

Research Article

Transfer Kernel Common Spatial Patterns for Motor Imagery Brain-Computer Interface Classification

Mengxi Dai,^{1,2} Dezhi Zheng ,^{1,2} Shucong Liu,^{1,2} and Pengju Zhang^{1,2}

¹Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, Beihang University, Beijing 100191, China

²Key Laboratory of Precision Opto-Mechatronics Technology, Ministry of Education, School of Instrumentation Science and Opto-Electronics Engineering, Beihang University, Beijing 100191, China

Correspondence should be addressed to Dezhi Zheng; zhengdezhi@buaa.edu.cn

Received 15 November 2017; Revised 2 February 2018; Accepted 12 February 2018; Published 18 March 2018

Academic Editor: Marta Parazzini

Copyright © 2018 Mengxi Dai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Motor-imagery-based brain-computer interfaces (BCIs) commonly use the common spatial pattern (CSP) as preprocessing step before classification. The CSP method is a supervised algorithm. Therefore a lot of time-consuming training data is needed to build the model. To address this issue, one promising approach is transfer learning, which generalizes a learning model can extract discriminative information from other subjects for target classification task. To this end, we propose a transfer kernel CSP (TKCSP) approach to learn a domain-invariant kernel by directly matching distributions of source subjects and target subjects. The dataset IVa of BCI Competition III is used to demonstrate the validity by our proposed methods. In the experiment, we compare the classification performance of the TKCSP against CSP, CSP for subject-to-subject transfer (CSP SJ-to-SJ), regularizing CSP (RCSP), stationary subspace CSP (ssCSP), multitask CSP (mtCSP), and the combined mtCSP and ssCSP (ss + mtCSP) method. The results indicate that the superior mean classification performance of TKCSP can achieve 81.14%, especially in case of source subjects with fewer number of training samples. Comprehensive experimental evidence on the dataset verifies the effectiveness and efficiency of the proposed TKCSP approach over several state-of-the-art methods.

1. Introduction

The brain-computer interface (BCI) offers a new pathway of communication between an external device and the brain through transforming metabolic or electrophysiological brain activities to control messages for devices and applications. The electroencephalogram (EEG) obtains time series data with multiple variants recorded at several sensors pressed on the scalp. It thereby presents electrical potentials under the induction of brain activities. These are used by noninvasive BCI systems to convert the mind or intention of a subject into a control message for certain device, such as a computer, a neuroprosthesis, or a wheelchair [1–4].

Currently, classification performance promotion of BCI systems based on the EEG has significant challenges. For one, it is necessary for a fresh subject to conduct a lengthy calibration session for sufficient training sample collection to establish classifiers and extractors of subject-specific features. The test session later employs the classifiers and extractors

to classify the subjects brain signals. In a recent study on BCIs, it was shown to be very important to reduce training sessions on account of the time-consuming, tedious process of a calibration session. As a result, conducting a performance promotion using a scarce labeled set is more desirable compared with using a large one. Nevertheless, suitable methods must be identified to strengthen the performance. This is because a short calibration session means the availability of merely a few training samples for target users, which may result in overfitting or suboptimal feature classifiers or extractors.

To address the above problem, transfer learning is a promising approach [5, 6]. It applies data represented in various feature spaces or obtained from various distributions for compensating the insufficient labeled data. In the BCI field, transfer learning has attracted considerable attention because it enables the establishing of subject-independent spatial classifiers and/or filters, and it lowers calibration times. Some studies concentrated on feature representation

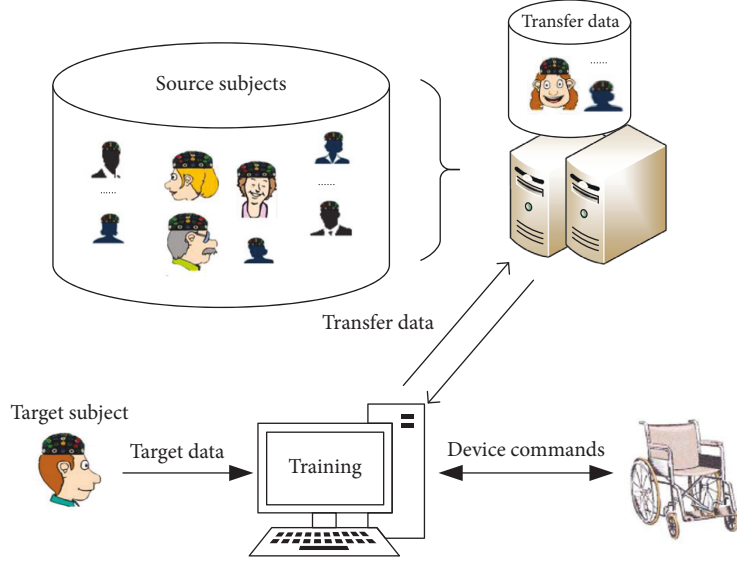


FIGURE 1: An illustration of subject transfer based BCI system.

transfer methods in EEG classification [7–11]. In this situation, we encode the knowledge that traverses domains into a fresh feature representation. Accordingly, precise classification performance is thereby expected in settings with a small sample.

A proposed schedule for practical applications of BCI systems based on subject transfer is presented in Figure 1 [12]. The datasets provided by the source subjects can be stored as a dataset group. Next, the BCI device can first acquire transfer data from the source subject groups when it is prepared to execute classification for the user. In this paper, we thus propose the transfer kernel common spatial patterns (TKCSP) method. The TKCSP computation is formulated by BCI as an optimization problem with multiple subjects, thereby incorporating data from other subjects to establish a common feature space.

2. Transfer Kernel Common Spatial Patterns

This section mentioned a new feature extraction method, TKCSP, which combines two previous approaches, kernel common spatial patterns (KCSP) [13] and transfer kernel learning (TKL) [6]. KCSP is an extraction approach for motor imagery, and TKL is a promising transfer learning method. In Sections 2.1 and 2.2, we describe the KCSP algorithm and the TKL algorithm, respectively. In Section 2.3, we would propose the TKCSP algorithm in combination with the above two algorithms.

2.1. Kernel Common Spatial Patterns. The KCSP algorithm based on CSP is used to find the components with the largest energy difference between the two experimental conditions [13–15]. Its basic idea is to find the optimal spatial filter to maximize the component energy under two sets of experimental conditions after the spatial filtering.

The first step is to calculate the covariance between the two signals. Consider E_i as an $M \times T$ matrix representing the

i th trial of EEG signals, wherein M represents the channel amount and T represents the points of time. The class-specific spatial covariance matrix can hence be acquired by the steps below.

$$R_i = \frac{K(E_i E_i^T)}{\text{trace}(K(E_i E_i^T))}, \quad (1)$$

where i represents the class label, $K(E_i E_i^T) = \langle \varphi(E_i), \varphi(E_i) \rangle$ represents the kernel function, and $\langle \cdot \rangle$ denotes the inner product. Thus we can replace the computation of the aggregate spatial covariance matrix with $R_c = R_1 + R_2$.

Additionally, we can factor R_c to be $R_c = U_0 \Lambda_c U_0^T$, where $U_0 \in \mathbf{R}^{n \times n}$ represents a matrix with eigenvectors in a row, while Λ_c represents the diagonal matrix of eigenvalues classified in declining order.

The variances can be equalized by using a whitening transmission P within space that the eigenvectors span in U_0 such that P equals

$$P = \lambda^{1/2} U_0^T. \quad (2)$$

Thirdly, the whitening matrix P can be used to transform R_1 and R_2 into S_1 and S_2 as

$$\begin{aligned} S_1 &= P R_1 P^T, \\ S_2 &= P R_2 P^T. \end{aligned} \quad (3)$$

S_1 and S_2 have the same eigenvectors, that is, if

$$\begin{aligned} S_1 &= B \lambda_1 B^T, \\ S_2 &= B \lambda_2 B^T, \\ \lambda_1 + \lambda_2 &= I, \end{aligned} \quad (4)$$

where I represents the identity matrix. At this point, the sum is always one for these two corresponding eigenvalues. Hence,

Input: Data E .
Output: Common spatial patterns W^{-1} ,
and common spatial filters W .

- (1) Compute spatial covariance matrix R_i , $i = 1, 2$ by (1)
and the total spatial covariance matrix is $R_c = R_1 + R_2$.
- (2) Eigen decomposition $R_c = U_c \Lambda_c U_c^T$;
whitening transformation $P = \lambda^{1/2} U_c^T$.
- (3) Transform covariance matrices $S_1 = PR_1 P^T$, $S_2 = PR_2 P^T$.
and eigen decomposition $S_1 = B\lambda_1 B^T$, $S_2 = B\lambda_2 B^T$.
- (4) Construct the spatial filter $W = (B^T P)^T$.

ALGORITHM 1: Kernel common spatial pattern algorithms.

the eigenvectors having the smallest eigenvalues for S_1 have the largest eigenvalues for S_2 and vice versa. This property enables eigenvector B to sort these two classes.

Finally, owing to $W = (B^T P)^T$ as the common spatial filters, the common spatial patterns are columns of W^{-1} , which can be regarded as the source distribution vectors for time-invariant EEG. Algorithm 1 shows the summary of a complete KCSP procedure.

2.2. Transfer Kernel Learning. TKL can directly match the source distribution and target distribution to learn a domain-invariant kernel space, using the knowledge of the source domain to help complete the learning tasks in the target domain. This section begins with definitions of terminology used, and Notations section presents a summary of commonly used notations.

Definition 1. A domain D includes a d -dimensional feature space \mathcal{F} as well as a marginal probability distribution $P(x)$; that is, $\mathcal{D} = \{\mathcal{F}, P(x)\}$, $x \in \mathcal{F}$.

In general, if two domains Z and X have different marginal distributions or feature spaces, they will have difference; that is, $\mathcal{F}_Z \neq \mathcal{F}_X \vee P(z) \neq P(x)$.

Definition 2. Given domain \mathcal{D} , a classifier $f(x)$ and a cardinality label set \mathcal{Y} compose a task \mathcal{T} ; that is, $\mathcal{T} = \{\mathcal{Y}, f(x)\}$, in which $y \in \mathcal{Y}$, and the interpretation of $f(x) = P(y | x)$ can be conditioned probability distribution.

In general, if two tasks \mathcal{T}_Z and \mathcal{T}_X have different conditioned distributions or label spaces, they will have a difference; that is, $\mathcal{Y}_Z \neq \mathcal{Y}_X \vee P(y | z) \neq P(y | x)$.

Problem 3 (transfer kernel learning). Given an unlabeled target domain $X = \{x_1, \dots, x_n\}$ and a labeled source domain $Z = \{(z_1, y_1), \dots, (z_m, y_m)\}$ with $\mathcal{F}_Z = \mathcal{F}_X$, $\mathcal{Y}_Z = \mathcal{Y}_X$, $P(z) \neq P(x)$, and $P(y | z) \neq P(y | x)$, a kernel $k(z, x) = \langle \phi(z), \phi(x) \rangle$ with an invariable domain is learned so that $P(\phi(z)) \simeq P(\phi(x))$. Suppose $P(y | \phi(z)) \simeq P(y | \phi(x))$, then a kernel machine targeting Z can effectively generalize X .

Firstly, calculate the target kernel function, the source kernel function, and the cross-kernel function. Assume an

input kernel function k is given to us, for example, Laplacian kernel $k(z, x) = e^{\gamma \|z-x\|}$ or Gaussian kernel $k(z, x) = e^{\gamma \|z-x\|^2}$, then the target kernel K_X , the source kernel K_Z , and the cross-domain kernel K_{ZX} can be computed. A domain-invariant kernel \bar{K}_{ZUX} can be learned by utilizing these three kernels. Under this challenging situation, the sufficient matching of marginal distributions plays an indispensable role in efficient learning of the domain transfer.

To require two datasets (for example, target data X and source data Z) to conform to similar distributions of the feature space, that is, $P(\phi(z)) \simeq P(\phi(x))$, requiring them to have similar kernel matrices is sufficient, that is, $K_Z \simeq K_X$ [16]. Nevertheless, kernel matrices depend on data and the direct evaluation of closeness between varied kernels is improbable because of the varying dimensions; that is, $K_Z \in R^{m \times m}$, $K_X \in R^{n \times n}$ [17]. To solve this issue, the Nyström kernel approximation idea is adopted for the generation of an extrapolated source kernel $\bar{K}_Z \in R^{m \times m}$ by an eigensystem of target kernel K_X . Next, \bar{K}_Z can arise to kernel K_Z as the ground truth source and can be comparable to a spectral kernel design. Figure 2 shows the whole learning procedure.

Secondly, Nyström kernel approximation is adopted to execute eigensystem extrapolation [16]. To this end, standard eigendecomposition is adopted on the target kernel K_X

$$K_X \Phi_X = \Phi_X \Lambda_X, \quad (5)$$

which provides the eigensystem $\{\Lambda_X, \Phi_X\}$ of target kernel K_X .

Thirdly, we assess the eigensystem on source data Z by utilizing the Nyström approximation theorem. The derivation of the eigenvectors $\bar{\Phi}_Z$ for extrapolated source kernel \bar{K}_Z is

$$\bar{\Phi}_Z \simeq K_{ZX} \Phi_X \Lambda_X^{-1}, \quad (6)$$

where $K_{ZX} \in R^{m \times n}$ is the cross-domain kernel matrix between Z and X , assessed by kernel function k .

The initial Nyström method directly utilizes target eigenvalues Λ_X and extrapolated source eigenvectors $\bar{\Phi}_Z$ to make approximation for the source kernel K_Z . In essence, the distribution difference across domains is embodied by the Nyström approximation error; that is, error is close to 0 if and only if $P(z) \simeq P(x)$. An invariant kernel extrapolated to varied domains will be achieved if an extrapolated kernel

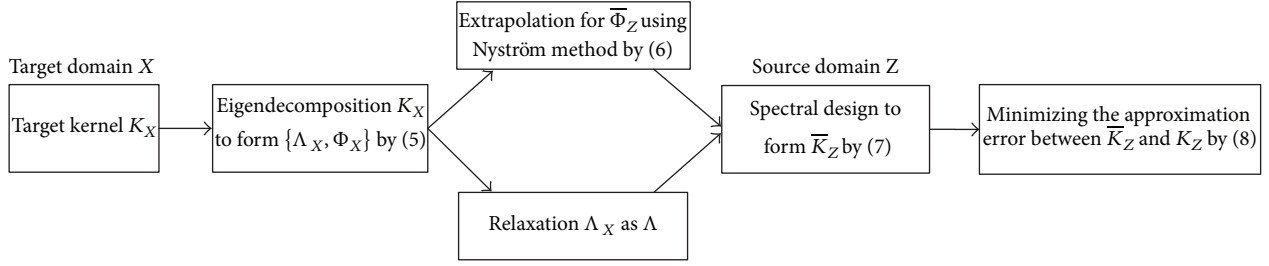


FIGURE 2: Complete procedure of transfer kernel learning.

can be found for realizing a minimized Nyström approximation error, thereby facilitating a more efficient cross-domain generalization.

The spectral kernel design idea is adopted to establish a new kernel matrix from extrapolated eigensystem to reduce the Nyström approximation error [18]. The key construction of target kernel K_X can thus be preserved by the kernel matrix generated via extrapolated eigensystem Φ_Z ; however, the flexibility of the reshaping could be retained to keep the distribution difference minimized.

Fourthly, eigenspectrum Λ_X can be relaxed in the primary Nyström approach to be parameters Λ that can be learned resulting in a kernel family extrapolated from the target eigensystem yet assessed on the source data. The extrapolated source kernel \bar{K}_Z is obtained as follows:

$$\bar{K}_Z = \bar{\Phi}_Z \Lambda \bar{\Phi}_Z^T. \quad (7)$$

The critical structures of the target domain can be preserved by this kernel family, that is, eigenvectors $\bar{\Phi}_Z$. Moreover, the free eigenspectrum Λ remains undetermined. Unlike a conventional spectral kernel design that learns the parameters through Λ trained on the spectral kernel towards a previous kernel calculated in the same domain, kernel matching can be performed across domains.

Fifthly, we strive to minimize the approximation error between the ground truth source kernel K_Z and the extrapolated source kernel \bar{K}_Z for explicitly minimizing the distribution difference herein by utilizing the squared loss

$$\begin{aligned} \min_{\Lambda} \quad & \|\bar{K}_Z - K_Z\|_{\mathcal{F}}^2 = \|\bar{\Phi}_Z \Lambda \bar{\Phi}_Z^T - K_Z\|_{\mathcal{F}}^2 \\ & \lambda_i \geq \zeta \lambda_{i+1}, \quad i = 1, \dots, n-1 \\ & \lambda_i \geq 0, \quad i = 1, \dots, n, \end{aligned} \quad (8)$$

where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ belongs to the n nonnegative eigenspectrum parameters, while $\zeta \geq 1$ belongs to the eigenspectrum damping factor [19].

The marginal distributions of multiple source domains can be matched with the target domain using the generalized transfer kernel learning (TKL) approach. This approach can be conducted by the source-specific eigenspectrum Λ learning for every source domain separately in the initial place. Secondly, existing learning algorithms of multiple sources are used to implement consensus forecasting for the target

domain on the basis of predicting multiple source domains [20, 21].

Sixthly, the standard quadratic programming possessing (QP) linear constraints are used herein to show the solution of the TKL optimization problem (8). Here, n eigenspectrum parameter is denoted as $\lambda = (\lambda_1, \dots, \lambda_n)$; that is, $\Lambda = \text{diag}(\lambda)$. Equation (8) is reformulated in the matrix form by linear algebra

$$\begin{aligned} \min_{\lambda} \quad & \lambda^T Q \lambda - 2\gamma^T \lambda \\ & C \lambda \geq 0 \\ & \lambda \geq 0. \end{aligned} \quad (9)$$

The following are the respective definitions of QP coefficient matrices Q , γ and constraint matrix

$$\begin{aligned} Q &= (\Phi_Z^T \Phi_Z) \odot (\Phi_Z^T \Phi_Z) \\ \gamma &= \text{diag}(\Phi_Z^T K_Z \Phi_Z) \\ C &= I - \zeta \bar{I}, \end{aligned} \quad (10)$$

where $\zeta \geq 1$ represents the eigenspectrum damping factor, which is also the only tunable parameter within TKL. Additionally, $I \in \mathbf{R}^{n \times n}$ denotes the identity matrix, and $\bar{I} \in \mathbf{R}^{n \times n}$ represents the first diagonal matrix with the nonvanishing elements.

Finally, constructing the domain-invariant kernel \bar{K}_A on the target and source data $A = Z \cup X$ is straightforward with the learned optimal eigenspectrum parameters Λ . According to spectral kernel design, we can generate \bar{K}_A from the eigensystem $\{\Lambda, \bar{\Phi}_A\}$ invariant to domain

$$\bar{K}_A = \begin{bmatrix} \bar{\Phi}_Z \Lambda \bar{\Phi}_Z^T & \bar{\Phi}_Z \Lambda \bar{\Phi}_X^T \\ \bar{\Phi}_X \Lambda \bar{\Phi}_Z^T & \bar{\Phi}_X \Lambda \bar{\Phi}_X^T \end{bmatrix} = \bar{\Phi}_A \Lambda \bar{\Phi}_A^T, \quad (11)$$

where $\bar{\Phi}_A \triangleq [\bar{\Phi}_Z; \bar{\Phi}_X]$ belongs to extrapolated eigenvectors on all data A . We can directly feed the kernel \bar{K}_A invariant to the domain to normal kernel machines, for example, KCSF, for facilitating the cross-domain generalization and prediction. Algorithm 2 shows the summary of a complete procedure.

Input: Data X ; kernel k ; eigen-damping factor ζ .
Output: Domain-invariant kernel \bar{K}_A .

- (1) Compute the target kernel function K_X , the source kernel function K_Z , and cross-domain kernel function K_{ZX} by kernel k .
- (2) Eigendecompose the target kernel function K_X for the eigensystem $\{\Lambda_X, \Phi_X\}$ by (5).
- (3) Extrapolate for source eigenvector $\bar{\Phi}_Z$ via Nyström approximation by (6).
- (4) Generate the extrapolated source kernel \bar{K}_Z by (7)
- (5) Minimize the approximation error between the ground truth source kernel K_Z and the extrapolated source kernel \bar{K}_Z by (8).
- (6) Solve QP problem (9) for eigenspectrum λ .
- (7) Construct domain-invariant kernel \bar{K}_A by (11).

ALGORITHM 2: TKL algorithm.

Input: Data E_X, E_Z .
Output: Common Spatial Patterns W^{-1} .

- (1) Compute transfer kernel \bar{K}_A by Algorithm 2.
- (2) Compute spatial covariance matrix $R_i = \bar{K}_A / \text{trace}(\bar{K}_A)$, $i = 1, 2$.
And the total spatial covariance matrix is $R_c = R_1 + R_2$.
- (3) Eigen decomposition $R_c = U_c \Lambda_c U_c^T$;
whitening transformation $P = \lambda^{1/2} U_0^T$.
- (4) Transform covariance matrices $S_1 = P R_1 P^T$, $S_2 = P R_2 P^T$
and eigen decomposition $S_1 = B \lambda_1 B^T$, $S_2 = B \lambda_2 B^T$.
- (5) Construct the spatial filter $W = (B^T P)^T$.

ALGORITHM 3: Transfer kernel common spatial pattern algorithm.

2.3. Transfer Kernel CSP. When transfer kernel \bar{K}_A replaces kernel $K(E_i E_i^T)$ in (1), we can build the TKCSP. For all methods based on the kernel, linear kernel is adopted by us; that is, $K(x_i, x_j) = x_i^T x_j$. Then $R_i = K_A / \text{trace}(K_A)$ can be used to estimate the spatial covariance. Algorithm 3 presents a summary of a complete TKCSP procedure.

We can compute the filtration of a trial E_j by $W = (B^T P)^T$ as the projection matrix [14]:

$$Z_i = W \times E_i. \quad (12)$$

Decomposing the EEG based on (6) can be used to obtain the features utilized for classification. For every imagined movement direction, the classifier construction employs the variances owned by merely a small amount of m signals that are the fittest for discrimination. The signals Z_p ($p = 1 \cdots 2m$) maximizing the variance difference of motor imagery EEG on the left versus the right belong to those associated with the largest eigenvalues λ_1 and λ_2 . These signals are blank in the last and first rows in Z because of the computation of W

$$f_p = \frac{\text{var}(Z_p)}{\sum_{i=1}^{2m} \text{var}(Z_i)}. \quad (13)$$

TABLE 1: Data description for dataset IVa in BCI Competition III.

Subject	aa	al	av	aw	ay
Number of training samples	168	224	84	56	28
Number of test samples	112	56	196	224	252

The linear classifier can be calculated by using the feature vectors f_p of right and left trials. The log-transformation contributes to approximating the standard data distribution.

3. Experiments

3.1. Data Preparation. In this study, we employed the IVa dataset from BCI Competition III [22]. The dataset includes EEG data containing a classification task of motor imagery with two levels: (1) imagery movement of the right hand (denoted by R) and (2) imagery movement of the right foot (denoted by F). We employed 118 electrodes to measure EEG signals in every trial from five different subjects, and each subject involved the performance of 280 trials. Table 1 presents a summary of the data descriptions, in which the number of subjects av, aw, ay of training samples is fewer than those of the test samples.

TABLE 2: Classification accuracy of TKCSP on the dataset.

Source	Target				
	aa (%)	al (%)	av (%)	aw (%)	ay (%)
aa	-	80.58	50.17	84.33	80.41
al	67.20	-	63.87	90.58	81.98
av	50.00	86.11	-	80.10	55.13
aw	62.23	93.88	54.39	-	50.00
ay	50.00	91.00	65.67	50.00	-
aa + al	-	-	57.43	87.69	84.65
aa + av	-	84.58	-	81.56	74.11
aa + aw	-	87.80	51.76	-	51.53
aa + ay	-	84.58	-	81.56	74.11
al + av	59.27	-	-	88.35	50.00
al + aw	68.10	-	61.96	-	50.00
al + ay	62.92	-	68.47	78.16	-
av + aw	52.44	93.11	-	-	50.00
av + ay	45.30	89.38	-	74.98	-
aw + ay	58.60	91.65	61.25	-	-
aa + al + av	-	-	-	88.10	73.25
aa + al + aw	-	-	50.00	-	58.60
aa + al + ay	-	-	58.47	80.01	-
aa + av + aw	-	83.37	-	-	50.00
aa + av + ay	-	85.68	-	78.10	-
aa + aw + ay	-	89.59	50.00	-	-
al + av + aw	62.56	-	-	-	50.00
al + av + ay	52.10	-	-	75.48	-
al + aw + ay	63.13	-	52.47	-	-
av + aw + ay	55.65	91.67	-	-	-
al + av + aw + ay	62.07	-	-	-	-
aa + av + aw + ay	-	87.68	-	-	-
aa + al + aw + ay	-	-	60.63	-	-
aa + al + av + ay	-	-	-	81.20	-
aa + al + av + aw	-	-	-	-	53.19

Each trial was considered an $M \times T$ matrix E_i , in which M represents the electrode amount and the time point amount sampled. EEG signals measured were band-pass decomposed (8–30 Hz). SVM (Support Vector Machine) involving linear kernel was utilized as the classifier. The proportion of the number of samples properly classified to the aggregate number of used samples in this test was employed to evaluate the classification precision.

Our establishment of a dataset (containing the target domain and source domain) for cross-domain classification is described as follows. The dataset of each subject could become the target domain (ay, aw, av, al, aa), while the dataset of other subjects could become the source domains. This strategy of dataset construction ensured the relevance between domains of unlabeled and labeled data, as they were located in the same top-level categories. Accordingly, $C_4^1 + C_4^2 + C_4^3 + C_4^4 = 15$ datasets of the source domains were generated for each target domain. It was possible to generate five dataset groups, including $5 \times 15 = 75$ datasets.

3.2. Experimental Results. In this section, TKCSP and six competitive methods are evaluated based on classification accuracy [8, 11, 23]. We established five dataset groups from the dataset described above. Each dataset group includes four source subjects in source domains and one target subject as target domain. If one subject is the target domain, it will no longer appear in the source domains, so that each target domain corresponds to 15 different source domains. The first column of Table 2 shows the different source domains, and the second column to the sixth column of Table 2 show the classification accuracy of each target domain in its source domains, respectively. Among them, the highest classification accuracy of target domain aa was 68.10% and the corresponding source domain was al + aw; the highest classification accuracy of target domain al was 93.88% and the corresponding source domain was aw; the highest classification accuracy of target domain av was 68.47% and the corresponding source domain was al + ay; the highest classification accuracy of target domain aw is 68.10% and the

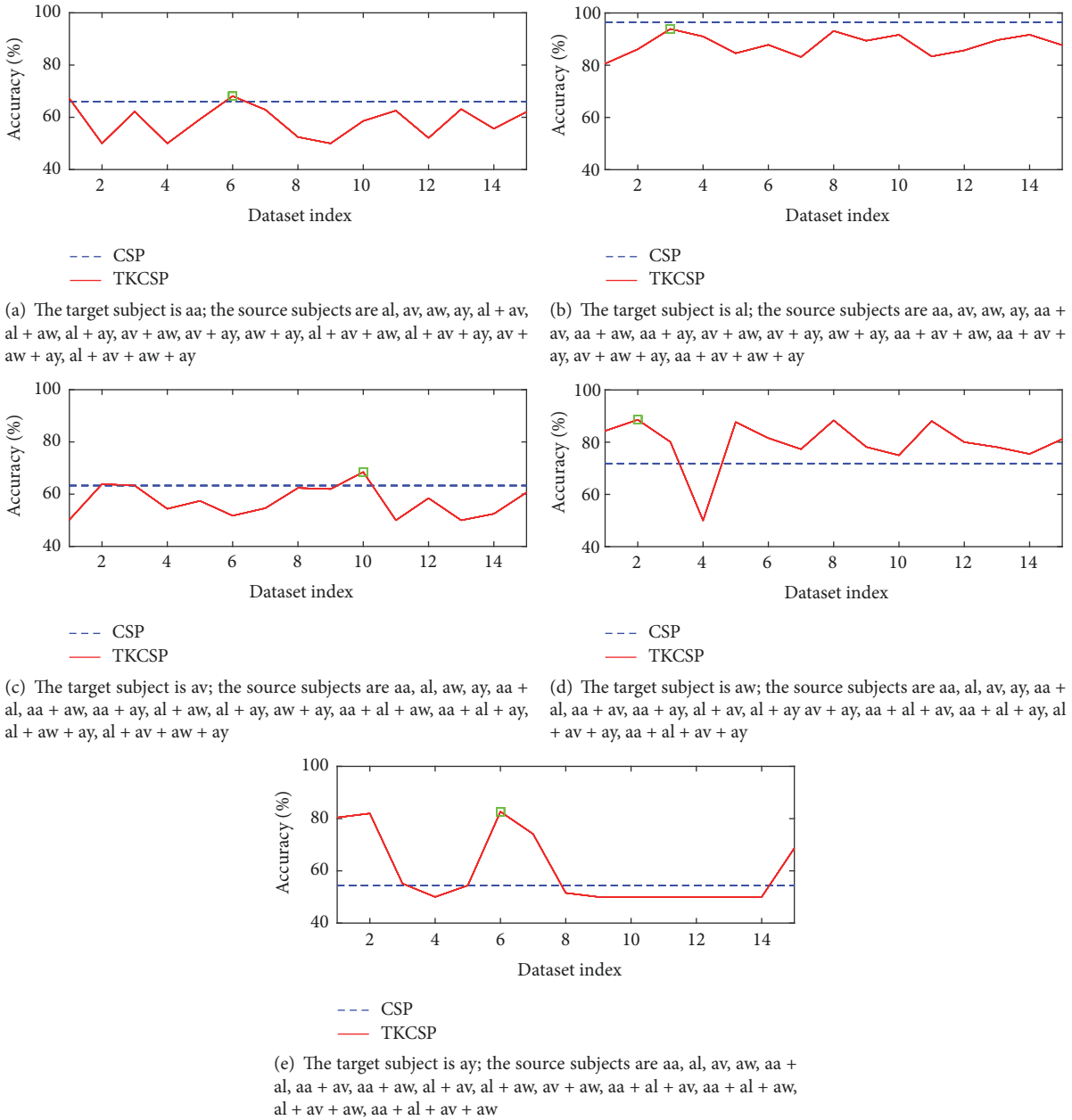


FIGURE 3: Classification accuracy of TKCSP and CSP on the dataset.

corresponding source domain is al; the highest classification accuracy of target domain ay is 68.10% and the corresponding source domain is aa + al.

To be complete, we detail the results of TKCSP method and CSP approach on all of 5 dataset groups in Figures 3(a), 3(b), 3(c), 3(d), and 3(e), where each figure presents the results of each group. In Figure 3, the blue dashed line indicates the classification accuracy of CSP algorithm, the red solid line indicates the classification accuracy of TKCSP algorithm in different source domains, and the green square indicates the best classification accuracy of TKCSP. The horizontal axis of green square is corresponding to the optimal source domain. The results show that the

classification accuracy of TKCSP method is better than that of CSP algorithm.

Table 3 lists the classification (recognition) precisions of five comparison approaches and TKCSP on dataset IVa. Figure 4 visually depicts the results for improved accessibility. The performance achieved by TKCSP is significantly better than those achieved by the five comparison approaches. Several observations can be made from these results.

Firstly, TKCSP achieves classification precision on the aw and aa datasets as 90.58% and 68.47%, respectively. These are higher than those of the five comparison approaches. Moreover, TKCSP achieves an average classification precision on these datasets as 81.14%, providing a significant performance

TABLE 3: Comparison of classification accuracy for TKCSP and 6 competitive methods.

Subject	aa (%)	al (%)	av (%)	aw (%)	ay (%)	Mean (%)
CSP	66.07	96.43	63.30	71.88	54.40	70.42
RCSP	71.43	96.43	63.30	71.88	86.90	77.98
CSP SJ-to-SJ	67.76	98.41	60.20	78.72	74.78	75.97
ssCSP	67.00	94.62	58.26	89.35	85.71	78.99
mtCSP	72.33	94.62	68.39	65.57	83.14	76.81
ss + mtCSP	71.43	94.63	66.32	88.40	74.93	79.17
TKCSP	68.10	93.88	68.47	90.58	84.65	81.14

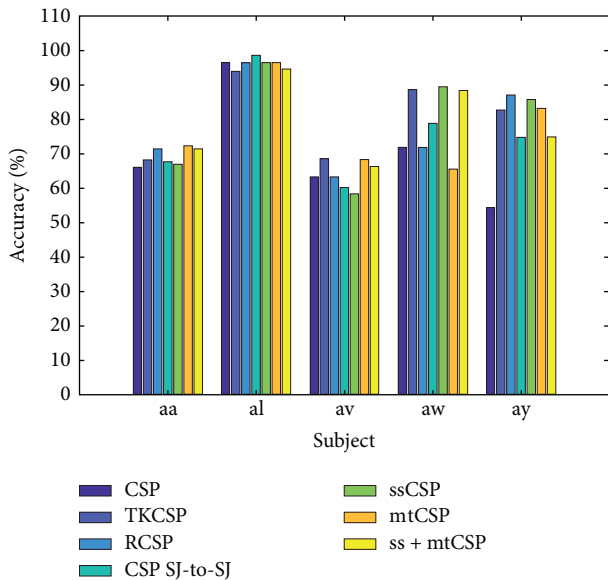


FIGURE 4: Comparison of classification accuracies for TKCSP and 6 competitive methods.

improvement of 1.97% over ss + mtCSP, the best competitive approach. It is strongly verified by the consistent performance improvements on these datasets that TKCSP can successfully establish powerful domain kernels for classification of cross-domain motor imagery.

Then, a composite covariance matrix as a weighted total of covariance matrices, including subjects resulting in a composite CSP, is determined by CSP for subject-to-subject transfer (CSP SJ-to-SJ). This approach thus achieves an average classification precision of 75.97%.

Thirdly, regularizing CSP (RCSP) is intended to regularize the covariance matrix to the mean covariance matrix of other subjects for improving its estimation performance. Such regularization is particularly promising in settings with small samples. Furthermore, this approach achieves an average classification precision of 77.98%.

Finally, the stationary subspace CSP (ssCSP) focuses on the nonstationarity issue while multitask CSP (mtCSP) focuses on the estimation issue. The combined mtCSP and

ssCSP (ss + mtCSP) method employs both approaches. That is, the nonstationary subspace acquired by ssCSP is firstly projected, and then the spatial filters are computed with mtCSP by regularization parameters acquired when it is applied to the initial data. The three above approaches achieve an average classification precision of 78.99%, 76.81%, and 79.17%, respectively.

In particular, TKCSP can assess the various cluster structures and naturally matches them between multiple domains. This procedure is achieved by TKCSP through the matching between the source domain kernel and kernel extrapolated from the target domain, while simultaneously increasing (declining) the domain-invariant (domain independent) eigenspectrum. The superior performance of TKCSP can be explained by this advantage.

4. Conclusion

In this paper, we proposed the TKCSP method to lower the training trial amount and improve the performance via learning a domain-independent kernel. To this end, direct matching of distributions between target subjects and source subjects within the kernel space is conducted. TKCSP and six competitive approaches were evaluated on EEG datasets provided by BCI Competition III. The results showed that the performance of the best approach, RCSP, was better than that of CSP by nearly 1.97% in terms of the mean classification precision. The results also revealed that RCSP can perform effective subject-to-subject transfer. Therefore, the behaviors matched with knowledge of neurophysiology could be classified by the TKCSP approach.

Notations

- Z, X : Source/target domain
- m, n : Source/target examples
- d, c : Features/classes
- γ, ζ : Eigenvectors/damping factor
- \mathbf{X} : Input data matrix
- \mathbf{K} : Kernel matrix
- Φ : Eigenvector matrix
- Λ : Eigenvalue matrix.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported in part by the Department of S and T for Social Development of China (2014BAF08B01); Program for Innovative Research Team of Ministry of Education of China (IRT1203); the youth talent support program of Beihang University, Key Laboratory of Precision Opto-Mechatronics Technology (Beihang University), Ministry of Education. The authors would like to thank the Laboratory of Brain-Computer Interfaces at the Graz University of Technology for providing their data.

References

- [1] A. Cichocki, Y. Washizawa, T. Rutkowski et al., “Noninvasive BCIs: multiway signal-processing array decompositions,” *The Computer Journal*, vol. 41, no. 10, pp. 34–42, 2008.
- [2] T. Ebrahimi, J.-M. Vesin, and G. Garcia, “Brain-computer interface in multimedia communication,” *IEEE Signal Processing Magazine*, vol. 20, no. 1, pp. 14–24, 2003.
- [3] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, “Brain-computer interfaces for communication and control,” *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.
- [4] Y. Ma, X. Ding, Q. She, Z. Luo, T. Potter, and Y. Zhang, “Classification of motor imagery EEG signals with support vector machines and particle swarm optimization,” *Computational and Mathematical Methods in Medicine*, vol. 2016, Article ID 4941235, 8 pages, 2016.
- [5] 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.
- [6] M. Long, J. Wang, J. Sun, and P. S. Yu, “Domain invariant transfer kernel learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 6, pp. 1519–1532, 2015.
- [7] A. Satti, C. Guan, D. Coyle, and G. Prasad, “A covariate shift minimization method to alleviate non-stationarity effects for an adaptive brain-computer interface,” in *Proceedings of the 2010 20th International Conference on Pattern Recognition, ICPR 2010*, pp. 105–108, Turkey, August 2010.
- [8] H. Kang, Y. Nam, and S. Choi, “Composite common spatial pattern for subject-to-subject transfer,” *IEEE Signal Processing Letters*, vol. 16, no. 8, pp. 683–686, 2009.
- [9] F. Lotte and C. Guan, “Learning from other subjects helps reducing brain-computer interface calibration time,” in *Proceedings of the 35th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '10)*, pp. 614–617, Dallas, Tex, USA, March 2010.
- [10] D. Devlaminc, B. Wyns, M. Grosse-Wentrup, G. Otte, and P. Santens, “Multisubject learning for common spatial patterns in motor-imagery BCI,” *Computational Intelligence and Neuroscience*, vol. 2011, Article ID 217987, 9 pages, 2011.
- [11] W. Samek, F. C. Meinecke, and K. Müller, “Transferring subspaces between subjects in brain—computer interfacing,” *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 8, pp. 2289–2298, 2013.
- [12] W. Tu and S. Sun, “A subject transfer framework for EEG classification,” *Neurocomputing*, vol. 82, pp. 109–116, 2012.
- [13] H. Albalawi and X. Song, “A study of kernel CSP-based motor imagery brain computer interface classification,” in *Proceedings of the 2012 IEEE Signal Processing in Medicine and Biology Symposium, SPMB 2012*, USA, December 2012.
- [14] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, “Optimal spatial filtering of single trial EEG during imagined hand movement,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, 2000.
- [15] X. Yin, B. Xu, C. Jiang et al., “A hybrid BCI based on EEG and fNIRS signals improves the performance of decoding motor imagery of both force and speed of hand clenching,” *Journal of Neural Engineering*, vol. 12, no. 3, Article ID 036004, 2015.
- [16] K. Zhang, I. W. Tsang, and J. T. Kwok, “Improved Nystrom lowrank approximation and error analysis,” in *Proceedings of the 25th International Conference on Machine Learning*, 2008.
- [17] C. Williams and M. Seeger, “Using the Nystrom method to speed up kernel machines,” in *Proceedings of the Advances in Neural Information Processing Systems*, 2001.
- [18] R. Johnson and T. Zhang, “Graph-based semi-supervised learning and spectral kernel design,” *IEEE Transactions on Information Theory*, vol. 54, no. 1, pp. 275–288, 2008.
- [19] R. Jin, T. Yang, M. Mahdavi, Y.-F. Li, and Z.-H. Zhou, “Improved bounds for the Nystrom method with application to kernel classification,” *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 59, no. 10, pp. 6939–6949, 2013.
- [20] L. Duan, D. Xu, and I. W.-H. 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.
- [21] W. Dai, G. Xue, Q. Yang, and Y. Yu, “Co-clustering based classification for out-of-domain documents,” in *Proceedings of the 13th ACM SIGKDD international conference*, p. 210, San Jose, Calif, USA, August 2007.
- [22] B. Blankertz, K.-R. Müller, D. J. Krusienski et al., “The BCI competition III: validating alternative approaches to actual BCI problems,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 153–159, 2006.
- [23] F. Lotte and C. Guan, “Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms,” *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2011.



Hindawi

Submit your manuscripts at
www.hindawi.com

