

# Multi-subject subspace alignment for non-stationary EEG-based emotion recognition

Xin Chai<sup>a</sup>, Qisong Wang<sup>a,\*</sup>, Yongping Zhao<sup>a</sup>, Xin Liu<sup>b</sup>, Dan Liu<sup>a</sup> and Ou Bai<sup>c</sup>

<sup>a</sup>*School of Electrical Engineering and Automaton, Harbin Institute of Technology, Harbin, Heilongjiang, China*

<sup>b</sup>*School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin, Heilongjiang, China*

<sup>c</sup>*Department of Electrical and Computer Engineering, Florida International University, Miami, FL, USA*

**Abstract.** Emotion recognition based on EEG signals is a critical component in Human-Machine collaborative environments and psychiatric health diagnoses. However, EEG patterns have been found to vary across subjects due to user fatigue, different electrode placements, and varying impedances, etc. This problem renders the performance of EEG-based emotion recognition highly specific to subjects, requiring time-consuming individual calibration sessions to adapt an emotion recognition system to new subjects. Recently, domain adaptation (DA) strategies have achieved a great deal success in dealing with inter-subject adaptation. However, most of them can only adapt one subject to another subject, which limits their applicability in real-world scenarios. To alleviate this issue, a novel unsupervised DA strategy called Multi-Subject Subspace Alignment (MSSA) is proposed in this paper, which takes advantage of subspace alignment solution and multi-subject information in a unified framework to build personalized models without user-specific labeled data. Experiments on a public EEG dataset known as SEED verify the effectiveness and superiority of MSSA over other state of the art methods for dealing with multi-subject scenarios.

Keywords: EEG, emotion recognition, domain adaptation, multi-subject learning, logistic regression

## 1. Introduction

Emotion recognition is a critical component in the design of man-machine interaction and clinical applications for the mental disorder. So far, research on emotion recognition has mainly focused on facial expression, speech, and certain physiologic signals [1,2]; of these, methods based on EEG signals have attracted great interest due to their objectivity and sensitivity to emotional reactivity.

Various types of feature extraction and machine learning methods have been applied with great success to the EEG-based emotion recognition [3,4]. However, distinguishing EEG-based emotional characteristics across subjects has remained being difficult due to the individual difference and high non-stationarity

---

\*Corresponding author: Qisong Wang, School of Electrical Engineering and Automaton, Harbin Institute of Technology, Harbin, Heilongjiang, China. E-mail: 11b901011@hit.edu.cn.

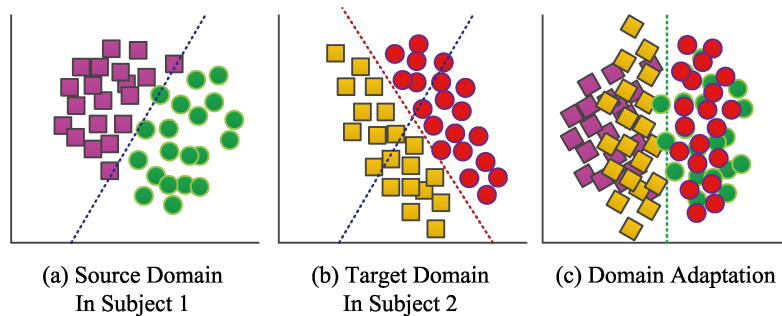


Fig. 1. An example to demonstrate domain adaptation algorithm. (a) Shows data distribution from Subject 1. (b) Shows data distribution from Subject 2. (c) Shows the transformed data after domain adaptation. It can be seen that the intention of domain adaptation is reducing the discrepancy in distribution and making classifiers is robust to both Subjects 1 and 2.

of EEG signals [5–7]. In other words, the performance of well-trained classifiers for Subject 1 usually degraded on Subject 2 when using the experimental setup shown in Fig. 1a and b. In recent years, domain adaptation (DA) algorithms, which have strong ability to minimize the domain discrepancy, have been widely studied for cross-subject classification problems [8].

In general, the domain with labeled data, which is used for training, is called the source domain, and the domain where the target task is conducted, with a different distribution, is called as the target domain. As shown in Fig. 1c, assuming the data distribution between the source and target domain is different, domain adaptation mainly focuses on learning common features across the source and target domains and training new classifiers, such as support vector machine (SVM) or logistic regression (LR), which are appropriate for both of two domains [9–12].

However, the domain adaptation strategies mentioned above can in fact only adapt one person to another person. In reality, there are many more individuals in the EEG dataset; therefore, concentrating all the subjects together to compose the source domain can make the distribution of training data inconsistent. If only one subject in the EEG dataset is used for training, the prediction for the new subject can be considered a single task. Obviously, this single task will waste the extra information provided by other subjects in the available dataset.

To handle these issues, a new DA strategy named as multi-subject subspace alignment (MSSA) is proposed in this work. The proposed method considers that there are  $N$  participants and associated labeled training samples in the training set; each participant can compose a single prediction task with the new participant who needs to be tested. Instead of learning each task independently or concentrating all the subjects together directly, all the information of these  $N$  subjects is utilized here to help in training a classifier for the new target subject, as shown in Fig. 2. Owing to an increased sample size for the prediction task, the prediction performance may be improved.

In this study, differential entropy (DE) [13] feature which has been proven effective in EEG-based emotion recognition was used to represent the EEG pattern by extracting frequency domain features from each channel band power, which are used to generate the input of the MSSA model. In MSSA, a domain adaptation method known as subspace alignment [9] was utilized to transform the source subject into the target subspace, and move the source and target subspace features closer together. Then, a set of  $N$  classifiers are learnt separately, parameterized by the vectors  $w_1, w_2, \dots, w_N$ . Meanwhile, a distance metric strategy is utilized to measure the distribution discrepancy between each pair of two subjects. In the next phase, the final classifier is obtained by applying regularized logistic regression (LR), and this means that the model of the target classifier is closer to the mean of  $M$  independent classifiers, which have a smaller distance from the target subject. As a result, classification accuracy will be improved.

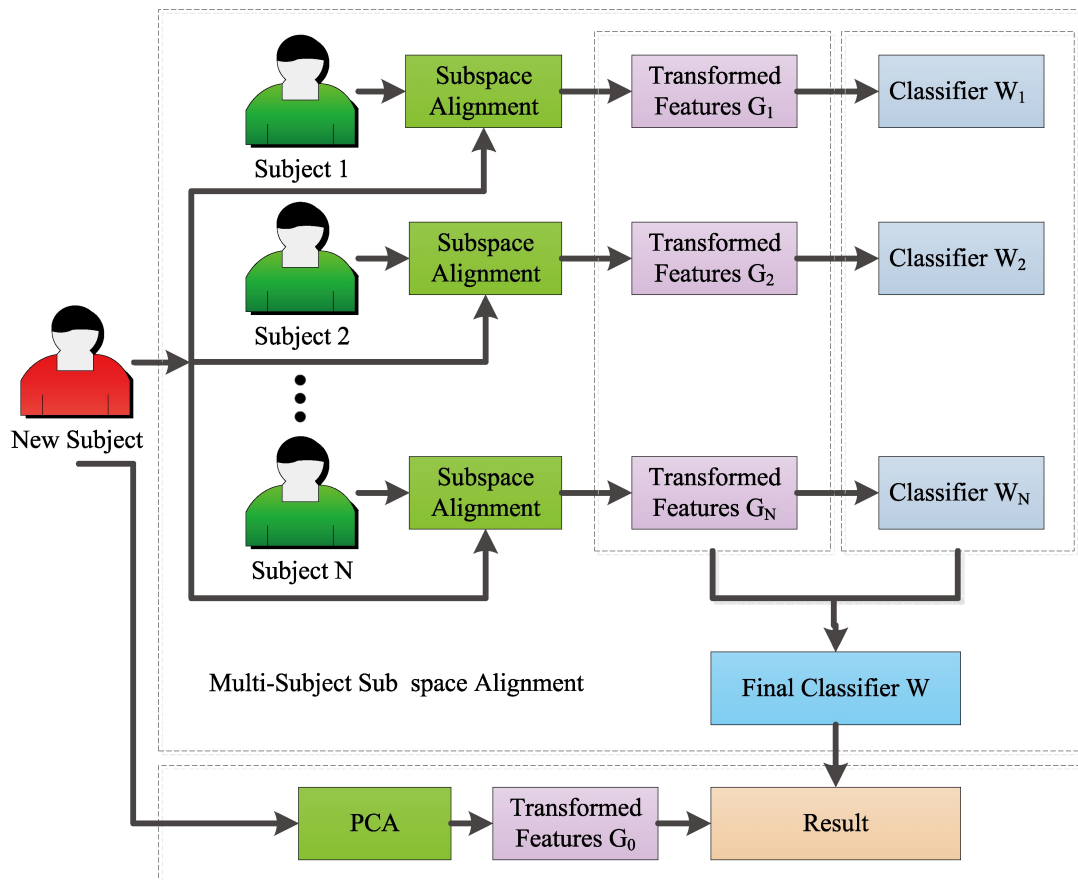


Fig. 2. Overview of multi-subject subspace alignment.

Our main contributions are two fold:

1. Subspace alignment is utilized to directly align the source and target subspaces, and to improve the consistency of training and test EEG data.
2. The proposed method takes advantage of all the information of the  $N$  subjects, and makes the final classifier more accurate by estimating the discrepancy in distribution between two subjects.

## 2. The proposed algorithm

This section introduces a multi-subject subspace alignment (MSSA) algorithm, which can be used to build personalized models without user-specific labeled data, and take advantage of the information of all the subjects in the training set. Given a set of EEG trails from a new subject, the problem to be addressed is to estimate the corresponding emotion state for each sample by using the above  $N$  subjects in the training set.

As shown in Fig. 2, MSSA is accomplished by two steps. In the first step, the normalized DE feature vector extracted from the training subject is aligned to the test subject using a subspace alignment strategy. If all of the  $N$  subjects in the training set are aligned to the new target subject using this

subspace alignment strategy, concentrating together all the new samples in the training set becomes feasible, because the subject discrepancy has been reduced. Even so, for the new target subject, the levels of similarity with these various subjects are different. Thus, in the second step, LR is applied to train  $N$  independent classifiers. Then, the distance metric is utilized to measure the discrepancy between distributions. Finally, the classifier is obtained by applying regularized logistic regression. As a result, the parameters of the target classifier are closer to the subject which is more similar to the target subject, and as a result, the classification accuracy will be improved.

### 2.1. Feature extraction and normalization

In this paper, differential entropy (DE) feature which has been proven effective in EEG-based emotion recognition [13], is utilized as the input of the proposed MSSA model. The DE feature proposes a hypothesis that segments of the EEG signal satisfy Gauss distribution after a band-pass filtering in the five frequency bands:  $\delta$  (1~3 Hz),  $\theta$  (4~7 Hz),  $\alpha$  (8~13 Hz),  $\beta$  (14~30 Hz) and  $\gamma$  (31~50 Hz). It is well known that: for information entropy of a continuous distribution, the Gaussian distribution contains the largest amount of information; therefore, it is more accurate to cover all the situations. In addition, since the energy of low-frequency is often much higher than the energy of high-frequency in EEG, DE feature can balance the ability of discriminating EEG pattern between different frequency bands. DE is defined as:

$$h(X) = - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma^2} \exp \frac{(x - \mu)^2}{2\sigma^2} \log \frac{1}{\sqrt{2\pi}\sigma^2} \exp \frac{(x - \mu)^2}{2\sigma^2} dx = \frac{1}{2} \log 2\pi e\sigma^2 \quad (1)$$

where the segment of the EEG signal  $X$  satisfies Gaussian distribution  $N(\mu, \sigma^2)$ . After a band-pass filtering, the corresponding DE feature can be calculated following Eq. (1).

Then, in order to normalize the DE features, the Min-Max strategy is utilized in this paper. The DE features from the source and target sets are denoted as  $F_S$  and  $F_T$ . Denoted as  $F = [F_S \ F_T]$ , the target is to normalize  $F$  to  $[L, U]$ . The equation is then given by:

$$X = \frac{F - \min(F)}{\max(F) - \min(F)} * (U - L) + L \quad (2)$$

In this paper,  $U$  and  $L$  are set as 1 and 0, and the normalized DE feature vector  $X$  was used as the input to our MSSA model.

### 2.2. Subspace alignment and cluster

Assume the training set contains  $N$  subjects, and that each subject contains  $ns$  labeled samples in  $d$  dimensions, denoted as  $X_S = [x_{s1}, x_{s2}, \dots, x_{sns}] \in \mathbb{R}^{d \times ns}$  with their class labels  $Y_S = [y_{s1}, y_{s2}, \dots, y_{sns}]$ , where  $x_{si}$  is the normalized DE feature vectors extracted from the  $i$ -th sample of the training subject  $M$ . Similarly, the normalized DE feature vector extracted from the new target set is denoted as  $X_T = [x_{T1}, x_{T2}, \dots, x_{Tnt}] \in \mathbb{R}^{d \times nt}$ . In order to handle the domain discrepancy between  $P(X_S)$  and  $P(X_T)$ , a linear transformation function [9] is learnt to align the subspace coordinate system of  $X_S$  to  $X_T$ .

According to the principal component analysis (PCA) theory,  $l$  eigenvectors corresponding to the  $l$  largest eigenvalues were defined as source subspace  $Z_S$  and target subspace  $Z_T$ . Thus, assuming that a linear relationship exists between  $Z_S$  and  $Z_T$ , the aim of this study is to learn the transformation  $S$  which can minimize the difference between  $Z_S S$  and  $Z_T$ :

$$\min(\|Z_S S - Z_T\|_F^2) \quad (3)$$

where  $\|\cdot\|_F^2$  is the Frobenius norm. Due to orthonormal operations of the Frobenius norm is invariable, Eq. (3) can be re-written as:

$$F(S) = \|Z_S^T Z_S S - Z_S^T Z_T\|_F^2 = \|S - Z_S^T Z_T\|_F^2 \tag{4}$$

Obviously, when  $S^* = Z_S^T Z_T$ ,  $F(S)$  is a minimum. The transformed subspace of  $X_S$  can be expressed as:

$$Z_{Trans} = Z_S Z_S^T Z_T \tag{5}$$

Then, samples in the subject  $iX_{iS}$  and the target subject  $X_T$  can be transformed into the new feature space by computing  $G_i = X_S Z_{Trans} = X_S Z_S Z_S^T Z_T$  and  $H = X_T Z_T$ , and the marginal distribution discrepancy between the source subject  $G_i$  and target subject  $H$  can be reduced. The transformed samples from the  $N$  subjects in the training set are then used to learn  $N$  independent classifiers  $w_i$  by minimizing the logistic loss function:

$$\min \sum_{j=1}^{ns} \log(1 + \exp(-y_j w^T g_j)) \tag{6}$$

The gradient-descent (GD) optimization approach is employed to solve this learning problem.

### 2.3. Multi-subject fusion with logistic loss

In order to train a target classifier whose parameters are closer to the subjects that are more similar to the target subject, a new DA strategy named as multi-subject subspace alignment is proposed in this paper. Firstly, a simple strategy is utilized to measure the distance between subject source subject  $G_i$  and the target subject  $G_0$ . To estimate the distance between  $G_i$  and  $G_0$ , first LR is applied to the target subject  $G_0$  to estimate the pseudo-labels. In this way, the cluster centroids of  $G_i$  and  $G_0$  can be computed as  $U_i = \{c_1^i, c_2^i, \dots, c_q^i\}$  and  $V = \{c_1^0, c_2^0, \dots, c_q^0\}$  respectively. The distance between subjects is computed as follows:

$$D_i(G_i, G_0) = \sum_{q=1}^Q \|c_q^i - c_q^0\| \tag{7}$$

Since the distance  $D_i$  between the subject source subject  $G_i$  and target subject  $G_0$  has been obtained,  $M$  minimum distances can be selected. The model of the target classifier is expected close to the mean of those  $M$  independent classifiers. Let  $T = [G_1, G_2, \dots, G_N]$  denote the training set; the hypothetical model is then given by:

$$\min \ell(w) = \sum_{j=1}^{ns} \log(1 + \exp(-y_j w^T t_j)) + \lambda \left\| w - \frac{1}{M} \sum_i^M w_i \right\|^2 \tag{8}$$

where  $t_j$  is the  $j$ -th sample from the training set  $T$ , and  $y_j$  is its corresponding class label. Then, the second order partial derivative of  $\ell(w)$  can be expressed as:

$$\frac{\partial^2 \ell(w)}{\partial w \partial w} = \sum_j^{ns} y_j^2 t_j^2 \frac{e^{-y_j w^T t_j}}{(1 + e^{-y_j w^T t_j})^2} + 2\lambda \geq 0 \tag{9}$$

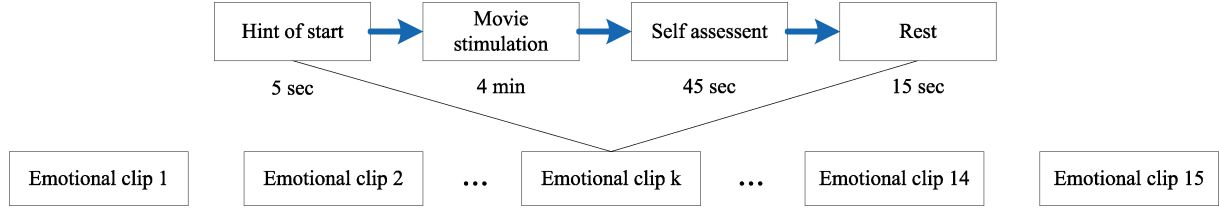


Fig. 3. Experimental protocol for emotion recognition based on EEG signal for one subject.

This proves that  $\ell(w)$  is convex. Therefore, the gradient-descent (GD) optimization approach is employed to solve this learning problem. We initialize  $w$  and learning rate  $\alpha$  randomly as follows:

$$w = w - \alpha \frac{\partial}{\partial w} (\ell(w)) = w - \alpha \sum_{j=1}^{ns} \left( \frac{1}{1 + \exp(-y_j w^T t_j)} - 1 + 2\lambda \left( w - \frac{1}{M} \sum_i^M w_i \right) \right) \quad (10)$$

Thus our regularization method finds a trade-off between the conventional model and closeness to the average of the  $M$  independent classifiers.

### 3. Experiment

#### 3.1. Datasets and experimental setup

In this paper, the performance of the proposed MSSA was evaluated using the emotion EEG dataset known as – SEED [14] (<http://bcmi.sjtu.edu.cn/~seed/>), since compared with other dataset, SEED is the only one designed for emotion recognition based on non-stationary EEG signals. In SEED, ESI NeuroScan system is utilized to record EEG signals from 15 participants with 62 electrodes. 15 Film clips in Chinese (positive, neutral and negative emotions) were utilized to stimulate each subject. For the feedback, participants were asked to assess their emotional reactions to each film clip immediately after watching each film clip. If the participant confirmed corresponding emotions had been elicited, the corresponding segments of the EEG signal were incorporated into the SEED dataset.

As shown in Fig. 3, each participant in SEED need watch 15 emotional film clips, and the duration of each film clip is about 4 minutes. In addition, there was a five-second warning before each clip and each participant had 45 seconds to assess their emotional reactions for the feedback. The SEED dataset contains a down-sampled (200 Hz), preprocessed (band-pass filter from 0.3 Hz to 50 Hz) and segmented (1 s without overlapping) version of the EEG data in Matlab (.mat file), and there are 3300 signal segments in each channel for per experiment. As there was 62 channels in total, the total dimension of features extracted from a group EEG signal segments is 310 by using the standard five frequency bands.

#### 3.2. Experiment results

In this study, SVM [15] and LR [16] without DA strategy were used in training the baseline. In addition, the proposed approach (MSSA) was compared with three state-of-the-art DA methods: auto-encoder (AE) [10], transfer component analysis (TCA) [12] and transfer joint matching (TJM) [11]. All results were conducted using the leave-one-subject-out cross validation method. For AE, TCA, and TJM, since it is impracticable to concentrate all the signal segments in the SEED dataset as the training data on account of the limits on the memory and time consuming. To avoid bias, 20 samples were

Table 1  
Comparison results of each subject on the SEED dataset (Accuracy in %)

Subject No.	SVM	LR	AE	TCA	TJM	MSSA
1	49.78	48.15	58.04	71.81	73.55	75.29
2	48.67	48.99	46.53	75.47	73.34	76.44
3	56.24	55.86	63.20	77.03	74.49	82.59
4	71.69	70.20	85.70	91.08	86.42	94.59
5	56.56	53.10	52.63	58.62	62.57	67.71
6	61.30	59.50	57.54	73.13	74.69	75.93
7	55.35	62.42	65.01	83.13	80.29	81.01
8	48.24	49.55	62.66	64.67	74.75	81.68
9	53.06	51.76	57.02	84.09	82.47	85.47
10	41.97	41.88	51.15	64.02	65.32	67.21
11	66.81	64.57	66.59	83.27	83.54	83.27
12	65.64	66.75	67.75	73.92	75.73	75.60
13	67.11	69.28	70.82	79.30	78.24	78.84
14	65.67	67.58	62.57	84.25	85.23	88.10
15	57.37	60.04	54.76	67.11	71.29	80.45
Average	57.70	57.98	61.46	75.39	76.13	79.61

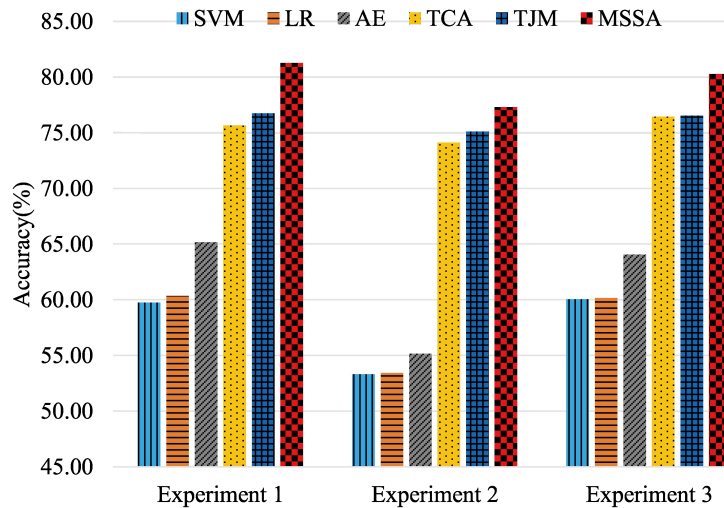


Fig. 4. The mean results of all of the subjects from same session on the SEED dataset.

selected randomly from each trial and 300 samples were obtained in total. Ultimately, 4200 samples were obtained from 14 subject as training data. Moreover, sample selection procedure was repeated five times to avoid bias in data.

Since each person undergoes three sessions in the SEED dataset, we verify our MSSA method using these three experiments. Table 1 shows the average classification accuracies of the three experiments for each subject. For more accurate illustration, the mean accuracies of the 15 subjects for each experimental session was also provided in Fig. 4.

Without the domain adaptation strategy, the two standard classifiers SVM and LR serving as the baseline for comparison, only achieved an average classification accuracy of 57.70% and 57.98%, respectively. With the benefits of a deep structure, which can learn domain-invariant feature from both training and test domains, the AE method achieves a mean accuracy of 61.46%, which is slightly better than the baseline methods. TCA and TJM, which are the most commonly used DA methods, indicate a signif-

Table 2  
Evaluation of the statistical significance between the performance of MSSA and other methods

MSSA Vs	SVM	LR	AE	TCA	TJM
<i>p</i> -Value	$1.91 \times 10^{-8}$	$2.08 \times 10^{-8}$	$7.95 \times 10^{-8}$	$7.38 \times 10^{-3}$	$8.93 \times 10^{-4}$

icant improvement with 75.39% and 76.13%. Both of these methods try to minimize the distribution divergence using an MMD constraint. However, without taking the differences among individuals into account, TCA and TJM might not be the best choices for EEG-based emotion recognition. By contrast, our MSSA achieves a mean accuracy of 79.61%. The above result suggests that, this strategy which aims at reducing the multi-subject discrepancy is effective in emotion recognition based on EEG signals. Furthermore, as shown in Table 1, the proposed MSSA method indicated improvements on most subjects, which confirms that MSSA can perform stably. Figure 4 shows the mean results of all of the subjects from same session on the SEED dataset. As shown, our MSSA method also indicated the highest accuracy. Moreover, compared with the transfer learning strategy reported in literature [17], also evaluated on SEED, the mean of 79.61% achieved by our MSSA method is higher than the 76.31% found in the literature [17]. The statistical significance between the proposed MSSA approach and other algorithms is evaluated using Student's t-test in Table 2. It could be seen that MSSA is significantly better than the methods with very low *p*-values.

#### 4. Conclusions

In this work, the MSSA strategy is proposed to make better use of multi-subject information in the EEG dataset. In MSSA, a subspace alignment strategy is utilized to reduce the marginal distribution discrepancy between each source subject and target subject. Following this, the distance metric is utilized to measure the discrepancy between the transformed source and target data, and a number of subjects which are more similar to the target subject can be selected. As a result, the parameters of the target classifier can be learnt as being closer to the subject which is more similar to the target subject, and the classification accuracy will be improved. Experiments on SEED dataset verify the effectiveness and superiority of MSSA over other state of the art methods for dealing with multi-subject scenarios.

#### Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 61401117, Grant No. 61402129 and Grant No. 61301012). Furthermore, Ou Bai was sponsored by the National Science Foundation (USA) (CNS-1552163).

#### Conflict of interest

None to report.

#### References

- [1] Agraftoti F, Hatzinakos D, Anderson AK. ECG pattern analysis for emotion detection. *IEEE Transactions on Affective Computing* 2012; 3: 102-115.



- [2] Gruebler A, Suzuki K. Design of a wearable device for reading positive expressions from facial EMG signals. *IEEE Transactions on Affective Computing* 2014; 5: 227-237.
- [3] Mühl C, Jeunet C, Lotte F. EEG-based workload estimation across affective contexts. *Frontiers in Neuroscience* 2014; 8: 114.
- [4] Murugappan M, Ramachandran N, Sazali Y. Classification of human emotion from EEG using discrete wavelet transform. *Engineering* 2010; 3: 390-396.
- [5] Buttfeld A, Ferrez PW, Millan JR. Towards a robust BCI: Error potentials and online learning. *IEEE Trans Neural Syst Rehabil Eng* 2006; 14: 164-168.
- [6] Li Y, Kambara H, Koike Y, Sugiyama M. Application of covariate shift adaptation techniques in brain-computer interfaces. *IEEE Transactions on Bio-Medical Engineering* 2010; 57: 1318.
- [7] Zheng WL, Zhu JY, Lu BL. Identifying stable patterns over time for emotion recognition from EEG. *IEEE Transactions on Affective Computing* 2016; 1-1.
- [8] Pan SJ, Yang Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* 2010; 22: 1345-1359.
- [9] Fernando B, Habrard A, Sebban M, Tuytelaars T. Unsupervised visual domain adaptation using subspace alignment. *IEEE International Conference on Computer Vision* 2014; 2960-2967.
- [10] Glorot X, Bordes A, Bengio Y. Domain adaptation for large-scale sentiment classification: A deep learning approach. *International Conference on International Conference on Machine Learning* 2011; 513-520.
- [11] Long M, Wang J, Ding G, Sun J, Yu PS. Transfer joint matching for unsupervised domain adaptation. *IEEE Conference on Computer Vision and Pattern Recognition* 2014; 1410-1417.
- [12] Pan SJ, Tsang IW, Kwok JT, Yang Q. Domain adaptation via transfer component analysis. *IEEE Transactions on Neural Networks* 2011; 22: 199.
- [13] Duan RN, Zhu JY, Lu BL. Differential entropy feature for EEG-based emotion classification. *International IEEE/EMBS Conference on Neural Engineering* 2013; 81-84.
- [14] Zheng WL, Lu BL. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development* 2015; 7: 162-175.
- [15] Chang CC, Lin CJ. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2012; 2: 1-27.
- [16] Fan RE, Chang KW, Hsieh CJ, Wang XR, Lin CJ. LIBLINEAR: A Library for Large Linear Classification. *Journal of Machine Learning Research* 2008; 9: 1871-1874.
- [17] Zheng WL, Lu BL. Personalizing EEG-based affective models with transfer learning. *International Joint Conference on Artificial Intelligence* 2016; 2732-2738.