WGAN Domain Adaptation for EEG-Based Emotion Recognition

Yun Luo[1], Si-Yang Zhang[1], Wei-Long Zheng[1], 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
[2] Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognition 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
{angeleader,zhangsiyang-sjtu,weilong,bllu}@sjtu.edu.cn

Abstract. In this paper, we propose a novel Wasserstein generative adversarial network domain adaptation (WGANDA) framework for building cross-subject electroencephalography (EEG)-based emotion recognition models. The proposed framework consists of GANs-like components and a two-step training procedure with pre-training and adversarial training. Pre-training is to map source domain and target domain to a common feature space, and adversarial-training is to narrow down the gap between the mappings of the source and target domains on the common feature space. A Wasserstein GAN gradient penalty loss is applied to adversarial-training to guarantee the stability and convergence of the framework. We evaluate the framework on two public EEG datasets for emotion recognition, SEED and DEAP. The experimental results demonstrate that our WGANDA framework successfully handles the domain shift problem in cross-subject EEG-based emotion recognition and significantly outperforms the state-of-the-art domain adaptation methods.

Keywords: EEG · Emotion recognition · Domain adaptation · GAN

1 Introduction

With rapid development of affective computing and emotional intelligence, affective brain-computer interfaces (aBCIs) have recently attracted widespread attention [13]. aBCIs aim to equip machines with the ability to detect users' affective states from neurophysiological signals and provide humanized interactions. Recently, many researchers have made significant progresses in EEG-based emotion recognition models, especially in subject-specific models [1,8,10,21]. However, due to domain shift [18] caused by the non-stationary nature of EEG signals

L. Cheng et al. (Eds.): ICONIP 2018, LNCS 11305, pp. 275–286, 2018.
https://doi.org/10.1007/978-3-030-04221-9_25

and structural variability between individuals [12,15], an EEG-based emotion recognition model trained with data from one specific subject usually does not generalize well to another. In practical aBCI applications, a cross-subject emotion recognition model which is capable of recognizing the emotions of a new subject with unlabeled data is required rather than a subject-specific one. To deal with the domain shift problem caused by inter-subject variability, we focus on developing cross-subject emotion recognition approach in this work.

A promising solution to the domain shift problem is to take advantage of the domain adaptation methods. The basic idea of these methods is to transfer knowledge from source domain to unlabeled target domain. Under the circumstance of domain shift, marginal probability distributions of source domain and target domain are different. Domain adaptation methods are able to handle this difference by mapping features of both domains into a common feature space, where the marginal probability distributions of the two mappings are similar.

Various domain adaptation methods have been developed to find the common feature space for source and target subjects. Most of them aim to minimize some metrics between two probability distributions, such as maximum mean discrepancy (MMD) [5]. For example, transfer component analysis (TCA) [14], a typical domain adaptation method, minimizes MMD between distributions of source and target domains by constructing kernel matrix. This method, along with kernel principle component analysis (KPCA) [17] and transductive parameter transfer (TPT) [16], has been successfully used for implementing personalized EEG-based emotion models [23].

An alternative way of finding the common space is to leverage the transferability of deep neural networks [9]. One of attractive approaches is to apply generative adversarial domain adaptation [19], which is closely related to generative adversarial networks (GANs) [4]. The adversarial training procedure of GANs can be formulated as a minimax problem. When the game achieves its equilibrium, the distribution of generated data is approximate to the distribution of real data. By taking advantage of the generative ability of GANs, generative adversarial domain adaptation methods have made considerable progresses in dealing with the domain shift problem in computer vision [19].

In this paper, we adopt the generative adversarial domain adaptation method to build a cross-subject EEG-based emotion recognition framework. Our work is based on Wasserstein GAN [2], which is an improved stable version of traditional GAN. Instead of using the source subject features, we consider their mappings in a new feature space as the real data distribution, which has been adopted in Adversarial Discriminative Domain Adaptation (ADDA) as well [19]. The features of the target subjects are mapped to the same feature space, in which their mappings are considered as the generated distribution. The distance between marginal probability distributions of the two mappings are reduced through adversarial training, and then the domain shift problem is fixed.

Our proposed Wasserstein GAN domain adaptation (WGANDA) framework aims to solve the domain shift problem in EEG-based emotion recognition caused by inter-subject variability. Compared with subject-specific models, the proposed

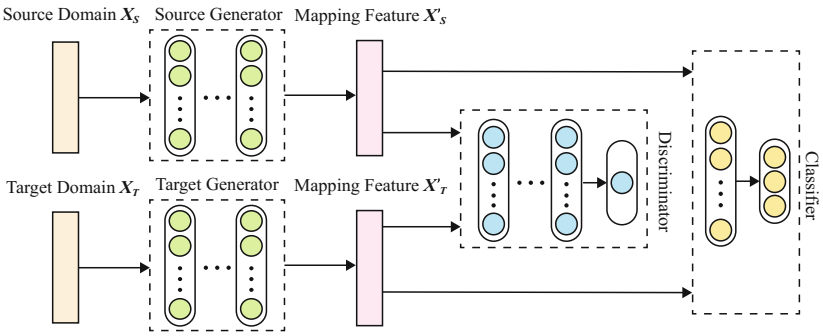

Fig. 1. Illustration of the proposed WGANDA framework, which consists of four parts: the source and target generators for mapping source domain and target domain to a common feature space, the discriminator for distinguishing source and target distribution in the common feature space, and the classifier for recognizing emotional states.

cross-subject framework makes better use of the EEG data collected from different subjects. The framework is also able to recognize the emotions of a new subject with unlabeled data more precisely. The application of Wasserstein GAN in this work overcomes the gradients vanish and instability problems of traditional GANs' training procedure. Besides, the implementation of the gradient-penalty Wasserstein GAN loss [6] speeds up the convergence process. According to experimental results on two public EEG datasets, the proposed WGANDA framework significantly outperforms the state-of-the-art domain adaptation methods.

2 Methods

2.1 Notations and Framework Structure

Our proposed framework consists of four components as shown in Fig. 1. Assume that a labeled dataset X_s is collected from the source subjects, and an unlabeled dataset X_t is collected from the target subjects:

$$X_s = \{x_s^i\}_{i=1}^m,\ Y_s = \{y_s^i\}_{i=1}^m,\ X_t = \{x_t^i\}_{i=1}^n \tag{1}$$

where m and n represent the numbers of data samples in source and target datasets, respectively.

The source generator ψ_s and the target generator ψ_t map source data X_s and target data X_t to a common feature space, respectively:

$$X'_s = \psi_s(X_s),\ X'_t = \psi_t(X_t) \tag{2}$$

where X'_s and X'_t are expected to have the same feature dimensions.

The classifier C takes X'_s and X'_t as inputs and outputs emotion predictions, Y_{sp} and Y_{tp}, as follows:

$$Y_{sp} = C(X'_s),\ \textit{and}\ Y_{tp} = C(X'_t). \tag{3}$$

Algorithm 1. The work flow of the proposed WGANDA framework

Input: Source domain dataset $X_s = \{x_s^i\}_{i=1}^m, Y_s = \{y_s^i\}_{i=1}^m$ and target domain dataset $X_t = \{x_t^i\}_{i=1}^n$

Output: Predicted target domain dataset labels Y_{tp}

1: Update θ_s and θ_c by descending along their gradients:

$$\nabla_{\theta_s,\theta_c}\left[-\frac{1}{m}\sum_{i=1}^{m}\sum_{h=1}^{H}\mathbb{I}(y_s^i = h)logC(\psi_s(x_s^i))\right]$$

2: Initialize θ_t with θ_s;

3: **repeat**

4: **for** *critic* iterations **do**

5: Update θ_d by ascending along its gradient:

$$\nabla_{\theta_d}\left[\frac{1}{m}\sum_{i=1}^{m}D(\psi_s(x_s^i)) - \frac{1}{n}\sum_{i=1}^{n}D(\psi_t(x_t^i)) - \frac{\lambda}{q}\sum_{i=1}^{q}(||\nabla_{\hat{x}^i}D(\hat{x}^i)||_2 - 1)^2\right]$$

6: **end for**

7: Update θ_t by descending along its gradient:

$$\nabla_{\theta_t}\left[-\frac{1}{n}\sum_{i=1}^{n}D(\psi_t(x_t^i))\right]$$

8: **until** convergence

9: Predict target domain dataset labels with the target generator and the classifier:

$$Y_{tp} = C(\psi_t(X_t))$$

10: **return** Y_{tp}

Then a discriminator D is applied to distinguish X_s' and X_t'. Note that all four components are parameterized by feedforward neural networks. The parameters of the source generator, the target generator, the classifier, and the discriminator are represented with θ_s, θ_t, θ_c and θ_d, respectively.

2.2 Training Procedure

The training procedure of the WGANDA framework consists of the following two steps:

(i) **Pre-training:** feed source data X_s through the source generator ψ_s to the classifier C, minimize cross-entropy loss with source dataset labels Y_s, and initialize target generator parameters θ_t with source generator parameters θ_s.

(ii) **Adversarial-training:** train the network through an adversarial way and update discriminator parameters θ_d as well as target generator parameters θ_t alternatively with source data X_s and target data X_t. Note that in each

adversarial iteration, the discriminator is updated a certain number of times denoted with *critic*, while the target generator is updated only once.

After the two-step training procedure, recognition accuracy can be calculated by feeding X_t' to the pre-trained classifier C. The whole work flow of the proposed WGANDA framework is described in Algorithm 1.

In the pre-training step, our goal is to minimize the cross-entropy loss by optimizing θ_s and θ_c:

$$\min_{\theta_s,\theta_c} L_C(X_s, Y_s) = -\mathbb{E}_{(x_s,y_s)\sim(X_s,Y_s)}\left[\sum_{h=1}^{H} \mathbb{I}(y_s = h) log C(\psi_s(x_s))\right] \tag{4}$$

where H is the number of emotion states. Then θ_s is fixed through the following adversarial-training step, and θ_c is fixed for the final target emotion prediction.

We initialize θ_t with θ_s when L_C is minimized. Without this target generator parameter initialization step, the discriminator can easily distinguish samples from X_s' and samples from X_t', which makes it hard to optimize target generator. Initializing θ_t with θ_s ensures the distribution of X_t is relatively close to X_s. The discriminator will thus not be able to distinguish the two distributions too easily, and target generator can be optimized faster.

In the adversarial-training step, the network is trained to narrow down the gap between marginal distributions $P(X_s)$ and $P(X_t)$. With fixed θ_s and θ_c, the framework can be treated as a typical GAN model. However, the traditional training procedure of GANs is prone to fall into model collapse, and it is troubled with gradients vanish as well. To prevent these two drawbacks, we implement Wasserstein GAN loss with gradient-penalty rather than traditional GANs' adversarial loss, which is applied in ADDA.

The training procedure of traditional GANs can be viewed as minimizing the Jensen-Shannon divergence between the real and generated distributions. As a metric for the distance of two distributions, Jensen-Shannon divergence is discontinuous, which makes it difficult to provide useful gradients for optimizing the generator. It is also the main reason of the GANs' instability. The Wasserstein GAN adopts Earth-Mover distance (EMD, also called Wasserstein-1) to eliminate the instability problem [2]. The EMD between two distributions is:

$$W(X_r, X_g) = \inf_{\gamma\sim\Pi(X_r,X_g)} \mathbb{E}_{(x_r,x_g)\sim\gamma}[||x_r - x_g||] \tag{5}$$

where $\Pi(X_r, X_g)$ denotes all possible joint distributions of real distribution X_r and generated distribution X_g defined in traditional GANs. The EMD is almost continuous and differentiable almost everywhere, and thus overcomes the instability problem. Since the infimum of Eq. (5) is computationally highly intractable, its Kantorovich-Rubinstein duality form is usually utilized [20]:

$$W(X_r, X_g) = \frac{1}{K} \sup_{||f||_L \le K} \mathbb{E}_{x_r\sim X_r}[f(x_r)] - \mathbb{E}_{x_g\sim X_g}[f(x_g)] \tag{6}$$

where f denotes the set of 1-Lipschitz functions. In realistic implementations, f is replaced by discriminator D and $||f||_L \leq K$ is replaced by $||D||_L \leq 1$. The loss function of Wasserstein GAN is then formulated by:

$$\min_{\theta_G} \max_{\theta_D} L(X_r, X_g) = \mathbb{E}_{x_r \sim X_r}[(D(x_r))] - \mathbb{E}_{x_g \sim X_g}[(D(x_g))] \tag{7}$$

where θ_D and θ_G represent the parameters of discriminator and generator in traditional GANs, respectively. The discriminator realizes 1-Lipschitz function by clipping the weights and constraining them within a bounded range.

Gulrajani *et al.* enforced Lipschitz constraint with gradient penalty instead of weight clipping to directly constrain the gradient norm [6], which makes the training procedure more stable and make convergence faster. An extra penalty term is appended to the loss function in their approach:

$$\begin{aligned} \min_{\theta_G} \max_{\theta_D} L(X_r, X_g) = \mathbb{E}_{x_r \sim X_r}[(D(x_r))] - \mathbb{E}_{x_g \sim X_g}[(D(x_g))] \\ - \lambda \mathbb{E}_{\hat{x} \sim \hat{X}}[(||\nabla_{\hat{x}} D(\hat{x})||_2 - 1)^2] \end{aligned} \tag{8}$$

where λ is a hyperparameter controlling the trade-off between original objective and gradient penalty, and $\hat{x}$ denotes the data points sampled from the straight line between real distribution X_r and generator distribution X_g:

$$\hat{x} = \alpha x + (1 - \alpha)\tilde{x}, \alpha \sim U[0, 1], x \sim X_r, \tilde{x} \sim X_g \tag{9}$$

In Algorithm 1, the number of sampled data points is denoted with q.

Our WGANDA framework can be treated as a Wasserstein GAN when the source generator is fixed. In this case, X'_s and X'_t correspond to the real data X_r and the generated data X_g in traditional GANs, respectively. We present our adversarial loss in Wasserstein GAN gradient penalty form as follows. First, the discriminator is trained by maximizing the discriminator loss (D-Loss) with target generator fixed:

$$\begin{aligned} \max_{\theta_d} L_D(X_s, X_t) = \mathbb{E}_{x_s \sim X_s}[D(\psi_s(x_s))] - \mathbb{E}_{x_t \sim X_t}[D(\psi_t(x_t))] \\ - \lambda \mathbb{E}_{\hat{x} \sim \hat{X}}[(||\nabla_{\hat{x}} D(\hat{x})||_2 - 1)^2] \end{aligned} \tag{10}$$

Then the target generator is trained by minimizing the generator loss (G-Loss) with discriminator fixed:

$$\min_{\theta_t} L_G(X_t) = -\mathbb{E}_{x_t \sim X_t}[D(\psi_t(x_t))] \tag{11}$$

The two losses are optimized in an alternating procedure, and the parameters of different components are updated in an interleaved manner. Note that in Wasserstein GANs, the discriminator aims to fit the 1-Lipschitz function. In each adversarial training iteration, the discriminator is fully trained to its optimization. Thus θ_d is updated for *critic* times and θ_t is updated only once in each adversarial training iteration. When the discriminator is fully trained, D-Loss represents the EMD between the marginal distribution of X'_s and X'_t. In our experiments, D-Loss is used as an indicator of training process.

The assumption of most domain adaptation methods is that, the conditional probability distributions of source domain and target domain equal when the marginal probability distributions of source domain and target domain are the same. When D-Loss converges, the marginal distribution of X'_s is approximate to the marginal distribution of X'_t:

$$P(X'_s) \approx P(X'_t) \tag{12}$$

According to the assumption mentioned above, the conditional distribution of X'_s and the conditional distribution of X'_t are also similar:

$$P(Y'_s|X'_s) \approx P(Y'_t|X'_t) \tag{13}$$

where Y'_t denotes the true labels of the dataset collected from target subject. Under this circumstance, the classifier pre-trained with X'_s is able to recognize the emotions of the target subject from X'_t. Thus after the adversarial-training procedure, we feed X_t to the pre-trained classifier and compare the output Y_{tp} with its true label to get the recognition accuracy.

3 Experiment Settings

3.1 EEG Datasets

We evaluate our framework on two public EEG datasets, SEED[1] [22] and DEAP[2] [7]. The SEED dataset consists of 15 participants. Each of them was required to watch 15 emotional film clips to elicit three emotions: positive, neutral, and negative. The EEG signals were recorded at a sampling rate of 1000 Hz with ESI NeuroScan System, which had a 62 electrode cap. The data in DEAP are formed with 8-channel peripheral physiological signals and 32-channel EEG signals. 32 participants watched 40 music videos and their EEG signals were collected by an international 10–20 system. The level of each video was rated 1–9 by the participants in terms of arousal, valence, like, and dislike.

The EEG signals of both datasets are preprocessed before feeding to the framework. Differential entropy (DE) features are extracted per second from five frequency bands for SEED dataset: δ: 1–3 Hz, θ: 4–7 Hz, α: 8–13 Hz, β: 14–30 Hz, and γ: 31–50 Hz [3,22]. The feature dimension is 310 (62 channels $\times$ 5 frequency bands) and the number of samples for each subject is 3394. For DEAP dataset, the DE features are extracted per second except for δ frequency since the low frequency band is filtered in this dataset. The feature dimension is 128 (32 channels $\times$ 4 frequency bands) and the number of samples for each subject is 2400. Valence model (high valence: level $>$ 5, low valence: level $\leq$ 5) and arousal model (high arousal: level $>$ 5, low arousal: level $\leq$ 5) are adopted in this work.

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

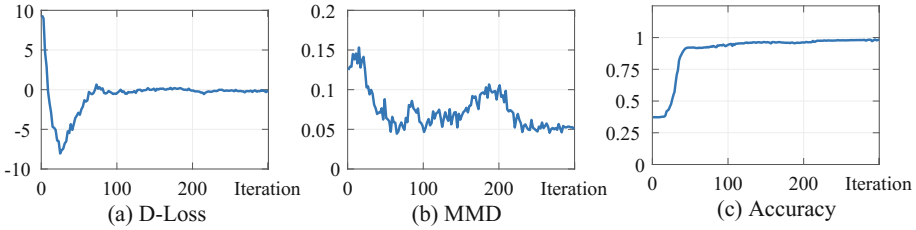

Fig. 2. Discriminator loss (D-Loss) (a), MMD (b) and accuracy (c) tendency along with training steps of SEED dataset.

3.2 Evaluation Details

To demonstrate the effectiveness of the proposed framework, a leave-one-subject-out cross validation is conducted. We chose one subject as the target subject and leave the others (14 for SEED, and 31 for DEAP) as source subjects.

To optimize the network structure, we perform grid search on the number of network layers. The numbers of layers are searched from 3 to 6 for both generator and discriminator. Each hidden layer of both the source generator network and the target generator network has 512 nodes for SEED dataset and 256 nodes for DEAP dataset. The outputs of the two generators have the same dimension as the input data, which is 310 for SEED dataset and 128 for DEAP dataset. Each hidden layer of discriminator network has the same number of nodes as the hidden layers of the generators, and the output has only one dimension. For the classifier, the numbers of network layers are searched from 1 to 3. The output dimension is 3 and 2 for SEED and DEAP datasets, respectively. Each hidden layer of the classifier network has 64 nodes. The ReLU activation function is used for all hidden layers.

In our experiments, we observe that the loss of discriminator is fluctuating with less discriminator training iterations in each round. And the discriminator should be fully optimized to ensure the convergence in each adversarial training iteration according to the theory of Wasserstein GANs. So the *critic* value is set to 20 to ensure the convergence and training speed. It means that we update discriminator 20 times and update target generator once in each adversarial-training iteration. Besides, Adam optimizer is more likely to cause fluctuation than RMSProp optimizer. Thus we use RMSProp optimizer during adversarial-training and Adam optimizer during pre-training. To speed up the training procedure, we use mini-batch instead of full batch shown in Algorithm 1. The size of mini-batch is set to 256. And the hyperparameter λ is set to 10.

MMD is frequently used as a measurement of the distance between two distributions [9,14], thus we adopt it in this work to evaluate the distance between the probability distributions of X'_s and X'_t, and demonstrate the effectiveness of our framework.

We use the recognition results before adversarial-training as baseline to show the ability of adversarial domain adaptation. In order to evaluate the effective-

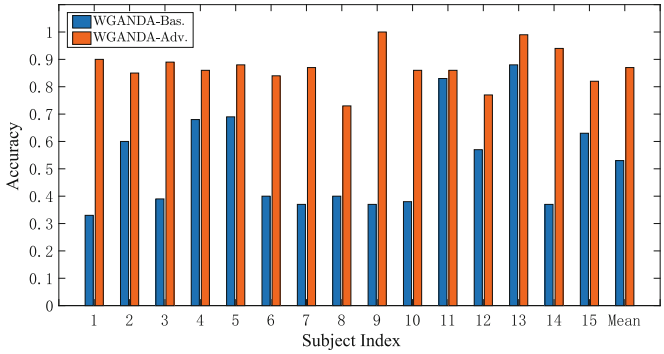

Fig. 3. Accuracy comparison between the strategy using adversarial-training and the baseline without using adversarial-training on SEED dataset.

ness of our framework, we compare it with the state-of-the-art methods including KPCA, TCA and TPT on SEED dataset [23]. We also implement these methods and evaluate their performances on DEAP dataset. All the hyperparameters are adjusted following the strategies used in [23].

4 Experimental Results

In this section, we demonstrate the effectiveness of our proposed WGANDA framework. Figure 2 depicts the training process of the adversarial-training procedure. The discriminator loss (D-loss) converges to a small value along with the training epoch as illustrated in Fig. 2(a). As the EMD between the distributions of source and target mappings, D-Loss converging to a small value demonstrates that the two marginal distributions are approximate to each other. The MMD curve in Fig. 2(b) has a similar converged tendency with D-Loss, which also implies that adversarial-training has reduced the distance between the two mapping distributions. Moreover, the recognition accuracy shown in Fig. 2(c) increases while MMD decreases. This phenomenon confirms the domain adaptation assumption. Since the classifier is optimized according to the conditional distribution of X'_s, only when the two conditional distributions are similar, X'_t can achieve high recognition accuracy with the same classifier.

We first compare our proposed framework with its baseline. Figure 3 shows the accuracy comparison of using adversarial-training (WGANDA-Adv.) and without using adversarial-training (WGANDA-Bas.) on SEED dataset. The recognition accuracy of the baseline WGANDA-Bas. is calculated with the target mappings X'_t directly fed into the classifier after target generator initialization. WGANDA-Bas. performs poorly due to the fact that domain shift exists when neglecting inter-subject variability. Without adversarial-training, the source mappings and the target mappings share no common marginal distributions as well as conditional distributions. The classifier trained with X'_s hence can not predict the emotion states of the target subject precisely according to X'_t. By

Table 1. Performance of different domain adaptation methods

Methods	SEED		DEAP-Arousal		DEAP-Valence	
	Mean	Std.	Mean	Std.	Mean	Std.
SVM	0.5673	0.1629	0.4922	0.1571	0.5036	0.1125
KPCA	0.6128	0.1462	0.5891	0.1521	0.5658	0.0980
TCA	0.6364	0.1488	0.5193	0.1539	0.5516	0.1069
TPT	0.7631	0.1589	0.5577	0.1496	0.5564	0.1221
WGANDA-Bas.	0.5260	0.1831	0.5183	0.1406	0.5164	0.0929
WGANDA-Adv.	**0.8707**	**0.0714**	**0.6685**	**0.0552**	**0.6799**	**0.0656**

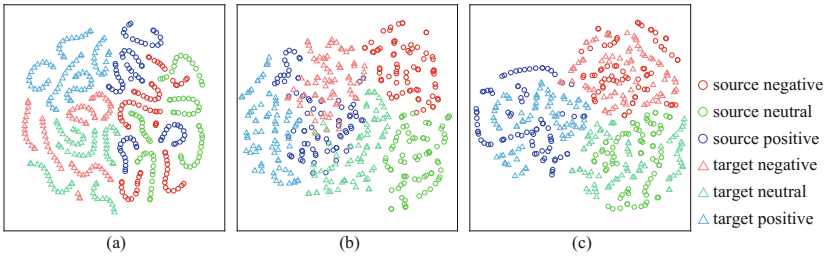

Fig. 4. Two-dimension visualization of source and target domain distributions in different training stages: (a) original distribution; (b) distribution after pre-training procedure; and (c) distribution after adversarial-training procedure. Small circles represent source data samples of three classes and small triangles represent target data samples of three classes.

using adversarial-training, the accuracy of WGANDA-Adv. shows a significant improvement for each subject compared with the baseline result.

Next, we compare our proposed framework with three state-of-the-art domain adaptation methods. Table 1 presents mean accuracies and standard deviations of our proposed framework WGANDA-Adv., the baseline WGANDA-Bas., and other three domain adaptation methods, KPCA, TCA, and TPT. The experimental results of KPCA, TCA and TPT on SEED dataset are referenced from [23]. From Table 1, we see that domain adaptation methods are effective when handling domain shift problem in EEG-based emotion recognition. Our framework significantly outperforms the state-of-the-art methods with mean accuracy of 87.07% and standard deviation of 0.0714 on SEED dataset. On DEAP dataset, our framework achieves mean accuracies of 66.85% and 67.99% and standard deviations of 0.0552 and 0.0656 on arousal and valence classifications, respectively, which is also superior to other methods.

In order to have a better view of the effectiveness of our proposed framework, the source and target data from SEED dataset at different training stages are visualized in a 2-dimension way by t-SNE [11] as shown in Fig. 4. To illustrate the influence of adversarial-training on marginal and conditional distributions more

intuitively, samples from different subjects and emotion categories are visualized with different markers. However, in both pre-training and adversarial-training procedures, target labels are unknown to the framework.

Figure 4(a) depicts the distributions of the original data from the source subjects and the target subjects, which have diverse distributions due to inter-subject variability. From Fig. 4(a), we see that there is no any overlapping between the source subjects samples (small circles) and the target subjects samples (small triangles). This means that the original data from the source subjects and target subjects have diverse distributions due to inter-subject variability. Figure 4(b) depicts the distributions of X'_s and X'_t before adversarial-training procedure, which are the mappings of the original data from source and target subjects after target generator initialization. Note that although the samples from three emotion categories have been successfully clustered, the marginal distributions are not the same. Under this circumstance, the pre-trained classifier can only recognize the three emotions of the source subject. Figure 4(c) depicts the distributions of X'_s and X'_t after the adversarial training procedure. Now the marginal distributions of the source mappings are approximate to the target mappings, while the conditional distributions are similar as well. Thus the pre-trained classifier can recognize different emotions on target subject correctly.

5 Conclusion

In this paper, we have proposed a novel Wasserstein GAN domain adaptation framework for building cross-subject EEG-based emotion recognition models. The framework adopts adversarial strategy by using Wasserstein GAN gradient penalty version. The performance of our framework has been evaluated by conducting a leave-one-subject-out cross validation on two public EEG datasets for emotion recognition. By narrowing down the gap between probability distribution of different subjects, this adversarial domain adaptation method successfully handles inter-subject variability and domain shift problems of cross-subject EEG-based emotion recognition. By taking advantages of adversarial training, the proposed framework significantly outperforms the state-of-the-art methods with a mean accuracy of 87.07% on SEED dataset, and reaches 66.85% and 67.99% on DEAP dataset for arousal and valence classifications, respectively.

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. Alarcao, S.M., Fonseca, M.J.: Emotions recognition using EEG signals: a survey. IEEE Trans. Affect. Comput. (2017)
2. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein GAN. arXiv preprint arXiv:1701.07875 (2017)

3. Duan, R.N., Zhu, J.Y., Lu, B.L.: Differential entropy feature for EEG-based emotion classification. In: IEEE EMBS NER 2013, pp. 81–84. IEEE (2013)
4. Goodfellow, I., et al.: Generative adversarial nets. In: NIPS 2014, pp. 2672–2680 (2014)
5. Gretton, A., Borgwardt, K.M., Rasch, M., Schölkopf, B., Smola, A.J.: A kernel method for the two-sample-problem. In: NIPS 2007, pp. 513–520 (2007)
6. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of Wasserstein GANs. In: NIPS 2017, pp. 5769–5779 (2017)
7. Koelstra, S., et al.: DEAP: a database for emotion analysis; using physiological signals. IEEE Trans. Affect. Comput. **3**(1), 18–31 (2012)
8. Lin, Y.P., Yang, Y.H., Jung, T.P.: Fusion of electroencephalographic dynamics and musical contents for estimating emotional responses in music listening. Front. Neurosci. **8**, 94 (2014)
9. Long, M., Cao, Y., Wang, J., Jordan, M.: Learning transferable features with deep adaptation networks. In: ICML 2015, pp. 97–105 (2015)
10. Lu, Y., Zheng, W.L., Li, B., Lu, B.L.: Combining eye movements and EEG to enhance emotion recognition. In: IJCAI 2015, pp. 1170–1176 (2015)
11. Maaten, L.V.D., Hinton, G.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**(Nov), 2579–2605 (2008)
12. Morioka, H., et al.: Learning a common dictionary for subject-transfer decoding with resting calibration. NeuroImage **111**, 167–178 (2015)
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. Samek, W., Meinecke, F.C., Müller, K.R.: Transferring subspaces between subjects in brain-computer interfacing. IEEE Trans. Biomed. Eng. **60**(8), 2289–2298 (2013)
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 Multimedia 2014, pp. 357–366. ACM (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. Sugiyama, M., Krauledat, M., MÃžller, K.R.: Covariate shift adaptation by importance weighted cross validation. J. Mach. Learn. Res. **8**, 985–1005 (2007)
19. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: CVPR 2017, vol. 1, p. 4 (2017)
20. Villani, C.: Optimal Transport: Old and New, vol. 338. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-71050-9
21. Wang, X.W., Nie, D., Lu, B.L.: Emotional state classification from EEG data using machine learning approach. Neurocomputing **129**(4), 94–106 (2014)
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. Ment. Dev. **7**(3), 162–175 (2015)
23. Zheng, W.L., Lu, B.L.: Personalizing EEG-based affective models with transfer learning. In: IJCAI 2016, pp. 2732–2738. AAAI Press (2016)