



Cross-Subject Emotion Recognition Using Deep Adaptation Networks

He Li¹, Yi-Ming Jin¹, Wei-Long Zheng¹, and Bao-Liang Lu^{1,2,3}(✉)

¹ Center for Brain-Like Computing and Machine Intelligence,
Department of Computer Science and Engineering, Shanghai Jiao Tong University,
800 Dong Chuan Road, Shanghai 200240, China
{polarsky, jinyiming, weilong, bllu}@sjtu.com

² Key Laboratory of Shanghai Education Commission for Intelligent
Interaction and Cognitive Engineering, Shanghai Jiao Tong University,
800 Dong Chuan Road, Shanghai 200240, China

³ Brain Science and Technology Research Center, Shanghai Jiao Tong University,
800 Dong Chuan Road, Shanghai 200240, China

Abstract. Affective models based on EEG signals have been proposed in recent years. However, most of these models require subject-specific training and generalize worse when they are applied to new subjects. This is mainly caused by the individual differences across subjects. While, on the other hand, it is time-consuming and high cost to collect subject-specific training data for every new user. How to eliminate the individual differences in EEG signals for implementation of affective models is one of the challenges. In this paper, we apply Deep adaptation network (DAN) to solve this problem. The performance is evaluated on two publicly available EEG emotion recognition datasets, SEED and SEED-IV, in comparison with two baseline methods without domain adaptation and several other domain adaptation methods. The experimental results indicate that the performance of DAN is significantly superior to the existing methods.

Keywords: Affective brain-computer interface
Emotion recognition · EEG · Deep neural network · Domain adaptation

1 Introduction

Emotion plays a critical role in human lives, which affects our behavior and thought almost anytime and anywhere. As a result, the technology of emotion recognition has various applications in many fields, including assistance for people everyday life, improvement of working performance, and even implementation of emotional intelligence. On the other hand, EEG signals are considered to reflect the internal temporal states of human brains and has been studied in the field of Brain-computer interface (BCI). In recent years, BCIs have also seen

H. Li and Y.-M. Jin—The first two authors contributed equally to this work.

progressive growth in affective Brain-computer interface (aBCI) that aims at recognizing emotions from brain signals [13, 18]. Many studies have been made in detecting human emotion states with EEG signals [4, 8, 12].

Though existing studies have achieved many successes in emotion recognition, most of them only focus on training specific models for particular subjects. These subject-specific models suffer from degraded performance when they are applied to new subjects. The phenomenon is caused by the large domain shift introduced by individual differences across subjects. The naive solution for the problem is to train subject-specific models for every subject, which takes a lot of effort to collect labeled dataset. Another path to solve the problem is to apply domain adaptation methods. As the matter of fact, domain adaptation methods have been applied in various fields to solve the domain shift problem.

There are already several studies on the application of domain adaptation methods to EEG-based emotion recognition. In our previous work [21], Zheng and Lu adopted Transfer component analysis (TCA) [14], Kernel principle analysis (KPCA) [17], and Transductive parameter transfer (TPT) [16] for emotion recognition. Lan *et al.* explored various domain adaptation methods applied on two EEG emotion recognition datasets [9]. Jin *et al.* proposed to use Domain-adversarial neural network (DANN) [6] to eliminate the subject differences and achieved appreciable improvement in recognition performance [7]. Lin *et al.* proposed a conditional transfer learning method for emotion recognition task to avoid negative transfer [10].

In this paper, we introduce Deep adaptation network (DAN) to EEG-based emotion recognition and compare DAN with two baseline methods without domain adaptation and several other domain adaptation methods. As far as we know, this is the first work to apply DAN to deal with the subject transfer problem in EEG-based emotion recognition on two publicly available datasets: SEED and SEED-IV. From experimental results we find that DAN achieves the best performance and improves the accuracy of recognition significantly in comparison with the baseline methods.

2 Materials and Methods

2.1 Dataset Description

Two publicly available emotion recognition datasets, SEED [21] and SEED-IV¹ [20], are used in this paper to evaluate the proposed methods.

The SEED dataset contains EEG signals of 15 subjects recorded while they were watching Chinese film clips. A total number of 15 film clips were selected by a preliminary study and labeled as being negative, positive, or neutral. For each of the subjects, three experiments were performed at an interval of no less than one week. During the experiments, the 15 film clips were played in 15 trials and the subjects were required to watch the clips patiently. The EEG signals

¹ The SEED and SEED-IV datasets are available at <http://bcmi.sjtu.edu.cn/~seed/index.html>.

were recorded with a 62-channel cap according to the 10–20 system using ESI Neuroscan system.

The SEED-IV dataset also contains EEG signals of 15 subjects while they were watching Chinese film clips. The main difference between SEED-IV and SEED is there are film clips of four emotion categories: happy, sad, fear, and neutral. A total number 72 film clips were selected by a preliminary study (18 clips for each emotion category). Each experiment contains 24 trials so that the subjects watched all of the 72 film clips. The EEG signals were recorded with 62-channel cap according to the 10–20 system using ESI Neuroscan system.

2.2 Data Preprocessing

The EEG signals are firstly downsampled to 200 Hz and processed with a 1–75 Hz bandpass filter. The filtered signals are then segmented into 1-s and 4-s segments for SEED and SEED-IV datasets, respectively. The segments are attached with the label of the corresponding film clips. Differential entropy (DE) features are extracted from the segment at the frequency band of delta (1–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), beta (14–31 Hz), and gamma (31–50 Hz) [5, 19]. The DE feature is a robust EEG feature that has been applied in our previous studies [7, 21]. The definition of the DE feature on a one-dimensional signal X drawn from a Gaussian distribution $N(\mu, \delta^2)$ is

$$h(X) = - \int_{-\infty}^{\infty} P(x) \log(P(x)) = \frac{1}{2} \log 2\pi e \delta^2. \quad (1)$$

For SEED, because there are three duplicate experiment for each subject, we select one of them to reduce the scale of the data. After the preprocessing, there are 3394 and 2505 data samples for each subject in the SEED and SEED-IV datasets, respectively. The feature dimension is 310 (62 EEG channels by 5 frequency bands).

2.3 Domain Adaptation Methods

According to Pan *et al.* [15], a domain $\mathcal{D} = \{\mathcal{X}, P(X)\}$ consists of a feature space \mathcal{X} and the corresponding marginal probability distribution $P(X)$, where $X \in \mathcal{X}$. Given a domain \mathcal{D} , a task $\mathcal{T} = \{\mathcal{Y}, f(\cdot)\}$ consists of a label space \mathcal{Y} and the corresponding objective predictive function $f(\cdot)$, where $y = f(x), x \in \mathcal{X}$, and $y \in \mathcal{Y}$.

Traditional machine learning approaches focus on solving the task with data samples and the corresponding labels from the same domain. However, in the field of transfer learning, the goal is to solve tasks in a domain when there is no or little observation of data sample while some data samples are available from related domains.

The objective domain where the task lies is called the target domain $\mathcal{D}_T = \{\mathcal{X}_T, P(X_T)\}$ and the related domain is called source domains $\mathcal{D}_S =$

$\{\mathcal{X}_S, P(X_S)\}$. Additionally, when source and target domains share the same feature space and task, the problem is a subset of transfer learning, and is called domain adaptation. In this paper, we study the domain adaptation problem with the target domains being the data from the target subjects, the source domains being the data from the other subjects.

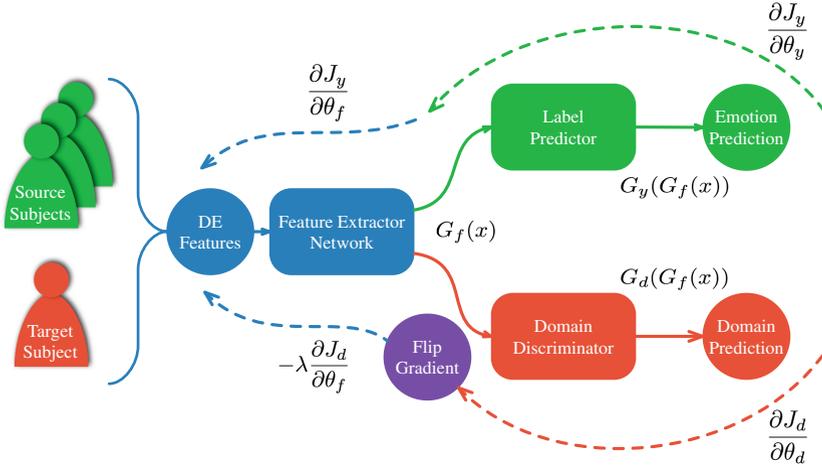


Fig. 1. Structure of DANN. The arrows in solid lines indicate the forward propagation path, while the arrows in dotted lines indicate the backpropagation path.

Domain-Adversarial Neural Network. Domain-adversarial neural network (DANN) is a domain adaptation method based on deep adversarial network. It is composed of three sub-networks as shown in Fig. 1: a feature extractor, a label predictor, and a domain discriminator whose network functions are denoted by $G_f(\cdot)$, $G_y(\cdot)$, and $G_d(\cdot)$ parameterized by θ_f , θ_y , and θ_d , respectively. The method aims to train a feature extractor that eliminates domain discrepancies as well as keep objective task related component of the input features.

In the forward propagation phase, the feature extractor projects the input features into a new feature space. The output is directed to the label predictor and the domain discriminator, simultaneously. The label predictor produces predictions of the labels according to the input, while the domain discriminator produces predictions of the corresponding domain. The loss of the whole network is

$$\frac{1}{n} \sum_{i=1}^n J_y(G_y(G_f(x_i)), y_i) + \alpha J_d(G_d(G_f(x_i)), d_i), \quad (2)$$

where $J_y(G_y(G_f(x_i)), y_i)$ denotes the loss for the label prediction $G_y(G_f(x_i))$ when the true label is y_i , $J_d(G_d(G_f(x_i)), d_i)$ denotes the loss for the domain prediction $G_d(G_f(x_i))$ when the true domain is d_i , α is a tradeoff hyperparameter, and n is the data sample number.

During the backpropagation phase, the label prediction and domain discrimination losses (J_y and J_d) are propagated along the network as in ordinary networks. However, the derivatives are inverted when it is passed from the domain discriminator to the feature extractor: the feature extractor is updated in the direction of maximizing the domain discrimination losses (i.e., deceiving the domain discriminator). In this way, the feature extractor finally discards the domain-specific component of the input (i.e., eliminates the domain discrepancies) in order to keep the domain discrimination losses J_d high. In the test phase, the prediction is made by the feature extractor and the label predictor.

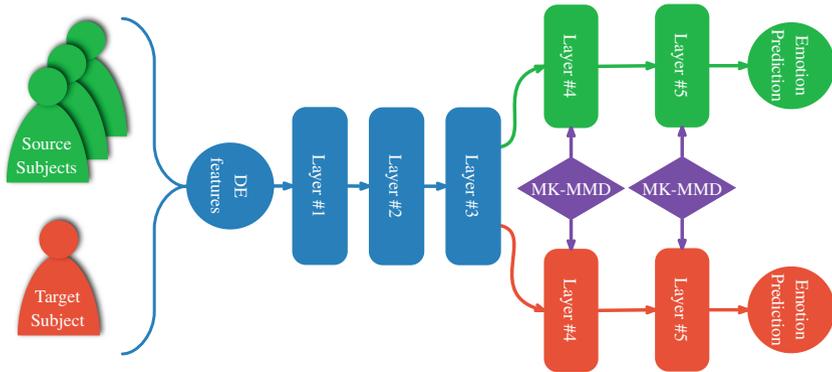


Fig. 2. Structure of DAN. The first 3 layers are ordinary layers. The MK-MMD values between source and target domains are calculated in the forth and fifth layers.

Domain Adaptation Network. According to the recent findings, neural networks extract features that transition from general to domain and task specific with the growth of their depths. Basing on this idea, Long *et al.* proposed to use multi-kernel Maximum mean discrepancies (MK-MMDs) as a measurement for domain discrepancies of hidden features extracted by deep layers in neural networks [11]. By jointly minimizing the MK-MMDs and the task related loss, the proposed Deep adaptation network (DAN) can eliminate domain discrepancies across domains as well as maintaining task related features.

The structure of DAN is shown in Fig. 2. The first several layers are ordinary ones that behave the same as in traditional networks in forward-propagation and back-propagation phases. Because the feature representation transition to be task and domain specific as the layers become deeper, the deep layers must be treated differently to eliminate domain discrepancies. MK-MMD is applied to achieve this goal in DAN. MK-MMD is multiple kernel variation of MMD that is used for distribution discrepancy measurement. The MK-MMD distance between two probability distributions p and q is defined as the distance between their mean embeddings in a reproducing kernel Hilbert space (RKHS) endowed with a characteristic kernel k :

$$d_k^2(p, q) \triangleq \|E_p[\phi_k(x)] - E_q[\phi_k(x)]\|_{\mathcal{H}_k}^2, \tag{3}$$

where $\phi_k(\cdot)$ is the corresponding projection function associated with the kernel. If the probability distributions p and q are the ones of the source and target domains, respectively, the MK-MMD value can then measure the domain discrepancy. In order to eliminate the domain discrepancies in the deep layers, the MK-MMDs between the distributions of source and target domain feature expressions in the deep layers should be minimized. As a result, the final objective for DAN is

$$\min_{\Theta} \frac{1}{n} \sum_{i=1}^n J(\theta(x_i), y_i) + \lambda \sum_{l=4}^5 d_k^2(\mathcal{D}_S^l, \mathcal{D}_T^l), \quad (4)$$

where $J(\theta(x_i), y_i)$ indicates the loss when the network predicts $\theta(x_i)$ for a data sample x_i with the true label y_i , $d_k^2(\mathcal{D}_S^l, \mathcal{D}_T^l)$ indicates the MK-MMDs between the distributions of source and target domain feature expressions in the l th layer, Θ is the set of all of the parameters, n is the size of the training set, and λ is a tradeoff hyperparameter that balance the objective loss and the MK-MMD loss. However, as (3) indicates, the calculation of MK-MMDs between two domains requires computation complexity of $O(n^2)$, which is not feasible during training of neural network. Here we propose to use an unbiased estimate of MK-MMD within a batch which can be computed with $O(n)$ cost.

3 Evaluation Experiments

3.1 Experiment Settings

We applied leave-one-subject-out cross validation to evaluate the domain adaptation methods on SEED and SEED-IV datasets: for each subject, an emotion recognition model is trained with the subject as target domain, and other subjects as source domain. The deep learning based methods are compared with several traditional methods and two baseline methods to show their advantages. Both DAN and DANN contain convolutional layers to extract features from images in there original papers [6, 11]. In this paper, general features are used instead of images, so the network structures are modified to adapt our problem. For DANN, the feature extractor consists of two fully connected layers, and the label predictor and the domain discriminator consist of three fully connected layers. For DAN, there are three ordinary fully connected layers, two specialized fully connected layers attached with MK-MMD losses, and one output layer for the label prediction. The specific structure of the two networks are described in Table 1. Other methods are described as follows:

- (1) KPCA projects the original features into a reproduce Hilbert kernel space (RHKS) with a projection function $\phi(\cdot)$ [17]. A low dimensional subspace of the RHKS is then found by maintaining the variance of the data distribution.
- (2) TCA is similar to KPCA in projecting the original features into a RHKS and find a low dimensional subspace [14]. The difference lies in that the subspace is found by minimizing the MMD distance between the source and target domain distributions as well as preserving data properties that are useful for the target supervised learning task.

- (3) TPT is a parameter based domain adaptation method on multiple source domains [16]. The method consists of three steps. First, domain-specific models are learnt on each domain. Then, a regression algorithm is applied to project the source domain distributions to the domain-specific model parameters. Finally, the domain-specific model for the target domain is constructed with the target domain distribution and the regression algorithm.
- (4) Baseline methods consist of training Support vector machine (SVM) and Multi-layer perceptron (MLP) models on the source domain and applying the trained models directly on the target domain.

Table 1. Structure description of DANN and DAN

Method	Description
DANN	The feature extractor has 2 layers, both with node number of 128. The label predictor and domain discriminator have 3 layers with node numbers of 64, 64, and C , respectively. C indicates the number of emotion classes to be recognized
DAN	There are 5 layers in total, each of them with node numbers of 128, 128, 64, 64, and C from the input end to the output end, respectively. The last two layers are attached with MK-MMD losses

3.2 Results and Discussion

The mean accuracies and standard deviations of each method for the two datasets are shown in Tables 2 and 3, respectively. The specific statistics when each subject is trained as target domain are shown in Figs. 3 and 4.

Table 2. Means and standard deviations of the accuracies for each method applied to the SEED dataset

Method	SVM	MLP	TCA	KPCA	TPT	DANN	DAN
Mean	0.5818	0.6101	0.6400	0.6902	0.7517	0.7919	0.8381
SD	0.1385	0.1238	0.1466	0.0925	0.1283	0.1314	0.0856

For SEED, we compare our previous results in [7] with the results of DAN. As Table 2 shows, DAN achieves the mean accuracy of 0.8381, which outperforms any other methods. DAN also achieves the smallest standard deviation value of 0.0856. It outperforms DANN, which was the best method in [7], by 4.62% in terms of mean recognition accuracy (but with no statistical significance with $p = 0.2645$ in ANOVA test). To show the advantages of DAN, we further compare it with results on the SEED dataset from other papers. Chai

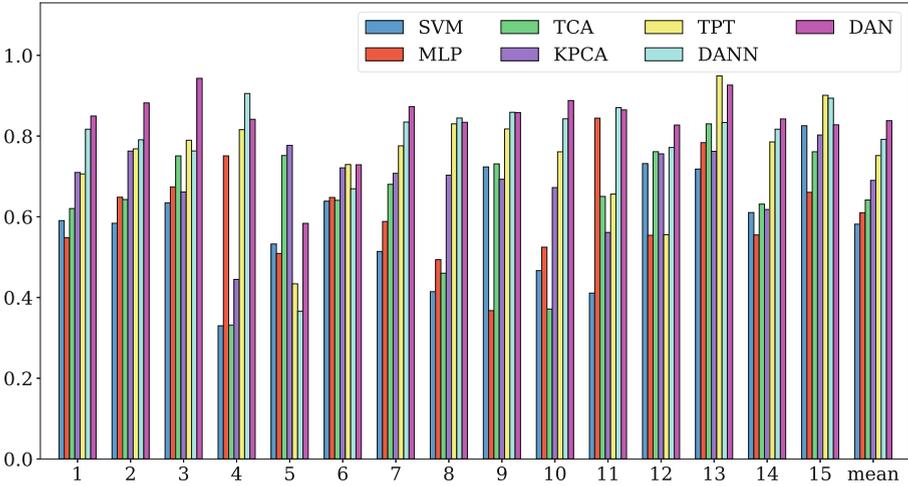


Fig. 3. The specific accuracies of each method for all the subjects and the averages in the SEED dataset.

and colleagues applied several novel domain adaptaiton methods to cross-subject emotion recognition from EEG data and evaluated those methods on the SEED dataset. They reported that the mean accuracies of 77.88%, 80.46%, and 79.61% were obtained in their studies [1–3]. Though the evaluation strategies are slightly different (mostly on the selection of the data), DAN outperforms all of the existing methods, which confirms it to be the state-of-the-art approach on the dataset for the cross-subject problem.

Table 3. Means and standard deviations of the accuracies for each method applied to the SEED-IV dataset

Method	SVM	MLP	TCA	KPCA	TPT	DANN	DAN
Mean	0.5178	0.4935	0.5397	0.5176	0.5243	0.5463	0.5887
SD	0.1285	0.0974	0.0805	0.1289	0.1443	0.0803	0.0813

As for SEED-IV, DAN still achieves the best performance, followed with DANN, TCA, TPT, SVM, KPCA, and MLP (in order of declining performance). The method outperforms the baseline SVM and DANN with 6.09% and 4.24% in terms of mean accuracy, respectively. The other deep learning based method, DANN, achieves the second best mean accuracy and the smallest standard deviation. TCA achieves the best performance among the three traditional methods, but still falls behind DAN with 4.90% of the mean accuracy. In the original paper of SEED-IV [20], Zheng and colleagues achieved a mean accuracy of 70.58% in a within-subject and within-experiment evaluation experiment (training and test

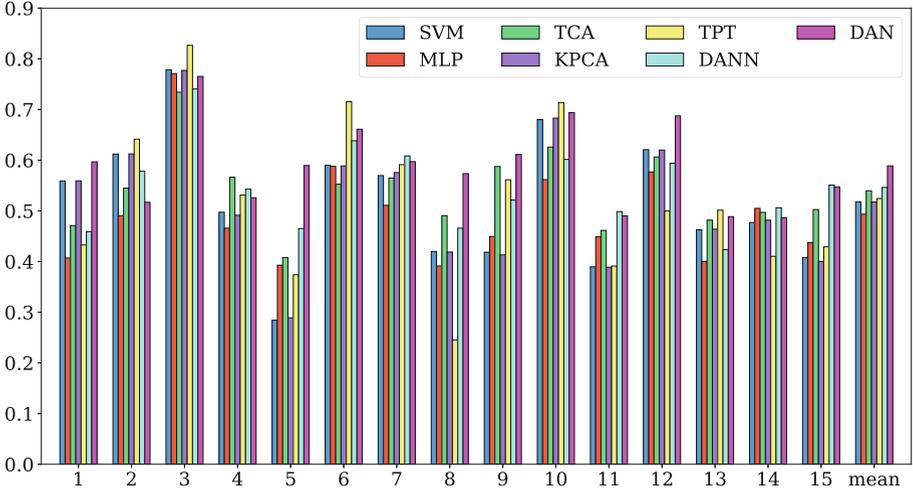


Fig. 4. The specific accuracies of each method for all the subjects and the averages in the SEED-IV dataset.

data are from the same subject, in the same experiment), compared with 58.87% in our results. There is an 11.72% gap of accuracy between the two mean accuracies. However, considering the great difference in the evaluation settings, our results should be a desirable one.

For both of the datasets, DAN outperforms DANN and achieves the best performance in terms of mean accuracy. It outperforms the baseline method significantly for SEED (with $p < 0.01$ in ANOVA test). For SEED-IV, it outperforms the baseline method with weaker statistical significance (with $p < 0.1$ in ANOVA test). Besides, it also has the smallest and the second smallest standard deviation of the recognition accuracies for the two dataset, respectively. These clues demonstrate that DAN is suitable for the EEG-based cross-subject emotion recognition, and can achieve more stable performance in comparison with the other domain adaptation methods.

By observing the accuracies on the two datasets, we find that the overall performance of the methods is worse in SEED-IV compared with those in SEED. There might be two reasons for this phenomenon. The first one is that SEED-IV contains four emotional states for recognition, which makes its task a harder one. The second one lies in that each subject has 2505 data samples in SEED-IV, compared with 3394 data samples in SEED, which adds to the difficulty for the methods to capture and eliminate the domain discrepancies.

4 Conclusion

In this paper, we have adopted Deep adaptation network (DAN) for dealing with the cross-subject problem in EEG-based emotion recognition. Two publicly available datasets SEED and SEED-IV have been used for performance evaluation.

The proposed method, DAN, was compared with several other domain adaptation approaches. The experimental results demonstrate that DAN achieves 4.62% and 4.24% accuracy improvements on three and four classes emotion recognition problems, respectively, and is suitable for the cross-subject emotion recognition from EEG data.

Acknowledgement. This work was supported in part by the grants from the National Key Research and Development Program of China (Grant No. 2017YF-B1002501), the National Natural Science Foundation of China (Grant No. 6167-3266), and the Fundamental Research Funds for the Central Universities.

References

1. Chai, X., et al.: A fast, efficient domain adaptation technique for cross-domain electroencephalography (EEG)-based emotion recognition. *Sensors* **17**(5), 1014 (2017)
2. Chai, X., Wang, Q., Zhao, Y., Liu, X., Bai, O., Li, Y.: Unsupervised domain adaptation techniques based on auto-encoder for non-stationary EEG-based emotion recognition. *Comput. Biol. Med.* **79**, 205–214 (2016)
3. Chai, X., Wang, Q., Zhao, Y., Liu, X., Liu, D., Bai, O.: Multi-subject subspace alignment for non-stationary EEG-based emotion recognition. *Technol. Health Care* **26**, 1–9 (2018)
4. Daniela, S., Maren, G., Thomas, F., Stefan, K.: Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music. *Psychophysiology* **44**(2), 293–304 (2007)
5. Duan, R., Zhu, J., Lu, B.: Differential entropy feature for EEG-based emotion classification. In: *International IEEE/EMBS Conference on Neural Engineering*, pp. 81–84. IEEE Press, San Diego (2013)
6. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: *International Conference on Machine Learning*, vol. 37, pp. 1180–1189. PMLR, Lille (2015)
7. Jin, Y.M., Luo, Y.D., Zheng, W.L., Lu, B.L.: EEG-based emotion recognition using domain adaptation network. In: *International Conference on Orange Technologies*, Singapore, pp. 222–225 (2017)
8. Knyazev, G.G., Slobodskoj-Plusnin, J.Y., Bocharov, A.V.: Gender differences in implicit and explicit processing of emotional facial expressions as revealed by event-related theta synchronization. *Emotion* **10**(5), 678–687 (2010)
9. Lan, Z., Sourina, O., Wang, L., Scherer, R., Müller-Putz, G.R.: Domain adaptation techniques for EEG-based emotion recognition: a comparative study on two public datasets. *IEEE Trans. Cogn. Dev. Syst.* **1** (2018)
10. Lin, Y.P., Jung, T.P.: Improving EEG-based emotion classification using conditional transfer learning. *Front. Hum. Neurosci.* **11**, 334 (2017)
11. Long, M., Cao, Y., Wang, J., Jordan, M.: Learning transferable features with deep adaptation networks. In: *International Conference on Machine Learning*, vol. 37, pp. 97–105. PMLR, Lille (2015)
12. Mathersul, D., Williams, L.M., Hopkinson, P.J., Kemp, A.H.: Investigating models of affect: relationships among EEG alpha asymmetry, depression, and anxiety. *Emotion* **8**(4), 560–572 (2008)

13. Mühl, C., Allison, B., Nijholt, A., Chanel, G.: A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges. *Brain-Comput. Interfaces* **1**(2), 66–84 (2014)
14. Pan, S.J., Tsang, I.W., Kwok, J.T., Yang, Q.: Domain adaptation via transfer component analysis. *IEEE Trans. Neural Netw.* **22**(2), 199–210 (2011)
15. Pan, S.J., Yang, Q.: A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22**(10), 1345–1359 (2010)
16. Sangineto, E., Zen, G., Ricci, E., Sebe, N.: We are not all equal: personalizing models for facial expression analysis with transductive parameter transfer. In: *ACM International Conference on Multimedia*, pp. 357–366. ACM Press, New York (2014)
17. Schölkopf, B., Smola, A., Müller, K.-R.: Kernel principal component analysis. In: Gerstner, W., Germond, A., Hasler, M., Nicoud, J.-D. (eds.) *ICANN 1997*. LNCS, vol. 1327, pp. 583–588. Springer, Heidelberg (1997). <https://doi.org/10.1007/BFb0020217>
18. Wang, X.W., Nie, D., Lu, B.L.: Emotional state classification from eeg data using machine learning approach. *Neurocomputing* **129**, 94–106 (2014)
19. Zheng, W., Lu, B.: Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Trans. Auton. Ment. Dev.* **7**(3), 162–175 (2015)
20. Zheng, W.L., Liu, W., Lu, Y., Lu, B.L., Cichocki, A.: Emotionmeter: a multimodal framework for recognizing human emotions. *IEEE Trans. Cybern.* **99**, 1–13 (2018)
21. Zheng, W.L., Lu, B.L.: Personalizing EEG-based affective models with transfer learning. In: *International Joint Conference on Artificial Intelligence*, pp. 2732–2738. AAAI Press, New York (2016)