Transfer Components between Subjects for EEG-based Emotion Recognition

Wei-Long Zheng, Yong-Qi Zhang, Jia-Yi Zhu, Bao-Liang Lu*
Center for Brain-Like Computing and Machine Intelligence
Department of Computer Science and Engineering
Key Lab. of Shanghai Education Commission for Intelligent Interaction and Cognitive Eng.
Shanghai Jiao Tong University
800 Dong Chuan Road, Shanghai 200240, China
Email: {weilong, zyq40137, happycathy, bllu}@sjtu.edu.cn

Abstract—**Addressing the structural and functional variability between subjects for robust affective brain-computer interface (aBCI) is challenging but of great importance, since the calibration phase for aBCI is time-consuming. In this paper, we propose a subject transfer framework for electroencephalogram (EEG)-based emotion recognition via component analysis. We compare two state-of-the-art subspace projecting approaches called transfer component analysis (TCA) and kernel principle component analysis (KPCA) for subject transfer. The main idea is to learn a set of transfer components underlying source domain (source subjects) and target domain (target subject). When projected to this subspace, the difference of feature distributions of both domains can be reduced. From the experiments, we show that the two proposed approaches, TCA and KPCA, can achieve an improvement on performance with the best mean accuracies of 71.80% and 79.83%, respectively, in comparison of the baseline of 58.95%. The significant improvement shows the feasibility and efficiency of our approaches for subject transfer emotion recognition from EEG signals.**

Keywords—**emotion recognition; transfer learning; EEG**

I. INTRODUCTION

The goal of affective computing (AC) is to narrow the gap between the highly emotional human and the emotionally challenged computer with the sensing of emotional states and the modeling of the emotional processing [1]. Affective brain-computer interfaces [2], [3] introduce affective factors into traditional BCI and attempt to detect and recognize emotions from neurophysiological signals using various methods in AC. There are two main traditional BCI applications, one is motor imagery based BCI for patients with spinal cord injury and stroke as a rehabilitation tool [4]. The other is ERP-based spelling systems for disabled people as a communicative tool [5]. In these systems, the rehabilitation training is necessary but annoying, especially for patients. The negative effect of fatigue or discomfort when training would degrade the efficiency of the rehabilitation. It is necessary to fuse BCI systems with cognitive monitoring, quantifying the fatigue or discomfort levels, providing valuable information about

subjects' emotional states to the BCI systems [6]. With the quick development of wearable EEG devices and dry electrode techniques, EEG-based emotion recognition has attracted increasing interest and various studies have shown its feasibility and effectiveness [7], [8], [9], [10].

To date, most machine learning based BCI systems rely on a calibration phase to train the models [11]. This calibration is time-consuming and annoying, especially for real world applications. The intuitive and straightforward approach is to train the classifiers on the collected data from a group of subjects and then make inference on the unseen data from a new subject. However, it is technically difficult due to the structural and functional variability between subjects as well as the nonstationary nature of EEG signals and the inherent changes of environmental variables [12], [13]. Traditional machine learning methods need a prior assumption that the distributions of training data and test data are independently and identically distributed (i.i.d.). However, due to this variability from subject to subject, this assumption can not be always satisfied.

Domain adaptation, one of the branches of transfer learning, is feasible to address this problem [14], [15]. Here, we apply the domain adaptation method to EEG-based emotion recognition across subjects. Let $X \in \mathcal{X}$ be the EEG recording of a sample (X, y), here $y \in \mathcal{Y}$ represents the corresponding emotion labels. In this case, $\mathcal{X} = \mathbb{R}^{C \times d}$, C is the number of channels, and d is the number of time series samples. Let $P(X)$ be the marginal probability distribution of X. According to [14], $\mathcal{D} = \{\mathcal{X}, P(X)\}$ is a domain, which in our case is a given subject from which we record the EEG signals. The source and target domain in this paper share the same feature space, $\mathcal{X}_S = \mathcal{X}_T$, but the respective marginal probability distributions are different, $P(X_S) \neq P(X_T)$. The key assumption in most domain adaptation methods is that $P(Y_S|X_S) = P(Y_T|X_T)$.

The major problem for subject transfer is how to reduce the difference between the distributions of the source and target domain data. It is crucial to extract good feature representations across subjects for EEG-based emotion recognition [16]. Several researchers proposed various approaches to tackle the subject to subject variability of EEG. For example, Matthias

et al. [11] proposed a zero-training method, which can exploit data from previous sessions in order to learn most of the calibration parameters. Wojciech *et al.* [12] proposed an approach to transfering information about nonstationarities in the data based on the assumption that principle nonstationarities are similar between subjects. Their results indicate that their method achieved an increase in performance and the extracted patterns allow for meaningful interpretation. Morioka *et al.* [13] proposed a method for extracting spatial bases shared by multiple subjects with dictionary learning. Their approach used the resting-state activity of target subjects as calibration data for compensating the variability. However, most of the existing methods are based on the common spacial patterns (CSP) methods for the motor imagery task. To our best knowledge, there are limited reports about the subject transfer for EEG-based emotion recognition.

In this paper, we apply two feature reduction methods, called transfer component analysis [17] and kernel principle analysis [18], to subject transfer emotion recognition from EEG. Although the distributions of source domain and target domain in high dimensional space are different, we may find a low dimensional manifold space where the distributions of both domains are similar [17]. These two methods try to learn a set of common transfer components underlying both the source domain and the target domain. When projected to this subspace, the difference of feature distributions of both domains can be reduced. There exists a transformation ϕ such that $P(\phi(X_S)) \approx P(\phi(X_T))$ and $P(Y_S|\phi(X_S)) \approx P(Y_T|\phi(X_T))$. We compare these two subject transfer methods with the baseline approach (training on the source domain and the testing directly on target domain). We show that TCA and KPCA can achieve an improvement on performance with the best mean accuracies of 71.80% and 79.83%, respectively, in comparison of the baseline of 58.95%.

II. METHODS

A. Feature Extraction

The raw EEG data is firstly downsampled to 200Hz sampling rate in order to reduce computing complexity. In order to filter the noise and remove the artifacts, the EEG data is processed with a bandpass filter between 0.3Hz and 50Hz. Each channel of EEG data is divided into the same-length segments of 1s without overlapping. After preprocessing the raw EEG data, feature extraction and smooth are further processed on each segment of the EEG data.

According to our previous work [19], differential entropy (DE) can achieve better performance than conventional EEG features such as power spectral density. Since DE features have the balance ability of discriminating the EEG patterns between low and high frequency energy, we employ the DE features in this study. If a random variable obeys the Gaussian distribution $N(\mu, \sigma^2)$, the differential entropy can be calculated as follows,

$$
\begin{aligned}
h(X) = & -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{(x-\mu)^2}{2\sigma^2} \log \frac{1}{\sqrt{2\pi\sigma^2}} \\
& \exp \frac{(x-\mu)^2}{2\sigma^2} dx = \frac{1}{2} \log 2\pi e\sigma^2.
\end{aligned}
\tag{1}
$$

We extract the DE features in five frequency bands (delta: 1-3Hz, theta: 4-7Hz, alpha: 8-13Hz, beta: 14-30Hz, gamma: 31-50Hz) with a 256-point short time fourier transform. The total dimension of DE features of a sample of 62 channels is 310. It should be noted that there is often contamination of electromyography (EMG) signals from facial expressions in EEG data [3], [20], [21]. Although the EEG data is preprocessed with a bandpass filter, it is also possible that the mixture of EMG and EEG signals affects the effectiveness of emotion recognition in later analysis.

B. TCA-based Subject Transfer

Here, we want to find a transformation $\phi(\cdot)$ such that $P(\phi(X_S)) \approx P(\phi(X_T))$ and $P(Y_S|\phi(X_S)) \approx P(Y_T|\phi(X_T))$. Since we have no labeled data in target domain (target subject), $\phi(\cdot)$ cannot be learned through minimizing the distance between $P(Y_S|\phi(X_S))$ and $P(Y_T|\phi(X_T))$. Pan *et al.* [17] proposed an efficient approach called transfer component analysis to learn $\phi(\cdot)$. An intuitive approach to find the mapping $\phi(\cdot)$ is to minimize the Maximum Mean Discrepancy (MMD) [22] between the empirical means of the two domains,

$$
MMD(X_S', X_T') = ||\frac{1}{n1} \sum_{i=1}^{n1} \phi(x_{S_i}) - \frac{1}{n2} \sum_{i=1}^{n2} \phi(x_{T_i})||_{\mathcal{H}}^2,
\tag{2}
$$

where $n1$ and $n2$ represent the sample numbers of source domain and target domain, respectively. However, it can get stuck in poor local minima.

TCA is a dimensionality reduction based domain adaptation method. It embeds both the source and target domain data into a shared low-dimensional latent space using a mapping ϕ. Specially, let the Gram matrices defined on the source domain, target domain and cross-domain data in the embedded space be $K_{S,S}$, $K_{T,T}$, $K_{S,T}$, respectively. The kernel matrix K is defined on all the data as

$$
K = \begin{bmatrix} K_{S,S} & K_{S,T} \\ K_{T,S} & K_{T,T} \end{bmatrix} \in \mathbb{R}^{(n1+n2)\times(n1+n2)}.
\tag{3}
$$

By virtue of kernel trick, the MMD distance can be rewritten as $tr(KL)$, where $K = [\phi(x_i)^\top \phi(x_j)]$, and $L_{ij} = 1/n_1^2$ if $x_i, x_j \in X_S$, else $L_{ij} = 1/n_2^2$ if $x_i, x_j \in X_T$, otherwise, $L_{ij} = -(1/n_1 n_2)$. A matrix $\widetilde{W} \in \mathbb{R}^{(n1+n2)\times m}$ transforms the empirical kernel map K to an m-dimension space (where $m \ll n1 + n2$). The resultant kernel matrix is

$$
\widetilde{K} = (KK^{-1/2}\widetilde{W}(\widetilde{W}^\top K^{-1/2}K)) = KWW^\top K,
\tag{4}
$$

where $W = K^{-1/2}\widetilde{W}$. With the definition of $\widetilde{K}$ in Eq.(4), the MMD distance between the empirical means of the two domain X_S' and X_T' can be rewritten as

$$
Dist(X_S', X_T') = tr((KWW^\top K)L) = tr(W^\top KLKW).
\tag{5}
$$

A regularization term $tr(W^\top W)$ is usually added to control the complexity of W, while minimizing Eq.(5).

Besides reducing the difference of the two distributions, ϕ should also preserve the data variance that related to the target learning task. From Eq.(4), the variance of the projected samples is $W^\top KHKW$, where $H = I_{n1+n2} - (1/(n1+n2))\mathbf{1}\mathbf{1}^\top$ is the centering matrix, $\mathbf{1} \in \mathbb{R}^{n1+n2}$ is the column vector with all 1's, and $I_{n1+n2} \in \mathbb{R}^{(n1+n2)\times(n1+n2)}$ is the identity matrix.

Therefore, the objective function of TCA is

$$\min_{W} \quad tr(W^\top KLKW) + \mu tr(W^\top W)$$
$$s.t. \quad W^\top KHKW = I_m \tag{6}$$

where $\mu > 0$ is a regularization parameter, and $I_m \in \mathbb{R}^{m \times n}$ is the identity matrix. According to [17], the solutions W are the m leading eigenvectors of $(KLK + \mu I)^{-1}KHK$, where $m \leq n1 + n2 - 1$. The algorithm of TCA for subject transfer is summarized in Algorithm 1. We recommend the readers to refer to [17] for the detailed descriptions of TCA. After obtaining the transformation matrix W, standard machine learning methods can be used in this subspace across domain.

Algorithm 1 TCA-based Subject Transfer

---

input : Source domain data set $\mathcal{D}_S = \{(x_{S_i}, y_{srci})\}_{i=1}^{n1}$, and target domain data set $\mathcal{D}_T = \{x_{T_j}\}_{j=1}^{n2}$.
output : Transformation matrix W.

1: Compute kernel matrix K from $\{x_{S_i}\}_{i=1}^{n1}$ and $\{x_{T_j}\}_{j=1}^{n2}$, matrix L, and the centering matrix H.
2: Eigendecompose the matrix $(KLK + \mu I)^{-1}KLK$ and select the m leading eigenvectors to construct the transformation matrix W.
3: **return** tranformation matrix W.

C. KPCA-based subject transfer

Kernel PCA extends standard PCA to nonlinear dimensionality reduction with kernel methods [18]. Assume that we have a nonlinear transformation $\phi(x)$ from the original D-dimensional feature space to an M-dimensional feature space, where $M \gg D$. Each sample x_i is projected to a point $\phi(x_i)$. The covariance matrix of the projected features can be calculated by $C = \frac{1}{N}\sum_{i=1}^{N}\phi(x_i)\phi(x_i)^\top$. Its eigenvalues and eigenvectors are given by $Cv_k = \lambda_k v_k$, where $k = 1, 2, \cdots, M$. Combining these two formulas, we have

$$\frac{1}{N}\sum_{i=1}^{N}\phi(x_i)\phi(x_i)^\top v_k = \lambda_k v_k, \tag{7}$$

All solutions v_k lie in the span of $\phi(x_1), \cdots, \phi(x_N)$. This implies that there exist coefficients $\alpha_1, \cdots, \alpha_N$ such that

$$v_k = \sum_{i=1}^{N}\alpha_{ki}\phi(x_i). \tag{8}$$

If we define the kernel function $\kappa(x_i, x_j) = \phi(x_i)^\top \phi(x_j)$, and combine Eq.(7) and Eq.(8), we have

$$\mathbf{K}^2 \mathbf{a}_k = \lambda_k N \mathbf{K} \mathbf{a}_k, \tag{9}$$

where $\mathbf{K}_{i,j} = \kappa(x_i, x_j)$, and $\mathbf{a}_k$ is the N-dimensional column vector of α_{ki}. α_{ki}, which can be obtained by

$$\mathbf{K}\mathbf{a}_k = \lambda_k N \mathbf{a}_k. \tag{10}$$

Thus the kernel principal components can be calculated by

$$p_k(x) = \phi(x)^\top v_k = \sum_{i=1}^{N}\alpha_{ki}\kappa(x, x_i). \tag{11}$$

The Gram matrix $\widetilde{\mathbf{K}}$ is used to substitute the kernel matrix $\mathbf{K}$. The Gram matrix is given by

$$\widetilde{\mathbf{K}} = \mathbf{K} - \mathbf{1}_N \mathbf{K} - \mathbf{K}\mathbf{1}_N + \mathbf{1}_N \mathbf{K}\mathbf{1}_N, \tag{12}$$

where $\mathbf{1}_N$ is the $N \times N$ matrix with all elements equal to $1/N$ [23]. The algorithm of KPCA for subject transfer is summarized in Algorithm 2.

Algorithm 2 KPCA-based Subject Transfer

---

input : Source domain data set $\mathcal{D}_S = \{(x_{S_i}, y_{srci})\}_{i=1}^{n1}$, and target domain data set $\mathcal{D}_T = \{x_{T_j}\}_{j=1}^{n2}$.
output : The kernel principal components p_k.

1: Concatenate the source and target domain data sets as the training data set, $\{x_i\}_{i=1}^{n1+n2} = [\{x_{S_i}\}_{i=1}^{n1}; \{x_{T_j}\}_{j=1}^{n2}]$.
2: Construct the kernel matrix $\mathbf{K}$ from the training data set $\{x_i\}_{i=1}^{n1+n2}$.
3: Compute the Gram matrix $\widetilde{\mathbf{K}}$ using Eq. (12).
4: Compute the vectors $\mathbf{a}_k$ using Eq. (10) (substitute $\mathbf{K}$ with $\widetilde{\mathbf{K}}$).
5: Compute the kernel principal components p_k using Eq. (11).
6: **Return** the kernel principal components p_k.

III. EXPERIMENT SETUP

A. EEG Dataset for Emotion Recognition

In this study, we employ a publicly available EEG dataset for emotion recognition called SJTU Emotion EEG dataset (SEED) [24] via the project website[1]. In SEED, the dataset includes the EEG data from 15 subjects (7 males and 8 females; MEAN: 23.27, STD: 2.37) when watching 15 four-minute emotional film clips and each subject performs the experiments three times with an interval of about one week. Therefore, there are totally 45 experiments in the dataset. The movie clips are carefully chosen as stimuli to help eliciting subjects' right emotions from a preliminary study. The categories of emotion for each clip are defined with the highest ratings of subjects. The dataset contains three discrete categories of emotions (positive, neutral, and negative). There are five emotional film clips for each emotion in one experiment. For each session, there is a 5s hint that indicates the beginning of next session before each clip, 45s for self-assessment, and 15s for rest after each clip. For self-assessment, the subjects are asked to rate their own emotional reactions in a questionnaire. The EEG is

---

[1]http://bcmi.sjtu.edu.cn/~seed/index.html

recorded using an ESI NeuroScan System from 62 channels according to the international 10-20 system. For more detailed descriptions of SEED, please refer to [24]. The sampling rate of the preprocessed EEG data is 200Hz. And we compute the DE features from the preprocessed EEG data with a time window of one second. There are about 3300 samples for one experiment.

B. Detailed Parameters for Training

In this section, we present the details of the parameters for training and the baseline for comparison. We aim to classify three emotions (positive, neutral and negative) from EEG signals. For subject transfer for emotion recognition, a straightforward method is to pool data from all subjects and employ the leave-one-subject-out cross validation, which has often been used for evaluating cross-subject performance in the literature. However, following this way, it is very time consuming to train the models with all the data from total subjects, whose sample numbers are up to 153 thousands. In this study, we randomly select a subset of samples from 14 subjects as the source domain and test the models with the data from the remaining subject as the target domain for all three methods. Since the sample numbers of different classes for each experiment are almost the same, the randomly selected subset of each class is balanced. Dimensionality reductions are further performed via the baseline method (dimension reduction in source domain and target domain with linear KPCA separately), TCA and KPCA. We further employ support vector machines with linear kernel and use new features extracted from different methods as input. We use Liblinear toolbox [25] to implement SVM classifiers with linear kernel. We search the parameter space $2^{[-10:10]}$ with a step of one for C to find the optimal value for training. All the algorithms are implemented in the Matlab. We evaluate how the performance of different methods varies with respect to the size of subset for training and dimensionality of new feature spaces.

For TCA, there are three parameters, kernel type with kernel parameters, regularization parameter μ, and the dimensionality of latent space D. Unless it is explicitly described, we set $\mu = 1$ and $D = 30$. We employ linear and radial basis function kernels for TCA. The rest of the parameters are kept the same as [17]. For KPCA, there are two parameters, kernel type with kernel parameters and the dimensionality of latent space D. We compare two types of kernels, linear and Gaussian kernels for KPCA. The Gaussian kernel width is set to 10 in all experiments. After learning the common transfer components from the source domain and the target domain using TCA and KPCA, the projected features of the source subjects and the target subject in the new subspace are further used to train linear SVM classifiers for cross-subject emotion recognition.

IV. EXPERIMENT RESULTS

We first compare the performance of three different methods mentioned above. Table I shows the mean classification accuracies and standard deviations of the baseline, TCA and
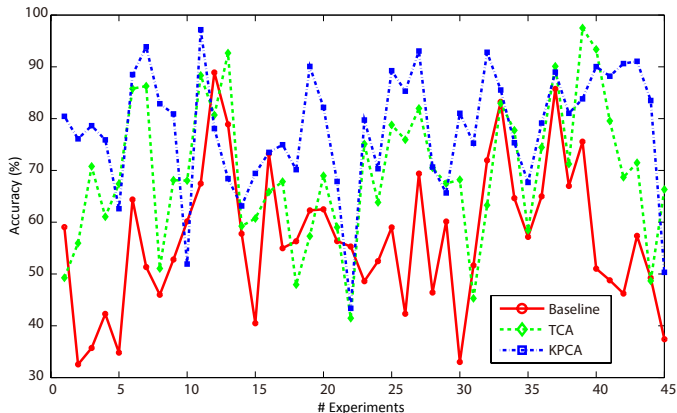

Fig. 1. The accuracies of three different methods (baseline with all features, TCA with linear kernel, KPCA with linear kernel) for total 45 experiments.

KPCA methods. For each target subject, we pool all the EEG features from the remaining 14 subjects and randomly select a subset of samples as the source domain. Here, we set the size of subset as 5000. From Table I, we can see that for the same feature dimension, TCA and KPCA outperform the baseline method without transfer. Figure 1 shows the accuracies of the three different methods for total 45 experiments. TCA and KPCA perform much better than the baseline method in most experiments. From the results, we can see that TCA and KPCA can achieve an increase in performance with respect to the baseline method. This improvement indicates that the source domain and target domain share common transfer components in the low-dimensional latent space. The new extracted features of target domain are separable in the latent space. Moreover, KPCA with linear kernel achieves the best performance among these three different methods. EEG data always contains much noise when collected, specially in different environments. KPCA can denoise when transferring the

TABLE I
CLASSIFICATION ACCURACIES (%) OF THE VARIOUS METHODS (THE NUMBER INSIDE PARENTHESES IS THE STANDARD DEVIATION)

Method		# dim	Accuracy
Baseline		310	56.82(13.79)
		10	**56.87(9.54)**
		20	54.50(10.38)
		30	53.56(10.76)
TCA	linear kernel	10	64.51(12.95)
		20	68.36(14.71)
		30	**69.44(13.48)**
	RBF kernel	10	65.50(11.44)
		20	**66.33(11.45)**
		30	65.37(12.65)
KPCA	linear kernel	10	67.74(9.24)
		20	72.54(11.43)
		30	**77.96(11.93)**
	Gaussian kernel	10	63.99(12.95)
		20	66.04(12.70)
		30	**66.54(12.64)**

common components between the source and target domains. Therefore, the classification models learned in the denoised latent space are more accurate than that of the original feature space. For kernel function, the linear kernel performs better than the RBF and Gaussian kernels here. This may imply that linear kernel is often adequate for high-dimensional EEG data.

Figure 2 shows the comparison results of the baseline, TCA and KPCA for different dimensionality of the latent space with the sample subset fixed to 5000. As we can see from the results, the accuracies of TCA and KPCA increase while the baseline decreases with the increasing dimensionality of the latent space. When the dimensions of the latent space are larger than 30, both TCA and KPCA can achieve relatively high and stable performance. TCA outperforms KPCA with very low dimensions (e.g. 3) of the latent space, while KPCA perform better than TCA with larger dimensions of the latent space. The performance of KPCA in low-dimensional latent space fluctuates more than TCA. This may be because KPCA can only project out noise in the original high-dimensional space but cannot explicitly reduce the difference between the source and target domains. However, in higher dimensions of the latent space, the role of reducing the distance between data distributions in the two domains may be weaker than that of filtering the unrelated noise.

Figure 3 shows the comparison results of TCA, KPCA and the baseline method by varying the numbers of samples randomly selected from the source domain for training. Here, we employ TCA with linear kernel and KPCA with linear kernel for comparison. We fixed the dimensionalities of TCA and KPCA as 30. From the results, we can see that the trends of the three curves are similar. All the three methods achieve the highest accuracies with 2000 samples selected from the source domain. The best mean classification accuracies and standard deviations of baseline, TCA and KPCA are 58.95%(12.23%), 71.80%(13.99%) and 79.94%(10.32%), respectively. From the

Fig. 3. Comparison of TCA, KPCA and the baseline method without dimensionality reduction by varying the number of samples selected from the source domain for training.

results, we can see that the performance of three methods are relatively stable with the varying numbers of samples randomly selected from the source domain for training. As can be seen, even with only a few samples selected from the source domain, the subject-transfer methods via common components, TCA and KPCA can perform better than the baseline method.

Figures 4 and 5 show the visualizations of the EEG features in 3-dimension latent space using TCA and KPCA, respectively. As we can see, the features from each emotion can be clustered together in the latent space. This indicates that there exist some common neural patterns for each emotion. Moreover, the cluster of positive emotion is significantly far from the other two clusters of neutral and negative emotions, while the clusters of neutral and negative emotions have much more overlap. This implies that the neural patterns of neutral and negative emotions are more similar to each other and it is easier to recognize positive emotions among these three emotions. These findings are consistent with our previous observations [26]. From these two figures, we can also observe that for each cluster of different emotions, it consists of several continuous manifold structures with shape like lines. For each manifold structure, the data points are from the same session in one experiment. This indicates that there even exist some variabilities among different recording sessions from the same experiments. For different experiments from different subjects, the EEG patterns may suffer with more variabilities. This is the challenge of robust affective brain-computer interfaces that we study in this paper.

V. CONCLUSION

In this paper, we have proposed a subject transfer framework for EEG-based emotion recognition via shared common components. We have applied two domain adaptation methods called TCA and KPCA to address the structural and functional variability across subjects. These two methods
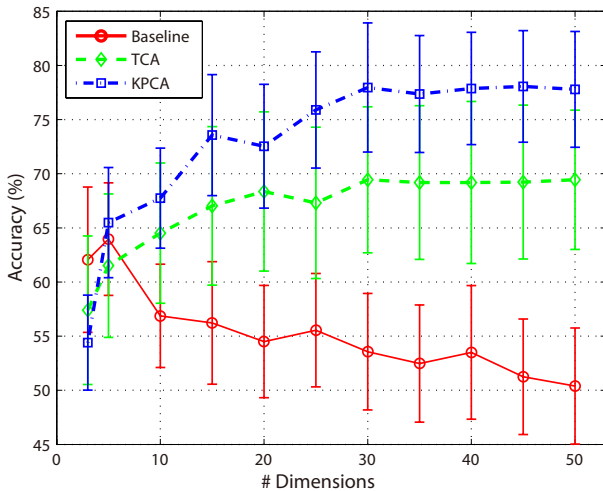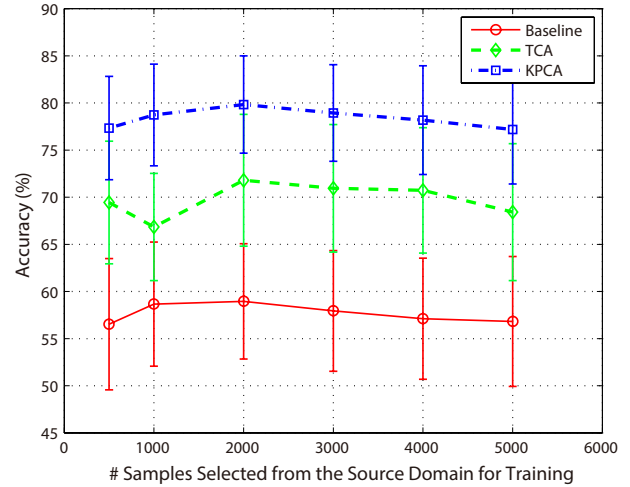

Fig. 2. Comparison of the baseline method (dimension reduction in source domain and target domain with linear KPCA separately), TCA and KPCA for different dimensionality of the latent space.
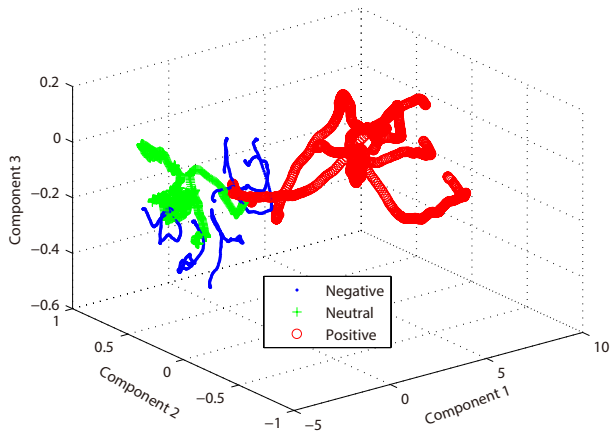
Fig. 4. Visualization of the EEG features of one experiment in 3-dimension latent space using TCA.
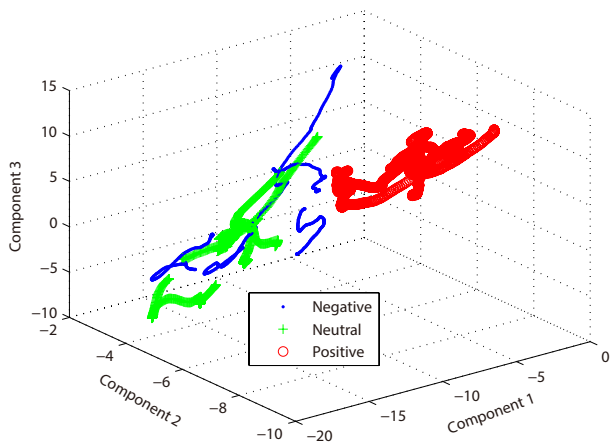

Fig. 5. Visualization of the EEG features of one experiment in 3-dimensional latent space using KPCA.

can learn transfer components in a moderate low-dimensional latent space from the source domain and target domain, where the difference of feature distributions of both domains can be reduced. The experimental results show that TCA and KPCA can achieve an improvement compared to the baseline method with mean accuracies. Moreover, we have plotted the visualizations of EEG features in three-dimensional latent space using TCA and KPCA. The EEG features of different emotions have specific manifold structures, which indicates that the neural patterns for different emotions do exist.

REFERENCES

[1] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Transactions on Affective Computing*, vol. 1, no. 1, pp. 18–37, 2010.

[2] G. G. Molina, T. Tsoneva, and A. Nijholt, "Emotional brain-computer interfaces," in *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*. IEEE, 2009, pp. 1–9.

[3] C. Mühl, B. Allison, A. Nijholt, and G. Chanel, "A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges," *Brain-Computer Interfaces*, vol. 1, no. 2, pp. 66–84, 2014.

[4] E. Buch, C. Weber, L. G. Cohen, C. Braun, M. A. Dimyan, T. Ard, J. Mellinger, A. Caria, S. Soekadar, A. Fourkas *et al.*, "Think to move: a neuromagnetic brain-computer interface (BCI) system for chronic stroke," *Stroke*, vol. 39, no. 3, pp. 910–917, 2008.

[5] C. Guger, S. Daban, E. Sellers, C. Holzner, G. Krausz, R. Carabalona, F. Gramatica, and G. Edlinger, "How many people are able to control a p300-based brain–computer interface (BCI)?" *Neuroscience letters*, vol. 462, no. 1, pp. 94–98, 2009.

[6] T. O. Zander and C. Kothe, "Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general," *Journal of neural engineering*, vol. 8, no. 2, p. 025005, 2011.

[7] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, and J.-H. Chen, "EEG-based emotion recognition in music listening," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1798–1806, 2010.

[8] D. O. Bos, "EEG-based emotion recognition," *The Influence of Visual and Auditory Stimuli*, pp. 1–17, 2006.

[9] D. Nie, X.-W. Wang, L.-C. Shi, and B.-L. Lu, "EEG-based emotion recognition during watching movies," in *5th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2011, pp. 667–670.

[10] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from EEG data using machine learning approach," *Neurocomputing*, vol. 129, pp. 94–106, 2014.

[11] M. Krauledat, M. Tangermann, B. Blankertz, and K.-R. Müller, "Towards zero training for brain-computer interfacing," *PloS one*, vol. 3, no. 8, p. e2967, 2008.

[12] W. Samek, F. C. Meinecke, and K.-R. Muller, "Transferring subspaces between subjects in brain–computer interfacing," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 8, pp. 2289–2298, 2013.

[13] H. Morioka, A. Kanemura, J.-i. Hirayama, M. Shikauchi, T. Ogawa, S. Ikeda, M. Kawanabe, and S. Ishii, "Learning a common dictionary for subject-transfer decoding with resting calibration," *NeuroImage*, vol. 111, pp. 167–178, 2015.

[14] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[15] L. Duan, D. Xu, and I. W. Tsang, "Domain adaptation from multiple sources: A domain-dependent regularization approach," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 3, pp. 504–518, 2012.

[16] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *IEEE Transactions on Affective Computing*, vol. 5, no. 3, pp. 327–339, 2014.

[17] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.

[18] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[19] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for EEG-based emotion classification," in *6th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2013, pp. 81–84.

[20] M. Fatourechi, A. Bashashati, R. K. Ward, and G. E. Birch, "EMG and EOG artifacts in brain computer interface systems: A survey," *Clinical neurophysiology*, vol. 118, no. 3, pp. 480–494, 2007.

[21] I. Goncharova, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "EMG contamination of EEG: spectral and topographical characteristics," *Clinical Neurophysiology*, vol. 114, no. 9, pp. 1580–1593, 2003.

[22] A. Gretton, K. M. Borgwardt, M. Rasch, B. Schölkopf, and A. J. Smola, "A kernel method for the two-sample-problem," in *Advances in neural information processing systems*, 2006, pp. 513–520.

[23] C. M. Bishop *et al.*, *Pattern recognition and machine learning*. springer New York, 2006, vol. 4, no. 4.

[24] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," to appear in *IEEE Transactions on Autonomous Mental Development*, 2015.

[25] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "Liblinear: A library for large linear classification," *The Journal of Machine Learning Research*, vol. 9, pp. 1871–1874, 2008.

[26] Y. Lu, W.-L. Zheng, B. Li, and B.-L. Lu, "Combining eye movements and EEG to enhance emotion recognition," to appear in *International Joint Conference on Artificial Intelligence (IJCAI)*, 2015.