Reducing the Subject Variability of EEG Signals with Adversarial Domain Generalization

Bo-Qun Ma[1], He Li[1], Wei-Long Zheng[2], and Bao-Liang Lu[1,3,4](✉)

[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
boqun.ma@hotmail.com, bllu@sjtu.edu.cn

[2] Department of Neurology, Massachusetts General Hospital,
Harvard Medical School, Boston, MA 02114, USA

[3] Key Laboratory of Shanghai Education Commission for Intelligent Interaction
and Cognition Engineering, Shanghai Jiao Tong University, Shanghai, China

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

Abstract. A major obstacle in generalizing brain-computer interface (BCI) systems to previously unseen subjects is the subject variability of electroencephalography (EEG) signals. To deal with this problem, the existing methods focus on domain adaptation with subject-specific EEG data, which are expensive and time consuming to collect. In this paper, domain generalization methods are introduced to reduce the influence of subject variability in BCI systems without requiring any information from unseen subjects. We first modify a deep adversarial network for domain generalization and then propose a novel adversarial domain generalization framework, DResNet, in which domain information is utilized to learn two components of weights: unbiased weights that are common across subjects and biased weights that are subject-specific. Experimental results on two public EEG datasets indicate that our proposed methods can achieve a performance comparable to and more stable than that of the state-of-the-art domain adaptation method. In contrast to existing domain adaptation methods, our proposed domain generalization approach does not require any data from test subjects and can simultaneously generalize well to multiple test subjects.

Keywords: Brain-computer interface · EEG subject variability · Domain adaptation · Domain generalization · Domain residual network · Emotion recognition · Vigilance estimation

1 Introduction

Brain-Computer Interface (BCI) systems focus on establishing a direct pathway between a human brain and an external device. As a reliable indicator of the human brain state, electroencephalography (EEG) has become a widely used modality in BCI systems [11]. In the past decades, EEG-based BCI systems have attracted researchers' interest and have been successfully applied in many applications [2]. However, the individual differences across subjects in the functional and anatomical connectivity of the

T. Gedeon et al. (Eds.): ICONIP 2019, LNCS 11953, pp. 30–42, 2019.
https://doi.org/10.1007/978-3-030-36708-4_3

brain, head shapes, mental states, etc., have become a major obstacle for BCI applications in real-life scenarios [15]. Conventional models trained with data recorded from one subject often fail to perform robustly on other subjects. Consequently, to obtain an effective model for a new subject, data recollecting and model retraining are required; unfortunately, such efforts are rather time consuming and expensive in practice.

Previous studies tackling the issue of subject variability can be classified into two categories: subject-dependent models with calibration and subject-independent models with features that are robust across subjects, according to the available information from new subjects. Several researchers have explored subject-dependent approaches in which the pretrained models are tuned with a small amount of calibration data recorded from test subjects [12]. The calibration phase needs to be repeated whenever the models are extended to new subjects; thus, good performance is achieved, but at a high cost. On the other hand, subject-independent approaches focus on extracting features that are robust across subjects for model training, thus achieving the necessary generalization ability to provide accurate predictions for new subjects [19]. However, if the calibration phase is removed, the models usually suffer compromised performance.

In recent years, efforts have been made to deal with the subject variability problem in BCI systems using transfer learning methods [7]. In traditional machine learning methods, it is assumed that the training data and test data are sampled from the same distribution; however, this assumption usually cannot be satisfied for cross-subject BCI systems. In contrast, transfer learning methods consider domain differences [18], thus allowing models trained on source-domain data to generalize well to the target domain. From the perspective of transfer learning, subject variability can be regarded as a kind of domain shift, i.e., distribution differences across several related domains.

Two of the main branches of transfer learning, domain adaptation (DA) and domain generalization (DG), are capable of reducing the influence of subject variability. DA methods enhance the performance of a model on the target domain by eliminating the domain shift between the source and target domains. Thus, these methods require acquaintance with the target-domain data in the training phase in order to measure the discrepancy between the source and target domains. Researchers have successfully applied DA methods in BCI systems [23]. In particular, deep adversarial models such as Domain-Adversarial Neural Network (DANNs) have achieved significant performance improvements [3,9]. However, since each individual is regarded as an independent domain in EEG-based BCI systems, DA methods, which require data collection and model training for each target domain (subject), are high in cost and low in efficiency. In particular, information from target subjects is usually unavailable in real-world cross-subject EEG-based BCI applications. One solution to these problems is to apply DG methods in EEG-based BCI systems. DG methods can extract domain-invariant features by exploiting domain differences across multiple source subjects without the need to acquire any data from the target subjects [1]. Therefore, systems based on DG models can perform robustly when applied to previously unseen domains.

In this paper, we aim to reduce subject variability in BCI systems without requiring any information from target subjects through two kinds of DG approaches. As the first approach, we adopt the Domain-Invariant Component Analysis (DICA) and Scatter Component Analysis (SCA) methods, proposed in [16] and [5], respectively.

We further apply deep adversarial networks to this problem by extending Domain-Adversarial Neural Network (DANN) to the DG condition (DG-DANN). These methods can project features from different domains into a domain-invariant feature space in which the dissimilarity among the domains can be reduced. Thus, the models can achieve a better generalization ability on new subjects. As the second approach, we exploit the information from the training domains to learn a set of regulated model weights. Inspired by [8], we propose a novel framework called the Domain Residual Network (DResNet) in which the network weights are explicitly divided into biased weights that are exclusive to each individual domain and unbiased weights that are shared by all domains. In this way, we can obtain a robust model that achieves a better generalization ability for unknown domains by means of the unbiased weights. In experiments, we evaluated the performance of these approaches on two different BCI tasks. We chose SEED, a public emotion recognition EEG dataset, for the classification evaluation and SEED-VIG, a public multimodal vigilance estimation dataset, for the regression evaluation.

2 Methods

2.1 Domain Generalization Problem

Given the input space $\mathcal{X}$ and the output space $\mathcal{Y}$, $\mathbb{P}_{XY}$ is the set of all joint distributions on $\mathcal{X} \times \mathcal{Y}$. We assume that the $P^i_{XY} \in \mathbb{P}_{XY}$ is observed from a distribution $\boldsymbol{P}$. A domain is denoted by $D_i = \{X_i, Y_i\}$, where the $\{X_i, Y_i\} = \{(x_1, y_1), (x_2, y_2), ..., (x_{n_i}, y_{n_i})\}$ are n_i samples from the joint distribution P^i_{XY}. Thus, we can obtain the marginal probability distribution P^i_X and the conditional probability distribution $P^i_{Y|X}$ of domain D_i. For k domains $D_1, D_2, ..., D_k$, we assume that the marginal distributions are different while the conditional distributions remain stable, i.e., $P^i_X \neq P^j_X, P^i_{Y|X} \approx P^j_{Y|X}$ when $i \neq j$. In the domain generalization problem, we aim to find a function $f : \mathcal{X} \to \mathcal{Y}$, which is insensitive to changes in P_X, to represent the conditional distribution $P_{Y|X}$. This f can be generalized to any previously unseen domain $D_t = \{X_t\}$, where the X_t are sampled from the unknown distribution P^t_X [1].

2.2 Domain-Invariant Component Analysis (DICA)

The goal of DICA is to find a low-dimensional feature subspace to minimize the discrepancy across domains [16]. Specifically, distributions can be represented as points in a reproducing kernel Hilbert space (RKHS) using the mean map function:

$$\mu : \mathbb{P}_x \to \mathcal{H} : P \mapsto \int_{\mathcal{X}} k(x, \cdot) dP(x) =: \mu_P. \tag{1}$$

Suppose that we have data samples $\mathcal{S} = \{S^i\}^k_{i=1} = \{(x^i_m, y^i_m)^{n_i}_{m=1}\}^k_{i=1}$ sampled from k domains. DICA can be applied to learn an orthogonal transformation $\mathcal{B}$ that minimizes the distributional variance across the different domains in a domain-invariant

m-dimensional feature subspace. The empirical distributional variance of $\mathcal{S}$ after the transformation can be calculated as:

$$\hat{\mathbb{V}}_{\mathcal{H}}(\mathcal{B}\mathcal{S}) = tr(B^T KLKB). \tag{2}$$

where K is the block kernel matrix, B is the coefficient matrix for transformation $\mathcal{B}$, and L is a coefficient matrix.

On the other hand, DICA also preserves the functional relationship P^i_{XY}. Given $\Phi_y = [\varphi(y_1), ..., \varphi(y_n)]$ and $U = \Phi_y^T \Phi_y$, the final objective function of DICA is

$$\max_{B \in \mathbb{R}^{n \times m}} \frac{\frac{1}{n} tr(B^T U(U + n\epsilon I_n)^{-1} K^2 B)}{tr(B^T KLKB + BKB)}, \tag{3}$$

where ϵ is a kernel regularizer. For further details, readers are referred to [16].

2.3 Scatter Component Analysis (SCA)

The goal of SCA is to find a projection B into an m-dimensional space where (1) the training domains are similar, (2) samples with the same label are similar, (3) samples with different labels are separated, and (4) the variance of the whole training set is maximized [5]. These constraints are quantified by means of a new concept called *scatter*:

$$\Psi_\phi(P) := \mathop{\mathbb{E}}_{x \sim P} \left[||\mu_P - \phi(x)||^2_{\mathcal{H}} \right], \tag{4}$$

where $|| \cdot ||_{\mathcal{H}}$ is the norm on $\mathcal{H}$. The four constraints mentioned above are quantified in terms of the following four scatters.

Domain Scatter. Given N samples $\{x_1, ..., x_N\}$ from a k-domain distribution set $\{P^i_X\}^k_{i=1}$ on $\mathcal{X}$, the domain scatter is defined with $\overline{\mu} = \frac{1}{k}\sum_{i=1}^k \mu_{P^i_X}$ as

$$\Psi(\{\mu_{P^i_X}\}^k_{i=1}) = \frac{1}{k}\sum_{i=1}^{k} ||\overline{\mu} - \mu_{P^i_X}||^2, \tag{5}$$

Class Scatter. Assuming the label set is $\{1, ..., C\}$, we denote the conditional distribution on $\mathcal{X}$ by $P^l_{X|t} = \frac{1}{k}\sum_{i=1}^k P^i_{XY}$, for $Y = t$. Therefore, the within-class scatter is defined as

$$\sum_{t=1}^{C} \Psi_{B \circ \phi}(\hat{P}^s_{X|y_t}) = Tr(B^T Q_s B) \tag{6}$$

and the between-class scatter is defined as

$$\Psi_B(\{\mu_{\hat{P}^l_{X|y_t}}\}^C_{t=1}) = Tr(B^T P_s B), \tag{7}$$

where $P_s = \sum_{t=1}^C n_t(m_t - \overline{m})(m_t - \overline{m})^T$ and $Q_s = \sum_{t=1}^C K_t H_t K_t^T$, with $m_k = \frac{1}{n_t}\sum_{n_t}^{i} k(\cdot, x_{it})$, $\overline{m} = \frac{1}{N}\sum_{i=1}^N k(\cdot, x_i)$, $[K_t]_{ij} = [k(x_{it}, x_{jt})]$, and $H_t = \mathbf{I}_{n_t} - \frac{1}{n_t}\mathbf{1}_{n_t}\mathbf{1}^T_{n_t}$.

Total Scatter. Given the total domain as calculated from the mean of the k domain distributions, namely, $P_X = \frac{1}{k}\sum_{i=1}^{k} P_X^i$, the total scatter can be derived by using B and K as follows:

$$\Psi_{B\circ\phi}(\hat{P}_X) = Tr(\frac{1}{N}B^T KKB). \tag{8}$$

The objective function of SCA for the DG problem is expressed as

$$\underset{B\in\mathbb{R}^{N\times m}}{argmax} \frac{\Psi_{B\circ\phi}(\hat{P}_X) + \Psi(\{\mu_{P^l_{X|t=k}}\}_{k=1}^{C})}{\Psi(\{\mu_{P^i_X}\}_{i=1}^{k}) + \sum_{t=1}^{C}\Psi_{B\circ\phi}(P^l_{X|t})}, \tag{9}$$

where $\beta, \delta > 0$ are hyperparameters. The objective function can be further rewritten as

$$(\frac{(1-\beta)}{N}KK + \beta P)B^* = (\delta KLK + K + Q)B^*\Lambda, \tag{10}$$

where $B^* = [b_1, ..., b_m]$ represents the first m eigenvectors and $\Lambda = diag(\lambda_1, ..., \lambda_m)$ represents the corresponding eigenvalues. According to [5], the solution to Eq. (9) consists of the m leading eigenvectors in Eq. (10).

2.4 Domain Generalization in Domain-Adversarial Neural Network

DANN is a deep adversarial domain adaptation model [3]. In this paper, we extend the DANN concept to the case of domain generalization (DG-DANN).

Specifically, there are three components in DG-DANN. Initially, the feature extractor G_f learns a feature mapping $G_f(x;\theta_f) = f(W_f x + b_f)$, where features are projected with an activation function f and parameters $\theta_f = \{W_f, b_f\} \in \mathbb{R}^{d\times p} \times \mathbb{R}^d$. Secondly, the label predictor G_y predicts the labels of the inputs by means of a function $G_y(G_f(X);\theta_y)$. The prediction loss on a sample (x_i, y_i) for a prediction $\hat{y_i}$ is denoted by $L_y(\hat{y_i}, y_i)$. Finally, the domain classifier $G_d(G_f(X);\theta_d)$ judges the source domain of each input feature.

In DA problems, the inputs are sampled from one source domain and one target domain. Thus, the domain classifier G_d is a binary classifier. According to [3], the loss of a binary G_d is defined as

$$L_d(G_d(G_f(x_i)), d_i) = d_i \log\frac{1}{G_d(G_f(x_i))} + (1-d_i)\log\frac{1}{1-G_d(G_f(x_i))} \tag{11}$$

for a sample (x_i, y_i, d_i), where d_i is the binary domain label of sample x_i.

In DG problems, the training data consist of N samples (x_i, y_i, d_i) from k different known domains. Following the idea of finding a domain-invariant feature space, we generalize the domain classifier G_d to a k-class domain classifier. Therefore, the loss of G_d can be modified as follows:

$$L_d(G_d(G_f(x_i)), d_i) = \log\frac{1}{G_d(G_f(x_i))_{d_i}}. \tag{12}$$

For brevity, we denote the loss of G_d by $L_d(\hat{d_i}, d_i)$, where $\hat{d_i}$ is the domain prediction for x_i. Therefore, the loss function of the DG-DANN is formulated as

$$E(\theta_f, \theta_y, \theta_d) = \frac{1}{N}\sum_{i=1}^{N} L_y(\hat{y}_i, y_i) - \lambda \frac{1}{N}\sum_{i=1}^{N} L_d(\hat{d_i}, d_i). \tag{13}$$

During optimization, the DANN is trained through a special layer called Gradient Reversal Layer (GRL), which connects G_f and G_d. The GRL can be ignored during forward propagation and reverses the gradient passed backward from G_d to G_f [3]. The optimization can be integrated as follows:

$$\begin{aligned} (\hat{\theta_f}, \hat{\theta_y}) &= arg \min_{\theta_f, \theta_y} E(\theta_f, \theta_y, \hat{\theta_d}), \\ (\hat{\theta_d}) &= arg \max_{\theta_d} E(\hat{\theta_f}, \hat{\theta_y}, \theta_d). \end{aligned} \tag{14}$$

2.5 Domain Residual Network (DResNet)

Another option for adversarial domain generalization is to utilize the domain information of the training domains to regulate the model parameters. We assume that each $P^i_{XY} \in \mathbb{P}_{XY}$ is a sample from a distribution $\boldsymbol{P}$. Thus, the domain shift can be regarded as the bias affecting observations of the true common space $\boldsymbol{P}$. According to [8], we can explicitly define the bias for each known training domain and approximate the parameters of the common space by undoing these biases. The common unbiased weights and the individual biased weights for each domain can be jointly trained to improve the generalization ability of the model. Based on the DG-DANN concept, we propose a novel model called the Domain Residual Network (DResNet) model.

In the DResNet model, the feature extractor G_f of the DG-DANN model is extended. The unbiased weights in D_f, which are shared by all domains, are denoted by θ^c_f. In contrast, the domain biases are explicitly described by biased weights $\theta^{\delta_i}_f$, which are unique for each known training domain D_i. Therefore, the parameters in each layer of G_f are formulated as follows:

$$\theta^i_f = \theta^c_f + \theta^{\delta_i}_f = \{W^c_f + W^{\delta_i}_f, b^c_f + b^{\delta_i}_f\}. \tag{15}$$

Hence, given an input x from domain i, each layer of the feature extractor G_f in DResNet is organized as follows:

$$G_f(x; \theta^i_f) = f\left((W^c_f x + b^c_f) + (W^{\delta_i}_f x + b^{\delta_i}_f)\right). \tag{16}$$

During backward propagation, for a sample (x_i, y_i, d_i) from domain D_i, the gradient in G_f simultaneously updates only the domain-specific $\theta^{\delta_i}_f$ and the common θ^c_f. After optimization, only the label predictor G_y and the common part of the feature extractor G_f are activated. The DResNet architecture is described in Fig. 1.

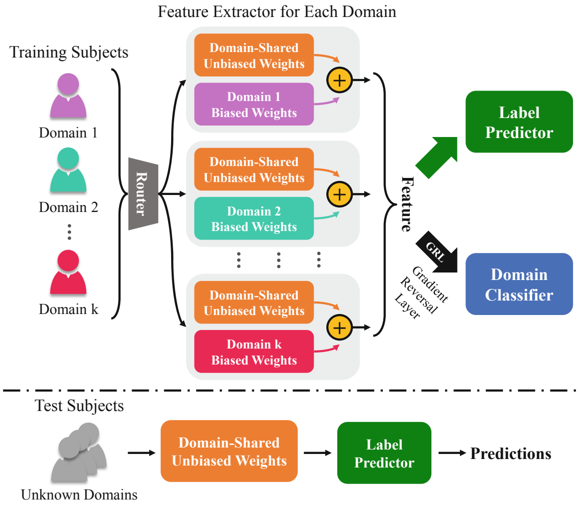


Fig. 1. The DResNet architecture. The colors for each domain in the training set indicate different domain shifts. For the test subjects, the domain shifts are unknown, and only the unbiased weights are activated. (Color figure online)

3 Experimental Setup

3.1 The SEED Dataset

The SEED[1] dataset is a public affective EEG dataset for emotion recognition. For SEED, 15 healthy subjects were recruited to be the participants in 3 sessions of experiments. Each experiment consisted of 15 trials of Chinese emotional film clips, which were selected in a preliminary study to induce 3 kinds of emotional states: positive, negative and neutral. The subjects were requested to exhibit their own corresponding emotions while watching the affective film clips. During the experiment, 62-electrode EEG signals were recorded in accordance with the international 10–20 system using an ESI Neuroscan system. The EEG signals were first downsampled to 200 Hz and then processed with a bandpass filter of 1–75 Hz. Finally, differential entropy (DE) features were extracted with nonoverlapping 1 s time windows in the five frequency bands (δ: 1–3 Hz, θ: 4–7 Hz, α: 8–13 Hz, β: 14–30 Hz, and γ: 31–50 Hz) [22]. For each subject, 3394 samples of 310-dimensional features were collected.

3.2 The SEED-VIG Dataset

The SEED-VIG[2] dataset is a public multimodal vigilance estimation dataset including EEG and electrooculography (EOG) signals [24]. Using SMI eye tracking glasses, the

[1] http://bcmi.sjtu.edu.cn/~seed/seed.html.
[2] http://bcmi.sjtu.edu.cn/~seed/seed-vig.html.

data were labeled with the percentage of eye closure (PERCLOS) [4], which is a continuous number varying from 0 (drowsy) to 1 (alert). For SEED-VIG, 23 subjects participated in the driving experiment, which was conducted in a simulation system consisting of a large screen, a real car and a corresponding software system. Forehead EEG and EOG signals were collected during the experiment with 4 electrodes using a ESI Neuroscan system. The data were first downsampled to 125 Hz and then segmented with nonoverlapping 8 s time windows. To separate the EEG and EOG components from the mixed signals, we applied the independent component analysis method. We extracted DE features from the EEG components in adjacent nonoverlapping 2 Hz bands within the range from 1 Hz to 50 Hz, thus obtaining 100-dimensional EEG features. On the other hand, the EOG components were processed with the Mexican hat wavelet transform to extract 36 eye movement features, including blinks, saccades and fixation. By concatenating the EEG and EOG features, 885 samples of 136-dimensional features were extracted for each subject.

3.3 Evaluation Details

To compare the DG and DA methods in terms of prediction accuracy, we adopted leave-one-subject-out cross-validation. In each iteration, for the DG methods, we selected one subject as the test domain and the others as the training domains, while for the DA methods, all training subjects were considered as one source domain and the test subject as the target domain. According to the total numbers of subjects, 15 and 23 iterations were performed for SEED and SEED-VIG, respectively.

For comparison in terms of generalization ability, we designed another setting called leave-multiple-random-subjects-out cross-validation, which also consisted of several iterations. In each iteration, one-third of the subjects were randomly selected as the test domains and the others as the training domains. To maintain the granularity of the evaluation, the numbers of iterations in this setting were the same as in the first setting.

For the kernel-based conventional methods, we adopted a linear kernel function. A subset of the samples in SEED (1000 samples for each subject) was randomly selected as the training data because of the practical infeasibility of loading all the training data due to the limited available memory and computation time. For dimensional reduction, the number of subspace dimensions was selected from the range of $\{10, 20, ..., 120\}$. For the shallow models, the parameters were randomly sought in the range $\{2^n | n \in \{-10, ..., 10\}\}$. For the deep models, we applied the Adam optimizer and a random search strategy. The search spaces for the learning rate and the hyperparameter λ for GRL were set to $\{2^n \times 10^{-4} | n \in [-10, 10]\}$ and $\{10^n | n \in [-5, -1]\}$, respectively.

4 Results and Discussion

4.1 Leave-one-subject-out Evaluation

Emotion Recognition. The performance of the DG methods for the classification task was evaluated on the SEED dataset. We adopted the leave-one-subject-out evaluation scheme and compared the DG methods with several conventional DA methods, such as

TCA [17] and TPT [20], as well as deep DA methods, such as DANN [3], DAN [10] and WGANDA [13]. Table 1 presents the mean accuracies (Avg) and standard deviations (Std). The baseline SVM method shows relatively poor performance due to the subject variability between the training subjects and the test subject. Among the shallow methods, TPT outperforms the other methods with an accuracy of 75.17% [23], while SCA and DICA achieve lower accuracy but more stable performance. Among the deep methods, WGANDA achieves the best performance with a mean accuracy of 87.07% [13]. In addition, the deep DG methods are also effective, exhibiting comparable performance, with DResNet being slightly better than DG-DANN. In general, the DA methods perform the best due to the additional information from the test subject. However, the DA methods require a large amount of unlabeled data from the test subjects to measure the discrepancy between the source and target domains. By comparison, the DG methods are capable of achieving the same level of prediction accuracy as the DA methods while requiring no data from the test subjects.

Table 1. Leave-one-subject-out evaluation results for classification on SEED

	Baseline	Domain adaptation methods					Domain generalization methods			
	SVM	TCA	TPT	DANN	DAN	WGAN-DA	DICA	SCA	DG-DANN	DResNet
Avg	0.5818	0.6400	0.7517	0.7919	0.8381	**0.8707**	0.6941	0.6633	0.8430	0.8530
Std	0.1385	0.1466	0.1283	0.1314	0.0856	**0.0714**	0.0779	0.1060	0.0832	0.0797

Vigilance Estimation. We also investigated the effectiveness of the proposed DG methods for the regression task on the SEED-VIG dataset. The Pearson correlation coefficient (PCC) and the root-mean-square error (RMSE) were calculated for the evaluation. Support vector regression (SVR) with a linear kernel was chosen as the baseline method for vigilance estimation. We compared the DG methods with two shallow DA methods, TCA [17] and GFK [6], as well as the latest deep DA methods, DANN [3] and ADDA [21]. As shown in Table 2, the DG models achieve stable performance that is comparable to that of the DA models; ADDA shows the best accuracy, with a PCC of 0.8442 and an RMSE of 0.1405 [9]. The performance of DResNet (PCC: 0.8440, RMSE: 0.1420) is quite similar to that of the state-of-the-art methods on the same task, even without additional data from the test subjects. In terms of performance stability, DResNet and DG-DANN outperform the other methods. These results are consistent with the conclusions summarized for the emotion recognition task.

Table 2. Leave-one-subject-out evaluation results for regression on SEED-VIG

		Baseline	Domain adaptation methods				Domain generalization methods		
		SVR	TCA	GFK	DANN	ADDA	DICA	DG-DANN	DResNet
PCC	Avg	0.7606	0.7786	0.7907	0.8402	**0.8442**	0.7733	0.8320	0.8440
	Std	0.2314	0.2152	0.1260	0.1535	0.1336	0.1382	0.1000	**0.0935**
RMSE	Avg	0.1689	0.1596	0.1910	0.1427	**0.1405**	0.2007	0.1470	0.1420
	Std	0.0673	0.0544	0.0636	0.0588	0.0514	0.0674	0.0444	**0.0402**

4.2 Leave-multiple-random-subjects-out Evaluation

As mentioned above, for practical BCI applications, DA methods become ineffective when extended to multiple unknown test subjects with only one well-trained model. To evaluate the generalization ability of the DG models under these circumstances, we adopted the leave-multiple-random-subjects-out cross-validation scheme. The experimental results of the baseline SVM method and all DG methods on SEED and SEED-VIG are shown in Tables 3 and 4, respectively. The performance drops slightly due to the decreased size of the training set. Here, DResNet outperforms the other methods on both datasets, achieving an accuracy improvement of 27.57% compared to the SVM model on SEED and a PCC improvement of 0.0887 compared to the baseline SVR model on SEED-VIG.

Table 3. Leave-multiple-random-subjects-out evaluation results on SEED

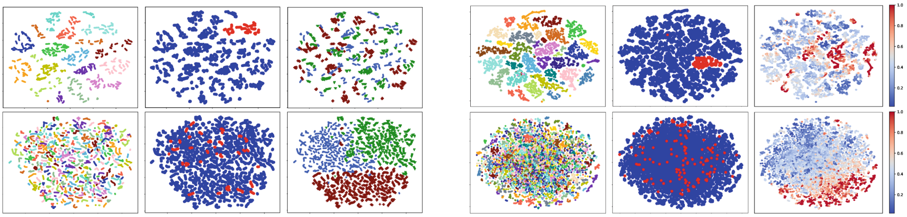
	SVM	DICA	SCA	DG-DANN	DResNet
Avg	0.5413	0.6435	0.6083	0.8146	**0.8170**
Std	0.1348	0.0896	**0.0505**	0.0788	0.0737

Table 4. Leave-multiple-random-subjects-out evaluation results on SEED-VIG

		SVR	DICA	DG-DANN	DResNet
PCC	Avg	0.7499	0.7719	0.8294	**0.8386**
	Std	0.1980	0.1841	0.1541	**0.1532**
RMSE	Avg	0.2068	0.1735	0.1604	**0.1569**
	Std	0.0587	**0.0468**	0.0782	0.0735

4.3 Discussion

To further investigate the effectiveness of the DG models on features extracted from different domains, we visualized the features from the leave-one-subject-out evaluation using the t-SNE algorithm [14]. The visualization results are depicted in Fig. 2. The first row shows the raw features from the datasets, while the second row shows the features extracted by the DResNet feature extractor. In the first column, the features are colored in accordance with their source subjects. In the second column, the blue points represent the training data, and the red points represent the test data. Finally, we visualize all features in accordance with their labels in the third column. The features from SEED are colored with red, blue and green, which represent positive, negative and neutral emotions, respectively. The features from SEED-VIG are colored in accordance with their PERCLOS labels, where the red points denote lower vigilance levels.



(a) SEED Feature Visualization

(b) SEED-VIG Feature Visualization

Fig. 2. Domain generalization feature visualization.

Firstly, the phenomenon of subject variability is clearly evident in the raw features in the first colume. After DResNet processing, the subject variability is significantly reduced since the data from different domains are evenly mixed together. In addition, the figures in the second colume demonstrate the reason for the remarkable performance of the DG models, since the training data and test data are aligned with similar distributions. Furthermore, it can be observed that the DResNet features vary smoothly with their labels in the third column and thus can be more easily predicted by the label predictor.

5 Conclusion

In this paper, we focused on reducing the influence of EEG subject variability on BCI systems for unknown subjects. DG methods were introduced to address this problem without needing to collect additional information from the test subjects. Following two different approaches to DG, we generalized the DANN concept and then proposed a novel framework called DResNet. In evaluations on classification and regression tasks, we compared our methods with other DA and DG methods on two public datasets related to different topics. We applied two different schemes for evaluation in terms of prediction accuracy and generalization ability. The experimental results show that the proposed methods are effective for solving the subject variability problem in cross-subject BCI systems for unknown users.

Acknowledgments. This work was supported in part by the grants from the National Key Research and Development Program of China (Grant No. 2017YFB1002501), the National Natural Science Foundation of China (Grant No. 61673266), and the Fundamental Research Funds for the Central Universities.

References

1. Blanchard, G., Lee, G., Scott, C.: Generalizing from several related classification tasks to a new unlabeled sample. In: Advances in Neural Information Processing Systems, pp. 2178–2186 (2011)
2. Brunner, C., et al.: BNCI horizon 2020: towards a roadmap for the BCI community. Brain-Comput. Interfaces **2**(1), 1–10 (2015)

3. Ganin, Y., et al.: Domain-adversarial training of neural networks. J. Mach. Learn. Res. **17**(1), 2030–2096 (2016)
4. Gao, X.Y., Zhang, Y.F., Zheng, W.L., Lu, B.L.: Evaluating driving fatigue detection algorithms using eye tracking glasses. In: 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER), pp. 767–770. IEEE (2015)
5. Ghifary, M., Balduzzi, D., Kleijn, W.B., Zhang, M.: Scatter component analysis: a unified framework for domain adaptation and domain generalization. IEEE Trans. Pattern Anal. Mach. Intell. **39**(7), 1414–1430 (2017)
6. Gong, B., Shi, Y., Sha, F., Grauman, K.: Geodesic flow kernel for unsupervised domain adaptation. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2066–2073. IEEE (2012)
7. Jayaram, V., Alamgir, M., Altun, Y., Scholkopf, B., Grosse-Wentrup, M.: Transfer learning in brain-computer interfaces. IEEE Comput. Intell. Mag. **11**(1), 20–31 (2016)
8. Khosla, A., Zhou, T., Malisiewicz, T., Efros, A.A., Torralba, A.: Undoing the damage of dataset bias. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part I. LNCS, vol. 7572, pp. 158–171. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33718-5_12
9. Li, H., Zheng, W.L., Lu, B.L.: Multimodal vigilance estimation with adversarial domain adaptation networks. In: 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1–6. IEEE (2018)
10. Long, M., Cao, Y., Wang, J., Jordan, M.: Learning transferable features with deep adaptation networks. In: the 32nd International Conference on Machine Learning, vol. 37, pp. 97–105. PMLR (2015)
11. Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., Arnaldi, B.: A review of classification algorithms for EEG-based brain-computer interfaces. J. Neural Eng. **4**(2), R1 (2007)
12. Lotte, F., Guan, C.: Learning from other subjects helps reducing brain-computer interface calibration time. In: IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), pp. 614–617 (2010)
13. Luo, Y., Zhang, S.-Y., Zheng, W.-L., Lu, B.-L.: WGAN domain adaptation for EEG-based emotion recognition. In: Cheng, L., Leung, A.C.S., Ozawa, S. (eds.) ICONIP 2018, Part V. LNCS, vol. 11305, pp. 275–286. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-04221-9_25
14. Maaten, L.V.D., Hinton, G.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**, 2579–2605 (2008)
15. Morioka, H., et al.: Learning a common dictionary for subject-transfer decoding with resting calibration. NeuroImage **111**, 167–178 (2015)
16. Muandet, K., Balduzzi, D., Schölkopf, B.: Domain generalization via invariant feature representation. In: International Conference on Machine Learning, pp. 10–18 (2013)
17. 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)
18. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (2010)
19. Samek, W., Kawanabe, M., Müller, K.R.: Divergence-based framework for common spatial patterns algorithms. IEEE Rev. Biomed. Eng. **7**, 50–72 (2014)
20. Sangineto, E., Zen, G., Ricci, E., Sebe, N.: We are not all equal: personalizing models for facial expression analysis with transductive parameter transfer. In: the 22nd ACM International Conference on Multimedia, pp. 357–366. ACM (2014)
21. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7167–7176 (2017)

22. Zheng, W.L., Lu, B.L.: Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. IEEE Trans. Auton. Mental Dev. **7**(3), 162–175 (2015)
23. Zheng, W.L., Lu, B.L.: Personalizing EEG-based affective models with transfer learning. In: The Twenty-Fifth International Joint Conference on Artificial Intelligence, pp. 2732–2738. AAAI Press (2016)
24. Zheng, W.L., Lu, B.L.: A multimodal approach to estimating vigilance using EEG and forehead EOG. J. Neural Eng. **14**(2), 026017 (2017)