption, the source number N is an important factor. More sources indicate integrating more classifiers to predict the target subject emotions. However, considering a weak correlation between the target subject and some source subjects, blind increasing of source subjects does not improve the accuracy and causes computational burden (Yao & Doretto, 2010). In order to evaluate appropriate source numbers, the accuracies achieved with the proposed model under different source numbers are compared. Fig. 5 shows that as the source number increases from 1 to 3, the accuracy of the three situations increases rapidly. This indicates more sources encourage better performance. From the number of 3 to 6, the accuracy generally stays stable, and after 6, the accuracy decreases because of involving more unrelated source data. The more the correction data, the smaller the N to obtain the best results. When K = 5, the obtained result is steady. Therefore, K is set as 5 in the following experience.

4.3.3. The trade-off of DCD loss

In order to investigate the importance of mapping different data sources to the same space, the results of α are tested with different weights in Eq. (5). Fig. 6 compares results obtained from α with different values. The value of 0.3 obtains the highest accuracy. Moreover, with the 0 two subjects get very lower accuracies and the AUC. During training, AUC goes wrong because of the single class result.

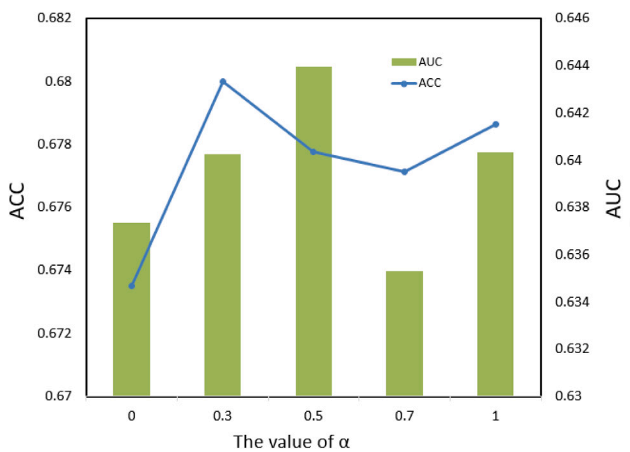

Fig. 6. Performance comparison of α with different values.

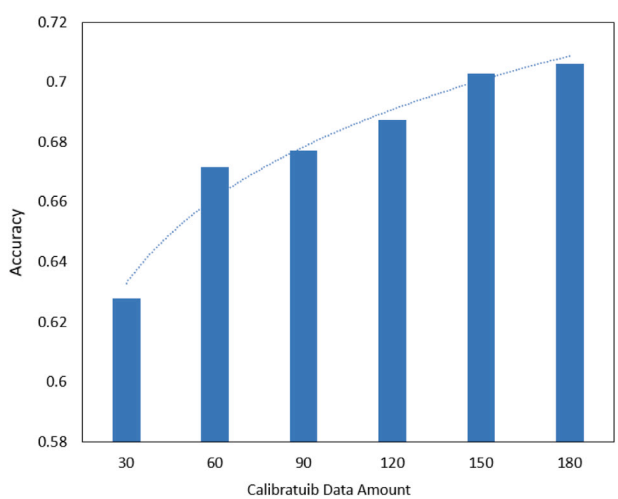

Fig. 7. Performance comparison of the proposed method with different calibration data amounts.

4.3.4. Amount of calibration data

The target domain has few labeled calibration data that play an important role in both source selection and Meta-learning. They are randomly picked out with a balanced number for each class. Fig. 7 shows the emotion classification accuracy with different numbers of calibration data. The posterior analysis of the proposed method shows that as the calibration data increase (30-60-90-120-150), the accuracy improves significantly, and remain stable afterward.

4.4. The SEED results

Every subject is recorded three times. In order to evaluate the proposed method, one of them is randomly selected in each subject as the experimental data. It should be indicated that there are fifteen subject data. Where one is set as the target data and the remaining 14 people data are set as the source data. In the target data, there are fifteen sessions. Three sessions from the three emotions are set as the calibration data, while the last twelve sessions as are considered as the test data. Table 3 presents the comparison results. It should be indicated that the source-only is the same with DEAP, the model trained by the source data, and tested by the target data. The next two methods are zero-shot by the domain adversarial with different base models. DANN applies the simple DNN as the simple network, and R2G-STNN uses the hierarchical spatial–temporal neural network for emotion recognition. In both methods, it is intended to learn the

Table 3
Mean accuracies (ACC) and standard deviations (STD) of SEED.

Method	Accuracy	ROC-AUC
Source-only	60.27(18.89)	65.6(14.8)
DANN (Ganin et al., 2016)	75.08(11.18)	86.5(10.4)
R2G-STNN (Li et al., 2019)	84.16(7.63)	85.7(5.5)
DGCNN (Song et al., 2018)	79.95(9.02)	81.3(8.9)
IAG (Song et al., 2020)	86.30(6.63)	89.7(5.5)
MTL (Li et al., 2019)	84.2(0.06)	86.3(0.06)
Proposed method	89.32(0.86)	97.52(0.63)

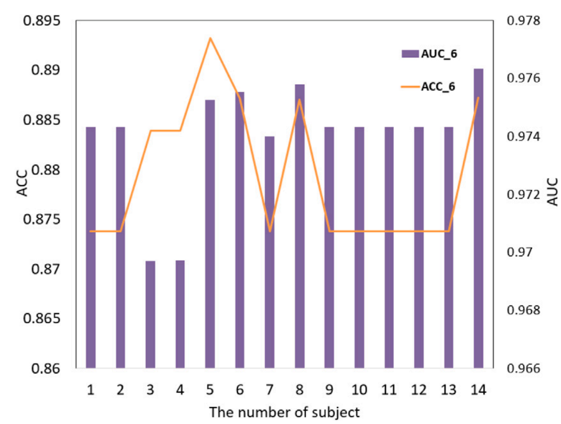

Fig. 8. Mean accuracy for different source numbers.

common features from all of the subjects. It is worth noting that emotion is a combination of unique features and a common feature. During the investigation of common features, the subject emotion classification accuracy may decrease because of omitting the unique features. The next two methods generate dynamic graphs based on the input data, using GCN to optimize features, and final classification based on the optimized features. They are the state-of-art methods in the emotion classification. The proposed method improves 5% on the accuracy and almost 4% on the AUC.

4.4.1. Source number evaluation

Fig. 8 shows the accuracies achieved by the proposed model under different source numbers with k = 0.7. It is observed that as the source number increases from 1 to 5, the corresponding accuracy increases and when k = 5 the corresponding reaches the highest value. However, when the source number exceeds five, the accuracy slightly decreases. The source numbers more than nine, an increment in the accuracy is not observed and it almost stays steady. In other words, considering the SEED dataset, the individual differences of 15 individuals are relatively small compared with that of the DEAP dataset.

Compared with the accuracy rate, the value of the AUC is easier to interpret. During the period from 1–5 to the source subject, the value of AUC is not steady. When the source number exceeds 5, the result of AUC stabilizes with a small change. It is the same as the accuracy rate, and the maximum value is obtained at 8. However, the difference is about 0.0001 compared with that of 5 sources.

4.4.2. Trade-off of DCD loss

In this section, the results of different weights of DCD in the SEED dataset with the same parameters are compared. It should be indicated that the high value of the DCD weight indicates a more important common space. Fig. 9 shows the SEED mean accuracy for different weights of DCD. It is observed that when the weight is less than 0.3, as the weight increases, the accuracy rate continuously rises. It is found that the highest accuracy is obtained for a weight of 0.7. Moreover, when weights exceed 0.3, as the weight increases, the quasi-delivery rate in the later period fluctuates. However, the change is not obvious.

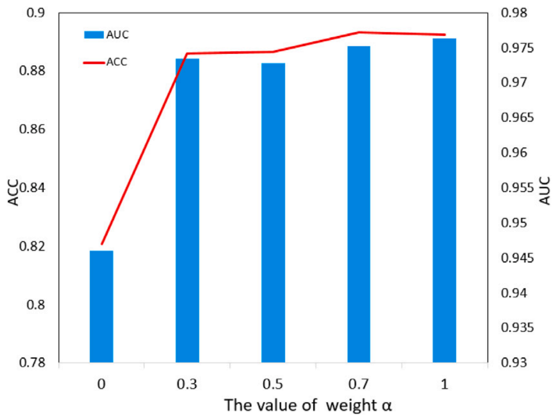

Fig. 9. Mean accuracy achieved with different weights of DCD.

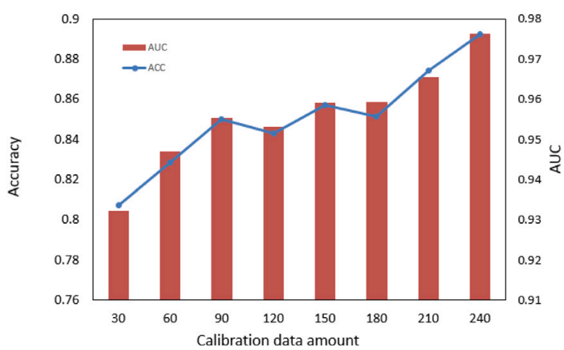

Fig. 10. Different calibration data of SEED performance result.

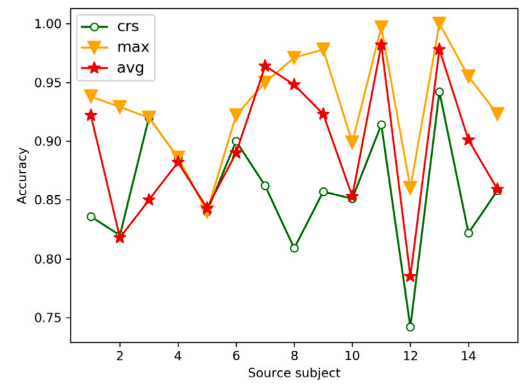

Fig. 11. Mean accuracy for each subject in the SEED dataset in the three cases.

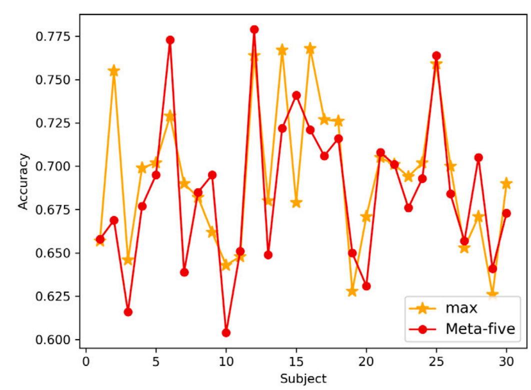

Fig. 12. Mean accuracy for each subject in DEAP dataset in the two cases.

4.4.3. Amount of calibration data

The target domain obtains a few labeled calibration data. On the other hand, these data play an important role in both source selection and Meta-learning. They are randomly selected with a balanced number for each class. In the SEED dataset, each session contains four minutes of EEG. For comparison, three sessions are used as source targets. This is a bit longer compared to the overall three minutes of the DEAP dataset. Therefore, in the later stage of verification, regarding the division of data time, a shorter time is emphatically considered. Fig. 10 compares the accuracy and AUC of the SEED dataset with different calibration data. It is observed that until the number of calibration data is less than 90, accuracy and AUC increase rapidly as the calibration data increases. It should be indicated that 90 to 180 calibration data is used in the present study. Moreover, it is found that with the further increase of the calibration data, accuracy and AUC relatively stabilize and enter a stage of slow growth. Finally, when the number of calibration data exceeds 120, the accuracy and AUC increase as the original speed increases, reaching the maximum value of 240.

5. Discussion

5.1. Max-result and average-result

In order to understand whether the proposed algorithm obtains the closest data source, the highest result is compared with the results of the proposed method. The highest result is the max result of the single-source subject when transfer learning happens. Fig. 11 illustrates the details of SEED comparison. In the experience, the same mount calibration data of 240 is used. Moreover, in the avg method, only one parameter $k = 5$ is changed. The green line shows the single-source domain adaption result, and the single source comes from the highest score source when no-transfer, that also the case when our method with $k = 1$. It is observed that the obtained results almost exceed all corresponding results, while the highest results are better than the mixture feature. Moreover, it is found that the emotional classification gap between different people is relatively large, and the twelfth person has the worst result in three cases.

Fig. 12 shows the details of the DEAP dataset comparison. In the experiment, the same parameter $K = 5$ is used, the verification data is three minutes. It is observed that the differences between different people are very large. All the maximum values exceed the fusion based on the five data subjects. Predictive value. Therefore, there is still a lot of work to be done here. How to go beyond the best results and simply merge the results is not the best method to automatically learn weights. Zhao et al. (2020) used the distance between source and target as the weight for each source feature. Further investigations are required to prove the performance of this idea.

5.2. Different layers of the feature extractor

In the proposed method, the base models are simple fully connected networks. In the DEAP dataset verification, there are three models. It should be indicated that for the feature extractor, only three layers fully connected network is used. Moreover, for the classifier, only two layers network is used and two layers are applied in the CDC classifier. In the SEED dataset verification, five layers network is utilized as the feature extractor. It is worth noting that the same parameters are applied in the other model. The SEED dataset obtains almost twice the features of the DEAP data as input. Therefore, in the feature selection, a multi-layer network is used. Fig. 13 shows the difference between the results of a five-layer network and a three-layer network. Five-layer network results surpass the three-layer network results by 0.2%. In other words, an appropriate model could improve the accuracy of the domain adaption. In the future, the results can be improved with an appropriate deep network model.

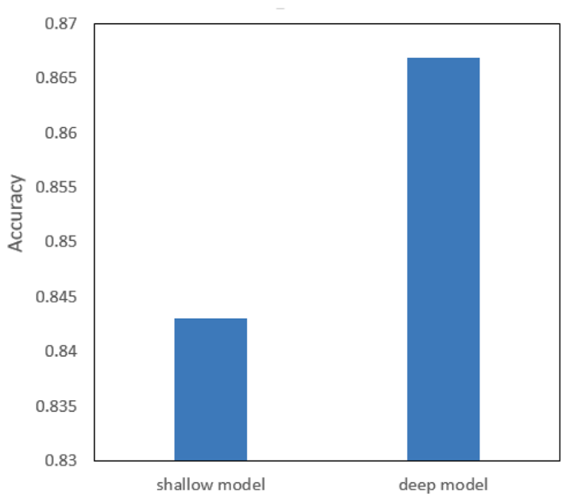

Fig. 13. Mean accuracy of SEED with different layers in feature extractor model.

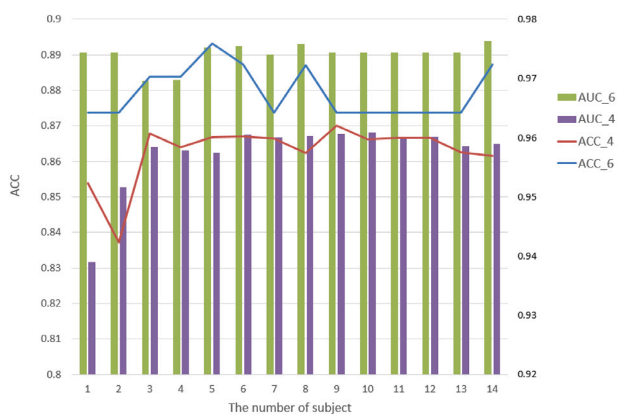

Fig. 14. Mean accuracy of SEED with four-group and six-group.

5.3. Four-group and six-group

In the process of group forming, a diverse number of groups, including four and six groups, can be applied. In order to investigate the influence of group numbers on the obtained results, the classification results of four and six groups are compared in the present study. It is worth noting that four group members, entitled G1, G2, G3 and G4, are part of six group data. The training process is the same as the six-group experience. The learning rate and the alpha value are set to 0.0001 and 0.7, respectively. Moreover, the three epoch numbers are 10, 150, and 80. The comparison results are shown in Fig. 14. It is observed that the classification of EEG emotion of ACC and AUC with six-group is much greater than that of four-group, even with different k values. When k is set to 5, the six-group-based method gets the highest ACC result of 0.893, which is 2.6% higher than the four-group of 0.867. Meanwhile, the six-group-based method of AUC is 0.975, which is 1.7% higher than that of the four-group-based method of 0.958. In other words, when the target subject-semantically consistent information is included, the common feature extraction model and classification model of target and source subject improve more.

5.4. Data agitation

In the last step of the training process, interlaced tags are initially applied as the labels of the distinguisher in the proposed method to optimize the feature extractors and classifiers. As a result, better

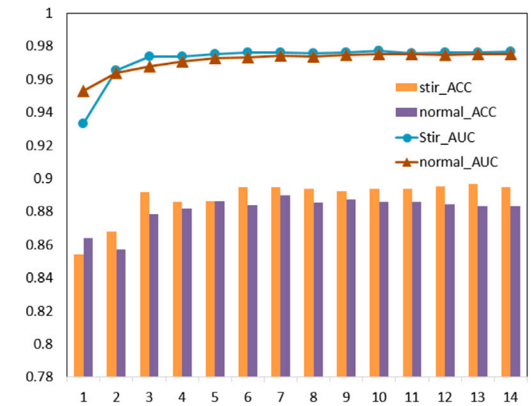

Fig. 15. The accuracy(ACC) and AUC of SEED with data agitation.

feature extractors and classifiers can be obtained for each source subject domain. In this case, the classifier cannot distinguish which sample comes from which domain, thereby making the model more robust to new data. In this process, different domains should be mapped into the same space to distinguish which sample comes from which domain. However, this process adversely affects the model accuracy. In order to ensure that the feature extractor and classifier have high semantic analysis capabilities, the distinguisher should change with the variation of the classifier to ensure high-precision semantic consistency. Therefore, after updating the feature extractor and classifier, the distinguisher should also be optimized. However, each set of data for these six categories has only 240 data. It means that this data overfits the training requirements of the neural network. Therefore, in the first step of the subsequent cross-label correction, interference error data is added to reduce the overfitting issue, and the model is corrected in the next step by the six-distinguisher training with different learning rates. This technique improves the performance of feature extraction and classifier, thereby improving the method accuracy up to 0.90. It should be indicated that G1 and G2 share the same label (see Fig. 15).

6. Conclusion

The EEG pattern variability across different subjects is a major challenge for the cross-subject EEG classification. In the present study, an efficient few-label domain adaption based on the multi-subject learning model is proposed. The output method includes four steps: (1) Train each source subject and use the target data to test the source model and obtain a list of source subjects. (2) Train the discriminator with the semantic shared feature model. (3) With the help of the frozen DCD, update the feature extractor and classifier to map the target subject and source subject in the same space and share the same feature extractor and classifier. (4) Aggregate the final classification results to further improve the classification accuracy. The four steps of the few-label domain adaption task enable the model to learn general characteristics based on multiple objects and then quickly adapt to the target object. The experiment results have proved that the proposed method has a more reasonable classification result than state-of-the-art methods. Moreover, in the proposed method, the training time is reduced significantly, and the number of the training epoch is shorter compared with other algorithms.

In real applications, labeled data is required when using an EEG wearable device for the first time. Accordingly, an incorrect annotation may adversely affect the model performance. Meanwhile, in the aspect of algorithm optimization, the weights of different source subjects are the same and there is no concern about more lost in the target data. In future work, it is intended to mainly focus on two contents: The first one is a combination of self-learning and contrastive learning to learn the unique and general features simultaneously and fit the target data, thereby eliminating the calibration data. Moreover, the main focus in the second content is on the allocation of dynamic weights and optimization of feature selection models.

CRediT authorship contribution statement

Yingdong Wang: Conceptualization of this study, Methodology, Writing - original draft. **Jiatong Liu:** Writing - review & editing. **Qunsheng Ruan:** Data curation, Software. **Shuocheng Wang:** Data curation, Software. **Chen Wang:** Resource, Review.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Alarcao, S. M., & Fonseca, M. J. (2019). Emotions recognition using EEG signals: A survey. *IEEE Transactions on Affective Computing*, *10*(3), 374–393.

Alnafjan, A., Hosny, M., Alohali, Y., & Alwabil, A. (2017). Review and classification of emotion recognition based on eeg brain-computer interface system research: A systematic review. *Applied Sciences*, *7*(12), 1239.

Ang, K. K., Chin, Z. Y., Zhang, H., & Guan, C. (2008). Filter bank common spatial pattern (FBCSP) in brain-computer interface. *International Joint Conference on Neural Network*, 2390–2397.

Caselles, V., Kimmel, R., & Sapiro, G. (1995). Geodesic active contours. *International Journal of Computer Vision*, *22*(1), 61–79.

Chen, Z., Zhang, B., Stojanovic, V., Zhang, Y., & Zhang, Z. (2020). Event-based fuzzy control for T-S fuzzy networked systems with various data missing. *Neurocomputing*, *417*, 322–332. http://dx.doi.org/10.1016/j.neucom.2020.08.063, URL: https://www.sciencedirect.com/science/article/pii/S0925231220313448.

Cheng, J., Chen, M., & Li, C. (2020). Emotion recognition from multi-channel EEG via deep forest. *IEEE Journal Biomed Health Inform*.

Cheng, P., He, S., Stojanovic, V., Luan, X., & Liu, F. (2021). Fuzzy fault detection for Markov jump systems with partly accessible hidden information: An event-triggered approach. *IEEE Transactions on Cybernetics*, 1–10. http://dx.doi.org/10.1109/TCYB.2021.3050209.

Cowie, R., Douglascowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. G. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, *18*(1), 32–80.

Duan, T., Chauhan, M., Shaikh, M. A., & Srihari, S. N. (2020). Ultra efficient transfer learning with meta update for cross subject EEG classification. arXiv: Learning.

Duan, R.-N., Zhu, J.-Y., & Lu, B.-L. (2013). Differential entropy feature for EEG-based emotion classification. In *6th international IEEE/EMBS conference on neural engineering (NER)* (pp. 81–84). IEEE.

Fazli, S., Grozea, C., Danoczy, M., Blankertz, B., Popescu, F., & Muller, K. (2009). Subject independent EEG-based BCI decoding. *Neural Information Processing Systems*, 513–521.

Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. CoRR abs/1703.03400. URL: http://arxiv.org/abs/1703.03400, arXiv:1703.03400.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., & Lempitsky, V. (2016). Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, *17*(1), 189–209.

Goodfellow, I., Pougetabadie, J., Mirza, M., Xu, B., Wardefarley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Neural Information Processing Systems*, 2672–2680.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., & Smola, A. J. (2006). A kernel method for the two-sample-problem. In B. Schölkopf, J. C. Platt, & T. Hofmann (Eds.), *Advances in neural information processing systems 19, Proceedings of the twentieth annual conference on neural information processing systems, Vancouver, British Columbia, Canada, December 4-7, 2006* (pp. 513–520). MIT Press, URL: http://papers.nips.cc/paper/3110-a-kernel-method-for-the-two-sample-problem.

Gunes, H., Schuller, B., Pantic, M., & Cowie, R. (2011). Emotion representation, analysis and synthesis in continuous space: A survey. In *Proc. IEEE Int. Conf. Autom. face gesture Recognit. workshops (FG)* (pp. 827–834).

Haeusser, P., Frerix, T., Mordvintsev, A., & Cremers, D. (2017). Associative domain adaptation. In *ICCV* (pp. 2784–2792).

Joshi, J., Goecke, R., Alghowinem, S., Dhall, A., Wagner, M., Epps, J., Parker, G., & Breakspear, M. (2013). Multimodal assistive technologies for depression diagnosis and monitoring. *Journal on Multimodal User Interfaces*, *7*(3), 217–228.

Koch, G., Zemel, R., & Salakhutdinov, R. (2015). Siamese neural networks for one-shot image recognition. In *Neural-networks one-shot-learning* (pp. 255–256).

Koelstra, S., Muhl, C., Soleymani, M., Lee, J., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., & Patras, I. (2012). Deap: A database for emotion analysis ;using physiological signals. *IEEE Transactions on Affective Computing*, *3*(1), 18–31.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86.

Kuusikko, S., Haapsamo, H., Janssonverkasalo, E., Hurtig, T., Mattila, M., Ebeling, H., Jussila, K., Bolte, S., & Moilanen, I. (2009). Emotion recognition in children and adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(6), 938–945.

Lan, Z., Sourina, O., Wang, L., Scherer, R., & Mullerputz, G. (2019). Domain adaptation techniques for EEG-based emotion recognition: A comparative study on two public datasets. *IEEE Transactions on Cognitive and Developmental Systems*, *11*(1), 85–94.

Laureanti, R., Bilucaglia, M., Zito, M., Circi, R., Fici, A., Rivetti, F., Valesi, R., Oldrini, C., Mainardi, L., & Russo, V. (2013). Emotion assessment using machine learning and low-cost wearable devices. In *6th international IEEE/EMBS conference on neural engineering (NER)* (pp. 81–84).

Li, J., Qiu, S., Du, C., Wang, Y., & He, H. (2020). Domain adaptation for EEG emotion recognition based on latent representation similarity. *IEEE Transactions on Cognitive and Developmental Systems*, *12*(2), 344–353.

Li, J., Qiu, S., Shen, Y., Liu, C., & He, H. (2019). Multisource transfer learning for cross-subject EEG emotion recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 1–13.

Li, Y., Zheng, W., Wang, L., Zong, Y., & Cui, Z. (2019). From regional to global brain: A novel hierarchical spatial-temporal neural network model for EEG emotion recognition. *IEEE Transactions on Affective Computing*, 1.

Li, Y., Zheng, W., Wang, L., Zong, Y., & Cui, Z. (2019). From regional to global brain: A novel hierarchical spatial-temporal neural network model for EEG emotion recognition. *IEEE Transactions on Affective Computing*, 1.

Li, Y., Zheng, W., Zong, Y., Cui, Z., Zhang, T., & Zhou, X. (2018). A bi-hemisphere domain adversarial neural network model for EEG emotion recognition. *IEEE Transactions on Affective Computing*, 1.

Lin, Y. (2019). Constructing a personalized cross-day EEG-based emotion-classification model using transfer learning. *IEEE Journal of Biomedical and Health Informatics*, 1.

Ma, B., Li, H., Luo, Y., & Lu, B. (2019). Depersonalized cross-subject vigilance estimation with adversarial domain generalization. *International Joint Conference on Neural Network*, 1–8.

Motiian, S., Jones, Q., Iranmanesh, S. M., & Doretto, G. (2017). Few-shot adversarial domain adaptation. *NIPS*, 6670–6680.

Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv: Learning.

Rodrigues, P. L. C., Jutten, C., & Congedo, M. (2019). Riemannian procrustes analysis: Transfer learning for brain-computer interfaces. *IEEE Transactions on Biomedical Engineering*, *66*(8), 2390–2401.

Rozgic, V., Vitaladevuni, S., & Prasad, R. (2013). Robust EEG emotion classification using segment level decision fusion. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 1286–1290).

Sejdinovic, D., Sriperumbudur, B. K., Gretton, A., & Fukumizu, K. (2013). Equivalence of distance-based and rkhs-based statistics in hypothesis testing. *The Annals of Statistics*, *41*(5), 2263–2291.

Shi, L., & Lu, B. (2010). Off-line and on-line vigilance estimation based on linear dynamical system and manifold learning. *IEEE Engineering in Medicine and Biology Society*, *2010*, 6587–6590.

Song, T., Zheng, W., Liu, S., Zong, Y., & Cui, Z. (2020). Instance-adaptive graph for EEG emotion recognition. *AAAI*.

Song, T., Zheng, W., Song, P., & Cui, Z. (2018). Eeg emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing*, *1*(1), 1.

Spuler, M., Walter, C., Rosenstiel, W., Moller, K., & Klein, E. (2016). Eeg-based prediction of cognitive workload induced by arithmetic: a step towards online adaptation in numerical learning. *Zdm*, *48*(3), 267–278.

Stojanovic, V., He, S., & Zhang, B. (2020). State and parameter joint estimation of linear stochastic systems in presence of faults and non-Gaussian noises. *International Journal of Robust and Nonlinear Control*, *30*(16), 6683–6700. http://dx.doi.org/10.1002/rnc.5131, URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/rnc.5131, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/rnc.5131.

Tan, C., Sun, F., & Zhang, W. (2018). Deep transfer learning for EEG-based brain computer interface. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 916–920).

Tao, H., Li, J., Chen, Y., Stojanovic, V., & Yang, H. (2020). Robust point-to-point iterative learning control with trial-varying initial conditions. *IET Control Theory & Applications*, *14*(19), 3344–3350. http://dx.doi.org/10.1049/iet-cta.2020.0557, URL: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-cta.2020.0557, arXiv:https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-cta.2020.0557.

Tripathi, S., Acharya, S., Sharma, R. D., Mittal, S., & Bhattacharya, S. (2017). Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset. In *AAAI*.

Wei, T., Xiaodi, L., & Stojanovic, V. (2021). Input-to-state stability of impulsive reaction–diffusion neural networks with infinite distributed delays. *Nonlinear Dynamics*, *103*, 1733–1755. http://dx.doi.org/10.1007/s11071-021-06208-6.

Yao, Y., & Doretto, G. (2010). Boosting for transfer learning with multiple sources. *Computer Vision and Pattern Recognition*, 1855–1862.
Yin, Z., Liu, L., Chen, J., Zhao, B., & Wang, Y. (2020). Locally robust EEG feature selection for individual-independent emotion recognition. *Expert Systems with Applications*, *162*, Article 113768. http://dx.doi.org/10.1016/j.eswa.2020.113768, URL: https://www.sciencedirect.com/science/article/pii/S0957417420305923.
Zanini, P., Congedo, M., Jutten, C., Said, S., & Berthoumieu, Y. (2018). Transfer learning: A Riemannian geometry framework with applications to brain-computer interfaces. *IEEE Transactions on Biomedical Engineering*, *65*(5), 1107–1116.
Zhao, S., Wang, G., Zhang, S., Gu, Y., Li, Y., Song, Z., Xu, P., Hu, R., Chai, H., & Keutzer, K. (2020). Multi-source distilling domain adaptation. In *AAAI*.
Zheng, W., & Lu, B. (2016). Personalizing EEG-based affective models with transfer learning. *International Joint Conference on Artificial Intelligence*, 2732–2738.
Zheng, W., Zhu, J., & Lu, B. (2019). Identifying stable patterns over time for emotion recognition from EEG. *IEEE Transactions on Affective Computing*, *10*(3), 417–429.

Yingdong Wang (S'17 - M'17) received the master degree in educational technology from Sun Yat-sen University, in 2012. She is currently pursuing the Ph.D. degree in computer science with Xiamen University. Her current research interests include big data, human–machine interface, and bioinformatics.

Jiatong Liu was born in Qiqihaer, Heilongjiang, China in 1977.received the B.S. degree in Heilongjiang University, Heilongjiang, China, in 2015. She is currently pursuing the M.S. degree in School of Informatics for Xiamen University, Xiamen, China. Her research interests are wide-reaching but mainly involve the areas of pattern recognition and natural language processing, particularly problems involving images and graphs.

Qunsheng Ruan (M'87) born in 1979, male, Associate Professor, is a Ph.D Candidate. His research interests include large scale graph data management and mining, intelligent processing of traditional Chinese medicine information.

Shuocheng Wang received the master degree in computer technology from Guizhou university, in 2019. He is currently pursuing the Ph.D. degree in computer science with Xiamen University. He current research interests computer vision, computer graphics, and machine learning.

Chen Wang was born in Tongliao City, Inner Mongolia, China in 2000. He studied at Xiamen University from 2018 to the present. He is a sophomore at Xiamen University. His current research interests include big data and bioinformatics.