Available online at www.sciencedirect.com
ScienceDirect
Neurocomputing 00 (2011) 1–11
NEUROCOMPUTING

A subject transfer framework for EEG classification

Wenting Tu, Shiliang Sun*

Department of Computer Science and Technology, East China Normal University
500 Dongchuan Road, Shanghai 200241, P. R. China

Abstract

This paper proposes a subject transfer framework for EEG classification. It aims to improve the classification performance when the training set of the target subject (namely user) is small owing to the need to reduce the calibration session. Our framework pursues improvement not only at the feature extraction stage, but also at the classification stage. At the feature extraction stage, we first obtain a candidate filter set for each subject through a previously proposed feature extraction method. Then, we design different criterions to learn two sparse subsets of the candidate filter set, which are called the robust filter bank and adaptive filter bank, respectively. Given robust and adaptive filter banks, at the classification step, we learn classifiers corresponding to these filter banks and employ a two-level ensemble strategy to dynamically and locally combine their outcomes to reach a single decision output. The proposed framework, as validated by experimental results, can achieve positive knowledge transfer for improving the performance of EEG classification.

Keywords:
EEG Classification, Transfer Learning, Ensemble Learning, Sparse Representation

1. Introduction

During recent years, the development of brain computer interface (BCI) technology has both theoretical and practical significance. BCIs have the ability to enable their users to manipulate an external device by means of translating brain activities into a command for a computer or machine [1]. They are communication and control systems that do not require any peripheral muscular activities, and therefore can be a very helpful aid to people suffering from motor disabilities [2]. As Fig. 1 shows, BCIs can be seen as a complex pattern recognition system [3], where the user's ability to reliably produce changes of electroencephalogram (EEG) signals and subsequent stages of feature extraction and classification are equally important and can complement one another.

Improving classification performances of EEG-based BCI systems faces a great challenge today. One problem is that, for a new subject (user), a long calibration session (e.g. more than one hour) is needed to collect sufficient training samples to construct subject-specific feature extractors and classifiers, which are used later in the test session to classify brain signals of this subject. In recent research in BCIs, reducing training sessions is a task of great sense since the calibration session is a boring and time-consuming process. Therefore, it is more desirable to conduct performance improvement with a small labeled set, rather than on an abundant one. However, because a short calibration session

*Corresponding author. Tel.: +86 21 54345186; fax: +86 21 54345119. E-mail address: slsun@cs.ecnu.edu.cn (S. Sun).

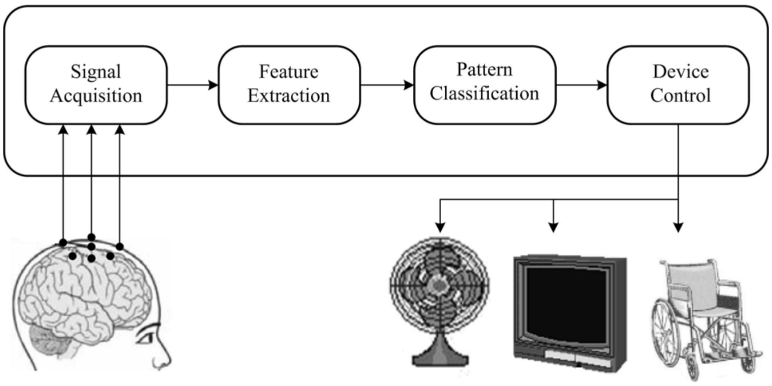


Figure 1: A general EEG-based BCI [4]

means only a few training samples of the target user are available, which may lead to suboptimal or overfitting feature extractors or classifiers, we have to find appropriate methods to enhance the performance.

One promising way to reduce training sessions is to utilize the samples collected from other subjects (we call them as “source subjects”) to aid the subject whose brain signals would be classified in the test session (we call this subject as “target subject”). This strategy can be termed “subject transfer”. However, owing to the large inter-subject variabilities, it is unwise to simply add training samples of source subjects to the training set of the target subject. This is often unhelpful and even degrades the performance. How to properly use the data of source subjects is the key to achieving positive subject transfer for EEG classification.

Several related works also focusing on the small training sample problem in EEG classification are described here. Semi-supervised learning with local temporal regularization [5] has been proposed to utilize test samples to solve the small training sample problem. However, for real BCI systems, collecting a lot of test samples are sometimes unpractical, so that semi-supervised technology may be unsuitable. Another important work for this problem is the session transfer strategy [6]. However, this method assumes the target user has already performed some training sessions, so that it can not work for a new user who did not perform any training session before. Moreover, there are several works on adaptive learning also related to this problem [7, 8, 9]. This paper proposes a framework for achieving subject transfer strategy. It can be applied to users with few training samples and it handles test samples with the one-after-one manner, which is similar to the real situation of most BCI systems.

This paper proposes a framework for improving EEG classification performance. It can achieve positive subject transfer with improvement on both feature extraction and classification stages. Traditional spatial filtering algorithms often have some important limitations: It is not flexible enough and often overfitts or underfits, especially in the small training sample situation (for details, see [10, 11, 12, 13]). Thus, we design a method to obtain two banks that ensure robustness and adaptiveness of spatial filtering, respectively. Specifically speaking, it uses a previous method called extreme energy ratio (EER) to obtain candidate filter banks, and then extracts its two subsets with 1-norm regularization and different performance criterions. Given multiple classifiers corresponding to these banks, we employ an ensemble strategy to combine them. The proposed fusion method assigns dynamic weights to these base models according to the local structure of a given test sample in the feature space. These weights represent the prediction consistency of the model. With this weighting approach, we can learn robust ensemble learner and adaptive ensemble learner, respectively. Finally, a parameter weighting combination of them makes the final prediction on the category of the given test sample. Fig. 2 provides a structural illustration of our framework. The experiment is performed on public datasets from nine subjects and the results demonstrate the excellent performance of our method.

The rest of this paper is organized as follows. Subject transfer based BCI systems are introduced in Section 2. Section 3 displays main notations used in this paper. Next, the feature extraction stage and classification stage of our framework are shown in Section 4 and Section 5, respectively. In Section 6, experimental results and performance analysis are provided. At last, conclusions and plans for future work are given in Section 7.

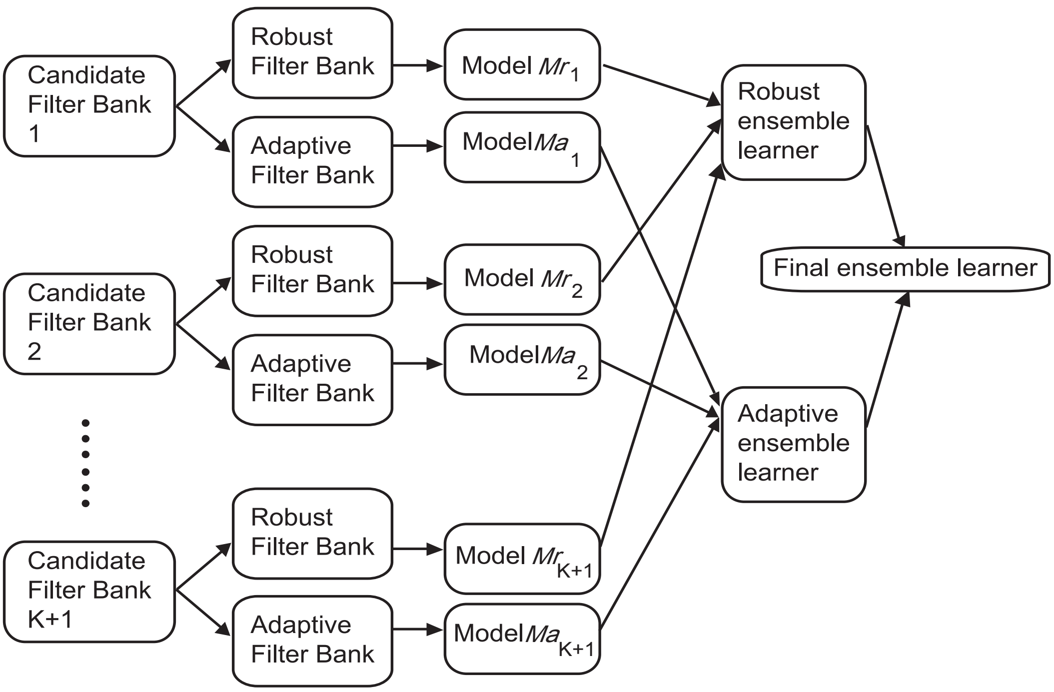


Figure 2: The proposed subject transfer framework for EEG classification.

2. Subject Transfer based BCI Systems

In this part, the subject transfer based BCI system is proposed. It wishes to use data from other subjects to reduce training burden of the current user.

It is known that the procedure of using BCI devices that includes two parts: training and test sessions. Before a BCI device can be served for users, users should firstly perform training sessions to provide enough training samples to systems. This can be achieved by two steps: First, devices show a command by offering a visual or audible cue and users should act accordingly By that way, BCI systems can collect enough training samples to learn a model for performing later tasks The training session of a BCI device often consumes half to one hour.

However, long time-consuming training sessions bring huge difficulties to the practical wide use of BCI devices. First, not all users can perform training sessions (e.g. people who have visual or audible handicaps) and sometimes users who wish to use the BCI devices in a hurry are not willing to perform long time-consuming sessions. All these situations can be solved by transferring other training sets to help the current test task. Subject transfer schedule aims to use training sets from other users to help the current user. Fig. 3 shows a recommend for practical applications of subject transfer based BCI systems. The datasets from source subjects can be stored as a dataset group. Then, when the BCI device is ready to perform classification task for the user, it can firstly obtain transfer data from source subject group. For example, it can select training sets of users whose characteristics are similar with the target user (e.g. same age or sex). Then, after training with these datasets, the device can perform test sessions for the target user. Owing to large inter-subject variances, how to extract and transfer knowledge from source datasets is the key challenge for the subject transfer schedule.

3. Notations

Denote an observed EEG sample as an $N \times T$ matrix x, where N is the number of recording electrodes and T is the number of total points during the recording period. For a subject, its training set is denoted as X.

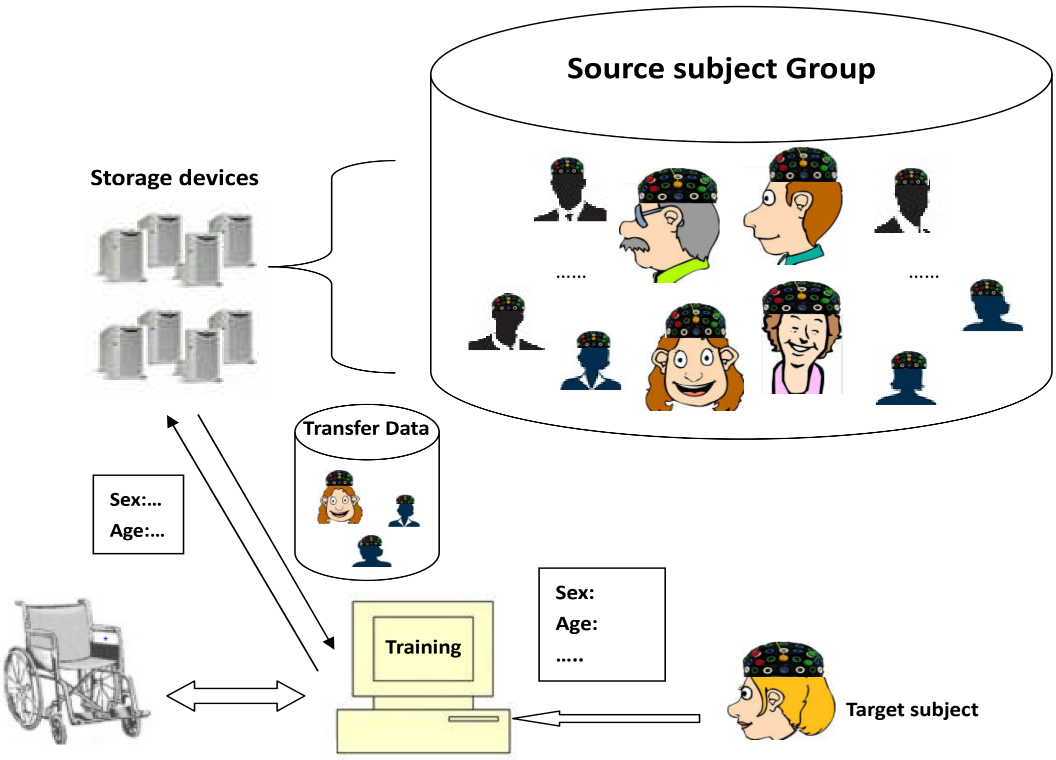


Figure 3: Subject transfer based BCI systems.

Here, suppose we have training sets from K source subjects denoted as $X_{S_1}, X_{S_2}, \cdots, X_{S_K}$ where $X_{S_j} = \{x_1^{S_j}, x_2^{S_j}, \cdots, x_{n_j}^{S_j}\}$ $(j = 1, 2, \cdots, K)$, and n_j is the number of samples in the training set X_{S_j}. The label set of the training set X_{S_j} is denoted as $Y_{S_j} = \{y_1^{S_j}, y_2^{S_j}, \cdots, y_{n_j}^{S_j}\}$, where $y_i^{S_k} \in \{-1, 1\}$ is the label of $x_i^{S_j}(i = 1, 2, \cdots, n_j)$. Moreover, we have a small training set of the target subject denoted as $X_{S_{K+1}} = \{x_1^{S_{K+1}}, x_2^{S_{K+1}}, \cdots, x_{n_{K+1}}^{S_{K+1}}\}$ and the corresponding label set $Y_{S_{K+1}} = \{y_1^{S_{K+1}}, y_2^{S_{K+1}}, \cdots, y_{n_{K+1}}^{S_{K+1}}\}$.

4. Feature Extraction Stage: Spatial Filter Bank Construction

4.1. Candidate Filter Bank Construction

We construct the candidate filter bank through a previous feature extraction method called EER algorithm [14]. The EER algorithm aims at learning spatial filters which maximize the variance of spatially filtered EEG signals from one class while minimizing the variance of signals from the other class, and has been proven to be theoretically equivalent and computationally superior to the commonly used common spatial patterns (CSP) method for EEG signal processing [14, 15, 16].

Assume only one latent signal source from each class is to be recovered. For an EEG sample x, the spatially filtered signal with a spatial filter denoted by $\phi_{(N\times1)}$ will be $\phi^\top x$. The signal energy after filtering can be represented by the sample variance as $(\phi^\top x)(\phi^\top x)^\top \triangleq \phi^\top C\phi$, where C is the normalized covariance of one EEG sample and given by

$$C = \frac{1}{T-1}\frac{xx^\top}{tr(xx^\top)}. \tag{1}$$

Ignoring the multiplicative factor $1/(T-1)$ in the following calculation of covariance, the covariances for specific classes can be computed as the average of all single covariances so as to get a more accurate and stable covariance estimate:

$$C_A = \frac{1}{T_A}\sum_{p=1}^{T_A}\frac{x_p x_p^\top}{tr(x_p x_p^\top)}, \tag{2}$$

$$C_B = \frac{1}{T_B}\sum_{q=1}^{T_B}\frac{x_q x_q^\top}{tr(x_q x_q^\top)}, \tag{3}$$

where T_A (T_B) is the number of samples from class A (B) and x_p (x_q) is the pth (qth) sample belongs to class A (B). Therefore, in order to maximize the difference of energy features under two conditions, EER finds a spatial filter which maximizes or minimizes their ratio to maximize the difference of their energy features. Thus, the discriminative EER criterion is defined as:

$$\max / \min \frac{\phi^\top C_A \phi}{\phi^\top C_B \phi}, \tag{4}$$

which can be extended readily to extract m latent source signals from each class. After eigen-decomposition, EER always selects m eigenvectors from top and end of the eigenvalue spectrum respectively to construct the spatial filter bank. In other words, the traditional spatial filter bank includes $2m$ spatial filters half of which are eigenvectors corresponding to the top of the eigenvalue spectrum and the other half are ones corresponding to the end of the eigenvalue spectrum. The spatial filter bank constructed by this strategy has some obvious limitations. First, the number of spatial filters in the bank is always even, which makes the method inflexible. Second, it always includes the first and the last eigenvector of the eigenvalue spectrum. However, due to the non-stationary nature of brain signals and the existence of outliers, those two spatial filters may overfit on training set and thus are probably unsuitable to be included in the filter bank. Third, the number of the spatial filters in the bank can be larger or smaller than the optimal but unknown number, which often cause overfitting or underfitting defects.

Therefore, we only use the group of eigen-vectors as the candidate filter bank. Sparse strategy is then employed to help us build spatial filter banks. Moreover, adaptiveness and robustness of our subject transfer framework are ensured with different objective functions.

4.2. Robust and Adaptive Filter Bank Construction

For X_{S_j} ($j = 1, 2, \cdots, K+1$), given its candidate filter set $\Phi_j = \{\phi_{j1}, \phi_{j2} \cdots, \phi_{jN}\} \in \mathbb{R}^{N\times N}$, we learn its two subsets called robust filter bank Φ_{rj} and adaptive filter bank Φ_{aj} by means of sparse representation. Note that each filter in the candidate filter set recovers a source signal:

$$S = \phi^T X. \tag{5}$$

The source signals recovered by different filters have different importance. Some of them are nonsensitive to inter-subject variabilities, and able to extract generic common discriminative features. These filters are potential to construct robust filter banks. In addition, there may be other filters that can recover source signals that are adaptive to the target subject. These filters should also be selected out to serve as the adaptive filter banks. The robust filter bank can alleviate the overfitting trend of the adaptive one. Meanwhile, the adaptive filter bank helps to improve the performance of classifying brain signals of the target subject.

Here we employ a penalized filter selection to construct these two kinds of filter banks. It includes a classification objective function and a penalty term. We design different classification performance criterions to learn banks that extract robust and adaptive source signals to construct robust and adaptive filter banks, respectively. Our choice for the second term is L1 (1-norm) penalty regularization [17]. It is a popular technology to enforce solution sparsity and has been the driving force for many emerging fields in signal processing, such as sparse coding and compressive sensing.

4.2.1. Robust Filter bank

Given source subjects with rich training samples, the generalization ability can be evaluated by the average classification ability on the training sets of all source subjects $S_1, S_2, \cdots, S_K$. Note that this definition loses effectiveness if there is considerable variability in the size of sample numbers of source subjects. However, this negative situation would not happen in the subject transfer based BCI systems since we can control the size regularity of source training sets. First, subjects with limit training samples should not be included in the source subject candidates. Second, since the training session of BCI is a boring and time-consuming process, it is impossible that a user would be pleased to provide huge training samples. Even if a user provides a huge training set, we can extract a subset whose size is close to other source training sets.

Assume we wish to select filters from Φ_j ($j = 1, 2, \cdots, K+1$) to construct a robust filter bank Φ_{rj}. We firstly express each sample $x_i^{S_k}$ ($i = 1, 2, \cdots, n_k$, $k = 1, 2, \cdots, K+1$) as its energy feature expression $\widetilde{x}_i^{S_k} = (\widetilde{x}_{i1}^{S_k}, \widetilde{x}_{i2}^{S_k}, \cdots, \widetilde{x}_{iN}^{S_k})^\top$, where $\widetilde{x}_{if}^{S_k}$ is the energy feature corresponding to the f-th filter in Φ_j : $\widetilde{x}_{if}^{S_k} = (\phi_{jf}^\top x_i^{S_k})(\phi_{jf}^\top x_i^{S_k})^\top$. Then, with that expression, the robust filter bank Φ_{rj} can be learned by the optimization problem of minimizing:

$$\frac{1}{K}\sum_{k=1}^{K}\frac{1}{n_k}\left(\sum_{i=1}^{n_k}(y_i^{S_k} - \beta_{rj}^\top \widetilde{x}_i^{S_k})^2\right) + \lambda_{rj}\|\beta_{rj}\|_1, \tag{6}$$

where n_k is the sample number of the training set from subject k, and $\beta_{rj} = [\beta_{rj1}, \cdots, \beta_{rjN}]^\top$, $\lambda_{rj} \geqslant 0$. The first term in (6) is an error function calculated on the training sets of all source subjects and $1/n_k$ is used to balance the influences of source training sets when constructing robust filter banks, so that the negative influence of imbalance source training set can be eliminated. The second one is an L1 penalty with $\|\cdot\|_1$ denotes L1 norm (sum of absolute values). Owing to the abundant training samples of source subjects, the filter bank corresponding to the excellence performance on them is assumed to capture generic common discriminative features among subjects. This means it is comparatively robust. An important property of the L1 penalty is that it can generate zero coefficients. In fact, β reveals the complexity of the model and minimizing its L1 norm means the complexity of the model is penalized. In our STF method, we can select the spatial filters corresponding to the non-zero coefficients of β_{rj} to construct the robust filter bank Φ_{rj}.

4.2.2. Adaptive Filter bank

Moreover, we also need to construct spatial filter banks adaptive to the target subject. In our STF method, "adaptive to a subject" means "subject-specific". Therefore, the loss function for learning adptive filter bank Φ_{aj} should be related to the classification ability on the training set of the target subject, and thus the optimization problem would be to minimize:

$$\sum_{i=1}^{n_{K+1}}(y_i^{S_{K+1}} - \beta_{aj}^\top \widetilde{x}_i^{S_{K+1}})^2 + \lambda_{aj}\|\beta_{aj}\|_1. \tag{7}$$

Analogously, the adaptive filter bank can be constructed by the filters corresponding to the non-zero coefficients of β_{aj}.

The relation of λ and the number of spatial filters to be selected should be noted: In principle, if we select the value of λ properly, we can obtain a filter bank containing any possible number of the spatial filters. Briefly speaking, the larger λ is, the more zero elements of β would have. When $\lambda \to 0$, more features will be selected. However, since the corresponding classifier would be bad complex, it may have unsatisfactory prediction and be less interpretable. When $\lambda \to +\infty$, fewer features will be selected. The case of $\lambda = +\infty$ corresponds to the simplest classifier where no input variable is used for classification. As a result, if we select the interval of λ values properly, we can obtain a filter bank containing any possible number of the spatial filters. The optimal number of spatial filters in one bank can be determined by the cross-validation technology.

5. Classification Stage: Ensemble Strategy of Multiple Models

Given robust filter banks and adaptive filter banks of all subjects, we can learn equal amounts of models (bank–classifier pairs). We denote the model obtained by the robust (adaptive) filter bank of subject j ($j = 1 \cdots K+1$) as M_{rj} (M_{rj}). It is trained with some classification strategy based on the training set of subject j projected by the robust (adaptive) filter bank Φ_{rj} (Φ_{aj}). Then, we should employ a ensemble strategy to combine their outcomes into a single one. However, most of existing model weighting approaches always assign static weights to models, which are either uniform (e.g., in Bagging [18]), or proportional to the training accuracy (e.g., in Boosting [19]), or fixed by favoring certain models (e.g., in single-model classification). Such a static weighting scheme may not perform well for subject transfer framework since different test examples may favor predictions from different base models, which may be caused by the time-varying property of target subject's brain patterns [20]. Here, we employ a dynamic ensemble strategy to assign different weights for distinct test samples. Specifically, weights are defined with source samples surrounding the given test sample, so that the weights can estimate similarities between source subjects and the target subject. The ensemble strategy can be decomposed to two levels: The first level constructs the robust ensemble learner and adaptive ensemble learner, whose weights are defined dynamically and locally. The second level combines two learners to a single learner to complement each other.

Given robust models corresponding to the robust filter banks of all subjects, we show the construction of the robust ensemble learner. When the test sample x_i is to be classified, the robust models of all subject M_{rj} ($j = 1 \cdots K + 1$) make up a robust ensemble learner:

$$RE(x_i) = \sum_{j}^{K+1} W_{rj} \times M_{rj}(x_i), \tag{8}$$

where $RE(x_i)$ denotes the robust ensemble result of test sample x_i, $M_{rj}(x_i)$ is the result of the robust model of subject i and W_{rj} is the weight of the model M_{rj}. Our weight assignment determines W_{rj} by firstly mapping the test sample and subject j's samples into the space projected by the robust bank of subject j, and then weighting each model locally according to its prediction consistency with the neighborhood structure of the test example surrounded by subject j's samples. We define another two weights for determining W_{rj}:

$$W1_{rj} = \frac{N_j(x_i)}{n_j}, \tag{9}$$

$$W2_{rj} = \frac{N_j^+(x_i)}{N_j(x_i)}, \tag{10}$$

where $N_j(x_i)$ is the number of neighborhoods of x_i, and $N_j^+(x_i)$ represent the maximum number of samples belonging to the same class. Here the neighborhoods of x_i can be defined as:

$$\{x \big| \, \|x - x_i\|_F < d, x \in X_{sj}\}, \tag{11}$$

where d is a parameter that controls the local range and $\| \cdot \|_F$ denotes the Frobenius norm (e.g. F=2).

Given $W1_{rj}$ and $W2_{rj}$, we define W_{rj} as follows:

$$W_{rj} = W1_{rj} \times W2_{rj}. \tag{12}$$

This expression is defined owing to the assumption that if a test sample plots more close with the samples of a source subject, and meanwhile, if most of the samples around it belong to the same class, the prediction of the trained model of the subject is more dependable. Note that, in training sessions of BCI systems, we can control the sample number of each class, so that the classes are nearly balance. However, for other application domains, the class imbalance problem may appear and the effectiveness of our method may be eliminated: Let's assume, one of the source training sets contains a very biased number of class labels and the test sample is at a location which most of the training samples are of that subject is inside the neighborhood. This may happen since d is likely to be larger than half the distance between cluster means. Then the filter will be skewed towards the biased labeled subject, even if the class label of the test sample is different than the biased label. This might cause undesired decision. Modifying our method to fit the class imbalance problem is a big challenge that is worth researching. Considering this negative situation would hardly happen, we do not provide the solution to this problem, but put it on our future work plan.

Then, we use the norm term of the above calculated weights to construct the robust ensemble learner. It should be noted that other methods for combining $W1_{rj}$ and $W2_{rj}$ are also possible, which is a challenge of our future work.

Analogously, we can get our adaptive ensemble learner:

$$AE(x_i) = \sum_{j}^{K+1} W_{aj} \times M_{aj}(x_i), \tag{13}$$

where $AE(x_i)$ denotes the adaptive ensemble result of test sample x_i and $M_{aj}(x_i)$ is the result of the adaptive model of Subject j. Its optimal model weights are also computing by firstly mapping and then measuring the models' prediction consistency of a given test sample as for robust ensemble learner.

Then, we use a parameter $\alpha \in [0, 1]$, which can be determined by cross-validation, to control the balance between robustness and adaptiveness of the final ensemble learner:

$$E(x_i) = (1 - \alpha)RE(x_i) + \alpha AE(x_i). \tag{14}$$

Finally, the ensemble learner E can dynamically classify brain signals of the target subject in the test session.

6. Experiments

6.1. Data Description

The EEG data used in this study were made available by Dr. Allen Osman of University of Pennsylvania during the NIPS 2001 BCI workshop [21]. There were a total of nine subjects denoted $S_1, S_2, \cdots, S_9$, respectively. For each subject, the task was to imagine moving his or her left or right index finger in response to a highly predictable visual cue. EEG signals were recorded with 59 electrodes mounted according to the international 10–20 system and the sampling rate is 100 HZ. A total of 180 trials were recorded for each subject. Ninety trials with half labeled left and the other half right were used for training, and the other 90 trials were for testing. Each trial lasted six seconds with two important cues. The preparation cue appeared at 3.75 s indicating which hand movement should be imagined, and the execution cue appeared at 5.0 s indicating it was time to carry out the assigned response.

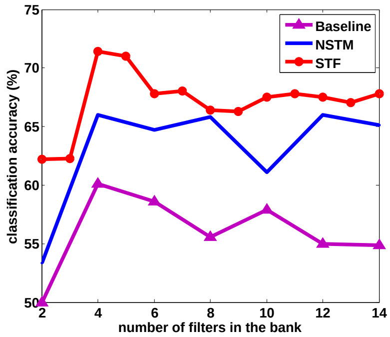
Signals from 15 electrodes over the sensorimotor area are used in this paper, so that the raw dimension of the energy feature space is 15. Moreover, for each trial, the time window from 4.0 s to 6.0 s is retained for analysis. Other preprocessing operations include common average reference [22], 8–30 Hz bandpass filtering, and signal normalization to eliminate the energy variation of different recording instants [16]. Moreover, we reduce the training set of each target subject by randomly extracting 20 samples from the original set to simulate the small calibration.

6.2. Experimental Setup and Result Analysis

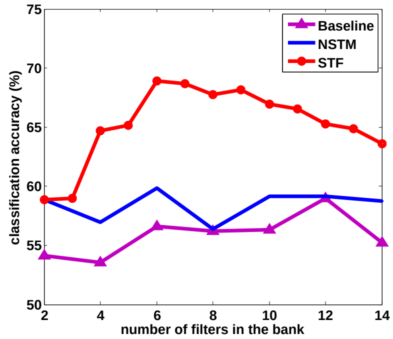
Here, we employ linear discriminant analysis (LDA) as the classification method in our experiment. This technique has been used with success in a number of BCI systems such as motor imagery based BCIs [23], P300 spellers [24], multiclass or asynchronous BCIs [25, 26]. Moreover, LDA algorithm is simple and has a very low computational requirement. Thus, it generally provides good generalization ability. In our subject transfer framework, after the construction of filter banks, LDA classifiers are trained with the feature expression of filtered training samples. Then, they give predictions of test samples as the outputs of the models. The final predictions for test samples are obtained by weighting model outputs. The parameters α are determined by 10-fold cross-validation on the training set and its value is selected among the sets $[0.1, 0.2, \cdots, 1]$ and d is defined as the average distance between samples in the training set of the corresponding subject. Moreover, experimental procedure was repeated 10 times and the averaged results are reported.

We compare the performance of our subject transfer framework (STF) against two methods. First, we employ the traditional EEG classification method as a baseline. It only uses target training samples to perform test tasks for the target user (baseline). Second, the result of a naive subject transfer method (NSTM) is also reported. It merges source subject training sets and the training set of the target subject as a whole training set. Almost all figures demonstrate that our new framework presents the best performance in these three methods. Some other indications are observed. For example, the NSTM may result negative transfer with worse performance than the baseline method (as S_2, S_4 and S_5 indicate). This fact illuminates the necessity of studying transfer learning methods. Moreover, when the NSTM achieves positive results, our method can further enlarge the performance improvement (as S_1, S_8 and S_9 depict). For statistical analysis of experimental results, we compare methods using t-test and find our method can remarkably get better performance (the accepted level satisfies $P < 0.05$).

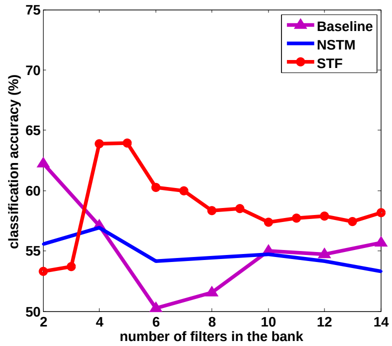
For parameter analysis, we report best α values selected by cross-validation in experiment iterations. As Fig. 5 shows, more than 70% of α values are between 0 and 0.5. This result illustrates that when training samples of the target sample are very few, the robust ensemble learner is more important than the adaptive one. It also demonstrates the benefits of using source data and the effectiveness of our subject transfer framework. Moreover, in the previous analysis, we assume filter numbers of filter banks are the same. However, in real applications of STF, they can be different. Table 1 shows the results of STF where filter numbers of filter banks can be different, and each of them is



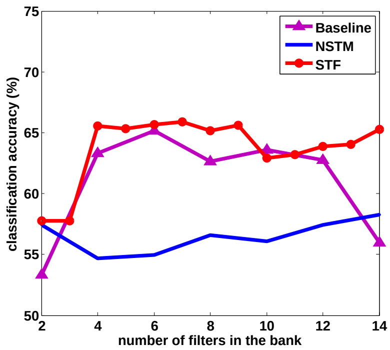
(a) S_1.



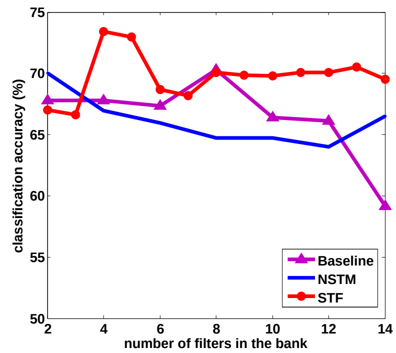
(b) S_2.



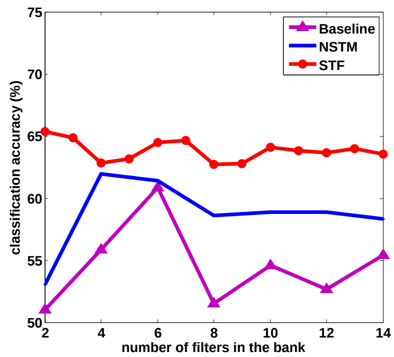
(c) S_3.



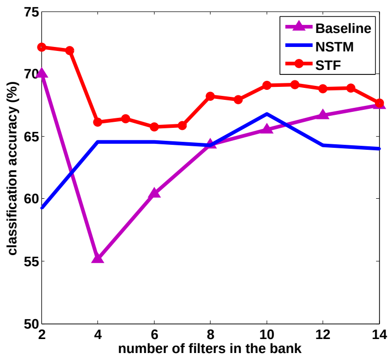
(d) S_4.



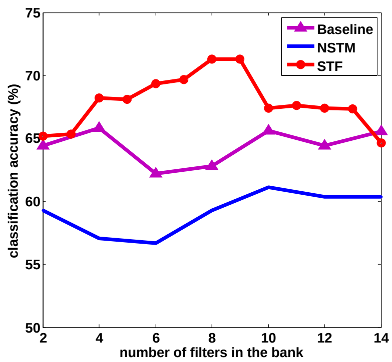
(e) S_5.



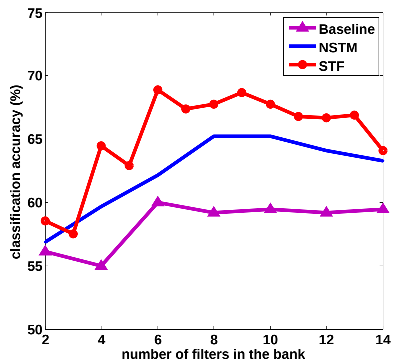
(f) S_6.



(g) S_7.



(h) S_8.



(i) S_9.

Figure 4: The performances of three methods on the data sets from nine subjects.

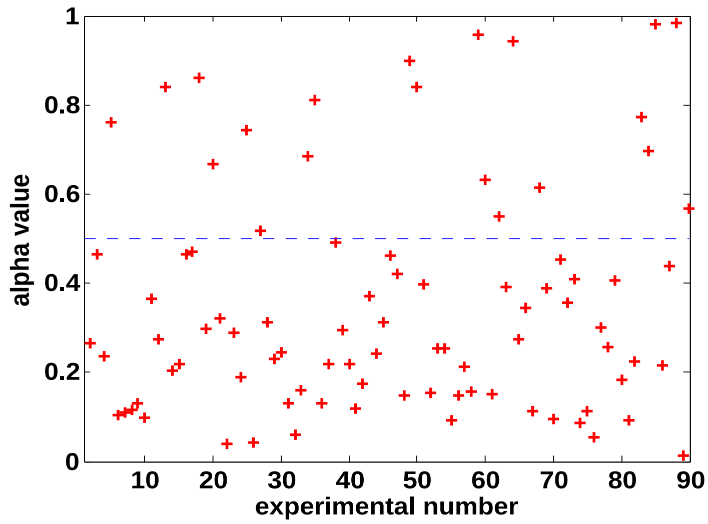


Figure 5: α value analysis. The x axis denotes the experimental number, and the y axis denotes the corresponding alpha values

Table 1: The classification accuracies (%) by using three methods, where filter numbers of models in the STF method can be different.

Method	Subject								
	1	2	3	4	5	6	7	8	9
baseline	58.73	56.28	55.31	63.24	67.25	53.72	65.11	58.54	57.93
	(±3.2)	(±2.9)	(±2.2)	(±2.7)	(±2.2)	(±3.6)	(±2.1)	(±2.4)	(±2.8)
NSTM	62.43	57.92	54.77	55.31	65.82	58.96	63.49	62.87	62.76
	(±2.6)	(±2.1)	(±1.3)	(±1.2)	(±1.3)	(±2.6)	(±2.3)	(±2.3)	(±3.1)
STF	**68.52**	**67.35**	**60.36**	**65.94**	**72.49**	**65.83**	**70.33**	**71.64**	**70.44**
	(±1.9)	(±1.7)	(±2.1)	(±1.4)	(±1.2)	(±0.9)	(±1.7)	(±2.8)	(±2.7)

selected based on the corresponding source training samples. The best performance of the STF method demonstrates its effectiveness compared to other methods. However, since these filter numbers are determined by cross-validation, the computational burden is an important problem that is worth studying in the future.

7. Conclusion and Future Work

This paper proposed a subject transfer EEG classification framework with improvements focused on feature extraction and classification. At the feature extraction stage, we employ a sparse approach to improve the construction of spatial filter banks. The previous method EER is only used to obtain the candidate filter set. In order to achieve positive transfer, we design different criterions with the 1-norm constraint to learn robust and adaptive banks. At the classification stage, a dynamic weight assignment strategy with considerations on the local structure of the given test samples is employed to learn the final ensemble learner. Experimental results demonstrated our subject transfer framework can perform positive transfer learning that results in improvements on EEG classification.

Although our framework in this paper is designed for EEG classification, it is promising to be a generic transfer learning method applicable to many other real-world domains, such as web-document classification, natural language processing or image classification. One of our future works is thus to extend it to a general method in transfer learning [27] that can be used in many areas. Moreover, generalizing the framework in this paper to cope with nonlinear feature extraction by means of the kernel trick [28, 29] is also one of our future research directions.

Acknowledgments.

This work is supported in part by the National Natural Science Foundation of China under Project 61075005, and the Fundamental Research Funds for the Central Universities.

References

[1] T. Vaughan, Guest editorial brain-computer interface technology: a review of the second international meeting, IEEE Transactions on Neural Systems and Rehabilitation Engineering 11 (2) (2003) 94–109.

[2] J. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller, T. Vaughan, Brain-computer interfaces for communication and control, Clinical Neurophysiology 113 (6) (2002) 767–791.

[3] A. Jain, R. Duin, J. Mao, Statistical pattern recognition: A review, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (1) (2000) 4–37.

[4] S. Sun, Extreme energy difference for feature extraction of EEG signals, Expert Systems with Applications 37 (6) (2010) 4350–4357.

[5] W. Tu, S. Sun, Semi-supervised feature extraction with local temporal regularization for EEG Classificationn., in: Proceedings of the 21th International Joint Conference on Neural Networks, 2011.

[6] M. Krauledat, M. Tangermann, B. Blankertz, K. Müller, Towards zero training for brain-computer interfacing, PLoS One 3 (8) (2008) 2967.

[7] S. Sun, Y. Lu, Y. Chen, The stochastic approximation method for adaptive Bayesian classifiers: towards online brain–computer interfaces, Neural Computing and Applications 20 (1) (2011) 31–40.

[8] P. Sajda, E. Pohlmeyer, J. Wang, L. Parra, C. Christoforou, J. Dmochowski, B. Hanna, C. Bahlmann, M. Singh, S. Chang, In a blink of an eye and a switch of a transistor: cortically coupled computer vision, Proceedings of the IEEE 98 (3) (2010) 462–478.

[9] Y. Huang, D. Erdogmus, M. Pavel, S. Mathan, I. Hild, E. Kenneth, A framework for rapid visual image search using single-trial brain evoked responses, Neurocomputing 74 (2011) 2041–2051.

[10] X. Yong, R. Ward, G. Birch, Robust common spatial patterns for EEG signal preprocessing, in: Proceeding of the 30th International Conference of the IEEE on Engineering in Medicine and Biology Society, IEEE, 2008, pp. 2087–2090.

[11] W. Tu, S. Sun, Spatial filter selection with LASSO for EEG classification, Lecture Notes in Computer Science 6441 (2) (2010) 142–149.

[12] H. Ramoser, J. Muller-Gerking, G. Pfurtscheller, Optimal spatial filtering of single trial EEG during imagined hand movement, IEEE Transactions on Rehabilitation Engineering 8 (4) (2000) 441–446.

[13] K. Ang, Z. Chin, H. Zhang, C. Guan, Filter bank common spatial pattern (FBCSP) in brain-computer interface, in: Proceedings of the 18th International Joint Conference on Neural Networks, IEEE, 2008, pp. 2390–2397.

[14] S. Sun, The extreme energy ratio criterion for EEG feature extraction, Lecture Notes in Computer Science 5164 (2008) 919–928.

[15] N. Hill, T. Lal, M. Schröder, T. Hinterberger, G. Widman, C. Elger, B. Schölkopf, N. Birbaumer, Classifying event-related desynchronization in EEG, ECoG and MEG signals, Pattern Recognition (2006) 404–413.

[16] J. Müller-Gerking, G. Pfurtscheller, H. Flyvbjerg, Designing optimal spatial filters for single-trial EEG classification in a movement task, Clinical Neurophysiology 110 (5) (1999) 787–798.

[17] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, SIAM Review 43 (1) (2001) 129–159.

[18] L. Breiman, Bagging predictors, Machine Learning 24 (2) (1996) 123–140.

[19] Y. Freund, R. Schapire, A desicion-theoretic generalization of on-line learning and an application to boosting, Lecture Notes in Computer Science 904 (1995) 23–37.

[20] A. Delorme, S. Makeig, M. Fabre-Thorpe, T. Sejnowski, From single-trial EEG to brain area dynamics, Neurocomputing 44 (2002) 1057–1064.

[21] P. Sajda, A. Gerson, K. Muller, B. Blankertz, L. Parra, A data analysis competition to evaluate machine learning algorithms for use in brain-computer interfaces., IEEE Transactions on Neural Systems and Rehabilitation Engineering 11 (2) (2003) 184–185.

[22] P. Nunez, R. Srinivasan, A. Westdorp, R. Wijesinghe, D. Tucker, R. Silberstein, P. Cadusch, EEG coherency:: I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales, Electroencephalography and Clinical Neurophysiology 103 (5) (1997) 499–515.

[23] G. Pfurtscheller, F. Lopes da Silva, Event-related EEG/MEG synchronization and desynchronization: basic principles, Clinical Neurophysiology 110 (11) (1999) 1842–1857.

[24] V. Bostanov, BCI competition 2003-data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram, IEEE Transactions on Biomedical Engineering 51 (6) (2004) 1057–1061.

[25] D. Garrett, D. Peterson, C. Anderson, M. Thaut, Comparison of linear, nonlinear, and feature selection methods for EEG signal classification, IEEE Transactions on Neural Systems and Rehabilitation Engineering 11 (2) (2003) 141–144.

[26] R. Scherer, G. Muller, C. Neuper, B. Graimann, G. Pfurtscheller, An asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate, IEEE Transactions on Biomedical Engineering 51 (6) (2004) 979–984.

[27] S. Pan, Q. Yang, A survey on transfer learning, IEEE Transactions on Knowledge and Data Engineering 22 (10) (2009) 1041–4347.

[28] S. Sun, C. Zhang, An optimal kernel feature extractor and its application to EEG signal classification, Neurocomputing 69 (13-15) (2006) 1743–1748.

[29] B. Nasihatkon, R. Boostani, M. Jahromi, An efficient hybrid linear and kernel CSP approach for EEG feature extraction, Neurocomputing 73 (1-3) (2009) 432–437.