2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Reducing the Calibration Effort of EEG Emotion Recognition using Domain Adaptation with Soft Labels

Zhunan Li[1,2], Hao Chen[1,2], Ming Jin[1,2], Jinpeng Li[1,2,*]

Abstract— **Electroencephalogram (EEG)-based emotion recognition has made great progress in recent years. The current pipelines collect EEG training data in a long-time calibration session for each new subject, which is time consuming and user unfriendly. To reduce the time required for the calibration session, there have been many studies using domain adaptation (DA) approaches to transfer knowledge from existing subjects (source domain) to the new subject (target domain) for reducing the dependence on the calibration session. Existing DA methods usually require substantial unlabeled EEG data of the new subject. However, the real scenario is that there are a small number of labeled samples in the calibration session of the target. Motivated by this, we introduce a novel domain adaptation architecture based on adversarial training to learn domain-invariant feature representations across subjects. To improve the performance when there are few labeled EEG data in the calibration session, we add a soft label loss to the architecture, which can ensure that the inter-class relationships learned from the source domain are transferred to target domain. We evaluate the method on the SEED dataset, and the experimental results show that our method uses only 15 examples per trial in the calibration session to achieve an average accuracy of 87.28%, indicating the effectiveness of our framework.**

I. INTRODUCTION

An increasing number of studies have shown that cognitive and emotional disorders could result in many diseases, including depression, autism and Alzheimer's disease [1]. With this concern, the study of human emotion is of great significance and become one of the research hotspots. The electroencephalogram (EEG) signal has been proven to be a useful tool for emotion recognition in the recent several years due to its outstanding characteristics of high time resolution and fast transmission speed. And then it provides constructive technical supports for establishing real-time emotion recognition systems [2].

The EEG-based BCI is the most popular type of BCIs because of its safety, low cost and convenience [3]. The procedure of EGG-based emotion recognition consists of the following components: (a) Signal acquisition. We use dry electrodes to collect EEG signals from the scalp. (b) Signal processing. This step mainly involves temporal filtering and spatial filtering. The purpose is to reduce noise and improve the signal-to-noise ratio. (c) Feature extraction. There are many ways to extract features of the processed signals, and the EEG-based emotion recognition usually use the time or frequency domain features. (d) Pattern recognition. We select the corresponding method according to the application. When we use deep learning for emotion recognition, both the feature extraction and pattern recognition could be integrated into a single model, and then optimized simultaneously and automatically.

Domain adaptation (DA) is a particularly promising method for emotion recognition tasks. Unlike supervised learning, DA transfers the knowledge from similar or relevant subjects to facilitate learning for a new subject. In domain adaptation, the source and target domains all share the same feature space but have different marginal probability distributions. There have been many studies about using the DA method for EEG-based emotion recognition. For example, Zheng *et al.* [4] first introduced the transfer component analysis (TCA) [5] and transductive parameter transfer (TPT) [6] for EEG-based cross-subject emotion recognition, and achieved the accuracy of 64.00% and 75.17% respectively. As domain adaptation based on deep neural network becomes increasingly popular, Li *et al.* [7] proposed the domain adversarial neural network (DANN) [8] to find the shared representations between source and target domain. Especially, the domain adaptation network (DAN) [9] pushed the accuracy up to 83.81%. Nevertheless, no matter how those methods achieve knowledge transfer, most of them demand the all target information, which is applicable to the offline datasets transfer, but cannot be reached in real-time BCI applications. Li *et al.* [10] noticed that current methods are limited by the number of labeled examples in training data, and proposed the Fast Online Instance Transfer (FOIT) to improve the accuracy of EEG-based emotion recognition. Zhao *et al.* [11] proposed a plug-and-play domain adaptation (PPDA) method for dealing with the inter-subject variability. However, although they reduce the calibration time, it brings the disadvantage of reducing recognition accuracy.

To tackle the problem, inspired by the work [12], this paper introduces a new method that can calibrate with a few labeled target data without sacrificing the recognition accuracy, which is outlined in Fig. 1. We divide the EEG representations of target domain into calibration session and inference session. The calibration session contains the first several trials for the target subject while the inference session consists of the rest of trials. We use the source domain data and a few examples sampled from calibration session to train the model, and the inference session is used to test the performance of our model. In the training phase, the domain confusion loss seeks to learn domain-invariant representations with respectively to the shift between different domains,

[1]HwaMei Hospital, University of Chinese Academy of Sciences, No.41 Northwest Street, Haishu District, Ningbo, Zhejiang, 315010, China.
[2]Ningbo Institute of Life and Health Industry, University of Chinese Academy of Sciences, Ningbo, Zhejiang, China.
*Corresponding author lijinpeng@ucas.ac.cn

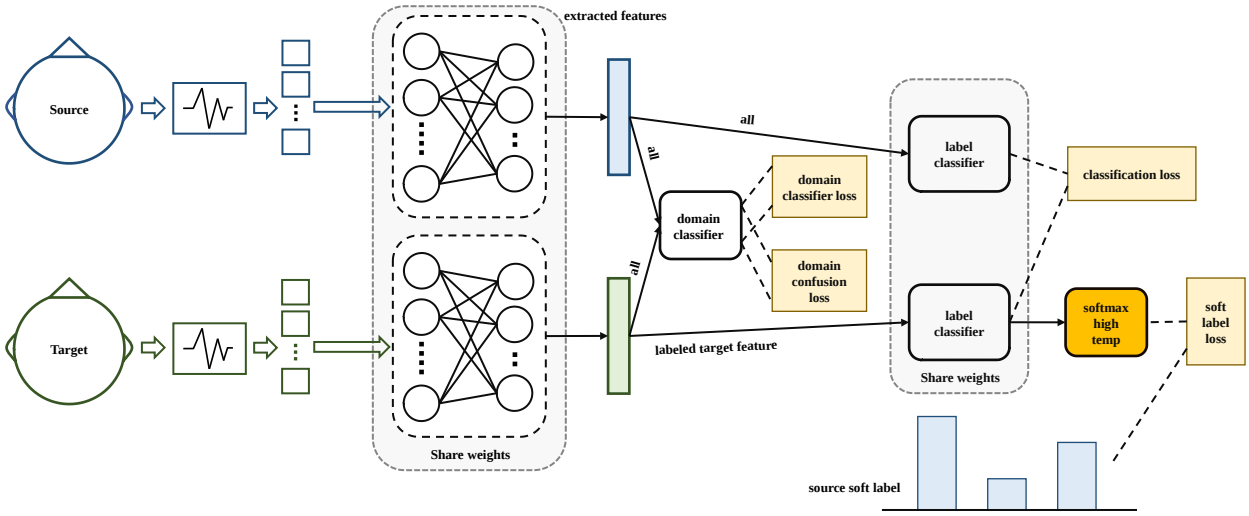

Fig. 1. The architecture of our proposed architecture. The shared encoder is a multi-layer perceptron (MLP). The domain confusion loss and domain classifier loss are used to learn a domain invariant representation over the source data and target data (both labeled and unlabeled). At the same time, The soft label loss is applied to transfer inter-class correlations knowledge from source domain to target domain.

while the domain classifier loss is used to contest with the domain confusion loss, and distinguish which domain the example comes from. Most of current domain adaptation methods only consider the marginal distribution information, and ignore the label distribution information between source and target domain. Since the label distribution holds key information about the relationships between categories, we add a soft label loss to our objective so as to transfer inter-class correlations knowledge from source domain to target domain.

II. METHOD

The architecture of our network is depicted in Fig. 1. The network takes as input the source data $\{x_s, y_s\}_1^n$ (blue one in Fig. 1) and the target data $\{x_t, y_t\}_1^m$ (green one in Fig. 1), where y_t are only provided for a small number of target data. The goal is to seek a label classifier θ_{cls} that could correctly classify target examples by operating on an EEG feature representation $f(x; \theta_{mlp})$.

Usually, the loss function of classification task is the cross entropy between the output predicted labels and ground-truth labels of examples, which can be defined as follows:

$$\mathcal{L}_{cls}(x_s, y_s, x_t, y_t; \theta_{mlp}, \theta_{cls}) = -\sum_k \mathbf{1}[y = k] \log p_k, \quad (1)$$

where k is the ground truth label of corresponding example, and p is the softmax of the label classifier activation: $p = softmax(\theta_{cls} \cdot f(x; \theta_{mlp}))$. The available source labeled data are used to train the shared encoder and label classifier parameters according to (1), yet it often result in overfitting to the source data in domain adaptation. However, if the source domain and target domain are similar enough that the classifier trained on the source will perform well on the target domain. Actually, under the learned representation θ_{mlp}, it is possible for both source and target domain data to be very similar. To address the problem, we add an additional domain classifier with parameters θ_{dom}, and directly train the domain classifier to identify whether a training example comes from source domain or target domain. Furthermore, in order to minimize the difference between the source and target data distribution, we also add a new loss $\mathcal{L}_{conf}$, called *domain confusion loss*, to the subjective by optimizing the representation θ_{mlp}.

Then for a particularly feature representation θ_{mlp}, we seek to learn the best domain classifier on the representation by optimizing the following objective:

$$\mathcal{L}_{dom}(x_s, x_t, \theta_{mlp}; \theta_{dom}) = -\sum_d \mathbf{1}[y_d = d] \log q_d, \quad (2)$$

where y_d denotes the domain that the example is originated from, and q corresponding to the softmax of the domain classifier activation: $q = softmax(\theta_{dom} \cdot f(x; \theta_{mlp}))$. For a particularly domain classifier θ_{dom}, the domain confusion loss is introduced to learn domain invariant by finding a representation in which the best domain classifier performs poorly. This can be formulated as:

$$\mathcal{L}_{conf}(x_s, x_t, \theta_{dom}; \theta_{mlp}) = -\sum_d \frac{1}{|D|} \log q_d, \quad (3)$$

where $|D|$ represents the number of domain. Since both two loss functions are completely opposite to each other, we use iterative update strategy to optimize the two objectives by fixing parameters from the previous iteration.

Though training the network to confuse the domain classifier acts to align the marginal distribution, there are no gurantees about the alignment of corresponding classes between source and target domain. To deal with the issue, we use a new label, called soft label, which is the average over a softmax with a high temperature τ of all activation of source examples, rather than the EEG data category hard label. We

denote the average as $l^{(k)}$ corresponding to the category k. Now, we can define the *soft label loss* as:

$$\mathcal{L}_{soft}(x_t, y_t; \theta_{mlp}, \theta_{cls}) = -\sum l^{(y_t)} \odot \log p, \quad (4)$$

where the symbol $\odot$ represents element-wise product, and p denotes the soft activation of the target labeled data: $p = softmax(\theta_{cls} \cdot f(x_T; \theta_{mlp})/\tau)$. By optimizing the loss function to match the expected source output distributions on the target data, we can transfer the learned inter-class correlations from the source domain to examples in the target domain.

To sum up, we firstly minimize the domain classifier loss to update the parameter θ_{dom} only:

$$\min_{\theta_{dom}} \mathcal{L}_{dom}(x_s, x_t, \theta_{mlp}; \theta_{dom}), \quad (5)$$

and then optimize the joint loss function to get the better feature representation and label classifier:

$$\min_{\theta_{cls}, \theta_{mlp}} \mathcal{L}_{cls}(x, y; \theta_{mlp}, \theta_{cls}) + \alpha \mathcal{L}_{conf}(x_s, x_t, \theta_{dom}; \theta_{mlp})$$
$$+\beta \mathcal{L}_{soft}(x_t, y_t; \theta_{mlp}, \theta_{cls}), \quad (6)$$

where α and β are the hyperparameters, which determine how strongly domain confusion and soft label influence the optimization. We can get the optimal representation parameters θ^*_{mlp} and label classifier parameters θ^*_{cls} by repeating the above procedure.

III. EXPERIMENTS

To analyze the effectiveness of our method, we use leave-one-subejct-out cross valudation to evaluate the approach on the SJTU Emotion EEG Dataset (SEED) [13], a dataset collection for various purposes using EEG signals. In the dataset, there are 15 Chinese movie clips to be used to elicit the desired target emotion among positive, negative and neutral. Fifteen subjects (7 males and 8 females) participated in the experiment three times on the different days. The experimental procedures involving human subjects described in this paper were approved by the Institutional Review Board.

A. Implementation Details

For each file (.mat) in this dataset, there are 15 trials and each trial consists of PSD, DE, DASM, RASM and DCAU features. The extracted differential entropy (DE) features of EEG signals are used to train the network. For each target subject, we divide the EEG representations into calibration session and inference session. We evaluate the performance under different four partitions of target domain: (1) the calibration consists of first 3 trials, (2) the calibration consists of first 6 trials, (3) the calibration consists of first 9 trials, (4) the calibration consists of first 12 trials. We follow the standard protocol for this dataset and sample 15 examples per trial in the calibration session. We use those examples and the rest of fourteen subjects as training data, and the inference session as testing data.

For our implementation of the model, we use a four-layer multi-layer perceptron of 512, 128, 128 and 64 hidden nodes respectively, with batch normalization and rectified linear units (ReLU) between each layer, and a domain classifier and label classifier following the output of multi-layer perceptron. The both two classifiers are linear layers whose nodes are 2 and 3 respectively. We also add L2 regularization and Dropout [14] to the model for avoiding the overfitting. The whole model parameters are updated using SGD with a learning rate of 0.001. The networks are trained with a mini-batch of 15 examples at every epoch, and tested with the unlabeled target domain data every 5 epochs. The hyper-parameters α and β are set to 0.01, 0.1 respectively. To allow for reproducible comparison, our results are reported over a random seeds. The main results are presented in Table I, and the results of ablation experiment are shown in Table II. In order to investigate the effect of the number labeled examples per target trial on our method, we also depict the accuracy with respect to the increase of labeled data amount in Fig. 2.

TABLE I
EXPERIMENTAL RESULTS OF DIFFERENT METHODS RUNNING ON THE SEED DATASET.

Method	Avg.	Std.
TCA[4]	64.00	14.66
TPT[4]	75.17	12.83
DANN[7]	79.19	13.14
DAN[7]	83.81	8.56
WGANDA[11]	87.10	7.10
PPDA[11]	86.70	7.10
Ours	87.28	5.75

TABLE II
EXPERIMENTAL RESULTS ON THE SEED DATASET. MAXIMUM VALUE FOR EACH TASK IS BOLDED.

Method	FOIT[10]	Ours[5]	Ours[6]	Ours[7]
3L-12U[1]	69.55±18.04	85.21±9.19	84.67±5.60	**87.28±5.75**
6L-9U[2]	80.61±10.80	88.73±7.09	89.42±6.91	**91.02±6.19**
9L-6U[3]	81.75±11.71	93.81±5.37	**94.24±5.10**	92.07±5.74
12L-3U[4]	86.54±11.56	91.46±9.64	90.43±10.57	**91.77±9.01**

[1] 3 labeled trials and 12 unlabeled trials
[2] 6 labeled trials and 9 unlabeled trials
[3] 9 labeled trials and 6 unlabeled trials
[4] 12 labeled trials and 3 unlabeled trials
[5] domain confusion loss only
[6] soft label loss only
[7] both domain confusion loss and soft label loss

B. Results and Discussion

To demonstrate the advantages of our method, we compare the performance with that of several popular methods on the cross-subject emotion recognition task on SEED, and report the average classification accuracy in Table I. As is shown, the proposed approach outperforms all other algorithms. From the experimental results, we have the following observations. Compared with WGANDA and PPDA respectively, our method could achieve an better accuracy of recognition, which demonstrates that our proposed architecture is capable of learning domain-invariant features. For further

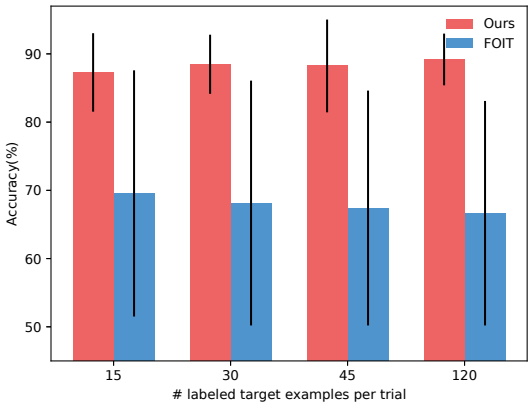

Fig. 2. Performance of our model with varying numbers of labeled target examples per trial.

verifying the effectiveness of soft label loss, we compare the performance of two variants of the proposed method: (1) using domain confusion loss only. (2) using soft label loss only. The average results on SEED are reported in Table II. From the results, our method (both domain confusion loss and soft label loss) achieves the highest classification accuracy among all variations. It indicates that it is not enough to only learn domain invariance, and simultaneously considering the relationships between classes in both source and target domain can achieve better performance.

The purpose of using calibration session is to enhance the model performance by directly transferring inter-class information from source to target. However, it is hard to know how many labeled target examples per trial is optimal. In order to find the appropriate number of labeled examples per trial in target domain, we depict the accuracy change of all methods with respect to the increase of labeled target examples in Fig. 2. As presented in Fig. 2, the performance of our model will increase with the extension of labeled target examples. However, we do not see a significant improvement of accuracy with the number of labeled target examples becomes larger. While the FOIT is the opposite of our method, its performance will decrease as the number of labeled target examples becomes larger. We compare our method with the FOIT, we can find that the most benefit of our method arises when there are a few labeled training examples per trial in the calibration session. In other words, our model can get the best with fifteen examples per trial in the calibration session when considering sparsely labeled target domain data.

IV. CONCLUSIONS

In this paper, we propose a novel domain adaptation architecture based on adversarial learning to extract domain-invariant features for cross-subject EEG-based emotion recognition. Different from other EEG classification methods, our approach exploits inter-class correlations from the source domain because it has key information about the relationships between categories. Moreover, since our method is suitable for solving the problem of a small amount of labeled EEG data in the target domain, in order to explore the appropriate number of labeled example for our method, we investigate the performance of the method with varying numbers of labeled target examples, and find that 15 labeled samples are the most suitable for our method. Finally, the experimental results show that the method can manage to decrease the number of labeled examples in target domain with the best accuracy about 87.28%, a comparable result to the state-of-the-art emotion recognition performance.

ACKNOWLEDGMENT

This work was supported in part by Zhejiang Provincial Natural Science Foundation of China (LQ20F030013), Research Foundation of HwaMei Hospital, University of Chinese Academy of Sciences, China (2020HMZD22), Ningbo Public Service Technology Foundation, China (202002N3181), and Medical Scientific Research Foundation of Zhejiang Province, China (2021431314).

REFERENCES

[1] R. S. Bucks and S. A. Radford, "Emotion processing in alzheimer's disease," *Aging Ment Health*, vol. 8, no. 3, pp. 222–232, 2004.

[2] D. Wu, Y. Xu, and B.-L. Lu, "Transfer learning for eeg-based brain-computer interfaces: A review of progress made since 2016," *IEEE Transactions on Cognitive and Developmental Systems*, pp. 1–1, 2020.

[3] W. Hu, G. Huang, L. Li, L. Zhang, Z. Zhang, and Z. Liang, "Video-triggered eeg-emotion public databases and current methods: A survey," *Brain Science Advances*, vol. 6, no. 3, pp. 255–287, 2020.

[4] W.-L. Zheng and B.-L. Lu, "Personalizing eeg-based affective models with transfer learning," in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 2016, pp. 2732–2738.

[5] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2010.

[6] E. Sangineto, G. Zen, E. Ricci, and N. Sebe, "We are not all equal: Personalizing models for facial expression analysis with transductive parameter transfer," in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 357–366.

[7] H. Li, Y.-M. Jin, W.-L. Zheng, and B.-L. Lu, "Cross-subject emotion recognition using deep adaptation networks," in *International Conference on Neural Information Processing*. Springer, 2018, pp. 403–413.

[8] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2096–2030, 2016.

[9] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *Proceedings of International Conference on Machine Learning*. PMLR, 2015, pp. 97–105.

[10] J. Li, H. Chen, and T. Cai, "Foit: Fast online instance transfer for improved eeg emotion recognition," in *2020 IEEE International Conference on Bioinformatics and Biomedicine*. IEEE, 2020, pp. 2618–2625.

[11] L.-M. Zhao, X. Yan, and B.-L. Lu, "Plug-and-play domain adaptation for cross-subject eeg-based emotion recognition," 2021.

[12] E. Tzeng, J. Hoffman, T. Darrell, and K. Saenko, "Simultaneous deep transfer across domains and tasks," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 4068–4076.

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

[14] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv:1207.0580*, 2012.