Pattern Recognition Letters xxx (2009) xxx-xxx

Contents lists available at ScienceDirect



Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Regularized tensor discriminant analysis for single trial EEG classification in BCI

Jie Li, Liqing Zhang*

Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, PR China

ARTICLE INFO

Article history: Received 3 May 2008 Received in revised form 19 October 2009 Available online xxxx Communicated by R.C. Guido

Keywords: Brain computer interfacing (BCI) Channel selection Electroencephalogram (EEG) Regularized tensor discriminant analysis (RTDA) Single trial classification Tensor factorization

ABSTRACT

In this paper, a tensor-based scheme is introduced for single trial electroencephalogram (EEG) classification in brain computer interfacing (BCI). Firstly, EEG signals are represented as third order tensors in the spatial-spectral-temporal domain by wavelet transform. Then, a regularized tensor discriminant analysis (RTDA) algorithm is proposed for a multi-way discriminative subspace extraction from tensor-represented EEG data. Unlike the conventional wavelet transform method, the proposed scheme includes the structural information in multi-channel time-varying EEG spectrums endorsed by tensor representation, and improves the performance for EEG classification. Compared with the common spatial pattern (CSP, the most successful algorithm in BCI) in the applications to two classes of datasets, the proposed scheme has the following advantages: (1) an optimal multi-way discriminative subspace can be extracted, obtaining significant spatial-spectral-temporal patterns for EEG classification; (2) the proposed scheme can identify discriminative characteristics robustly, and works well without prior neurophysiologic knowledge. This is a valuable property for developing new paradigms in BCI whose discriminative neural correlates are not known and (3) the proposed scheme is able to find the most significant channels for classification, and can be applied to channel selection in BCI. Computer simulations show that the number of used channels can be reduced to 2 in two datasets with very little loss in performance. Therefore, it has great potential for the practical application of BCI.

© 2009 Elsevier B.V. All rights reserved.

瘤

Pattern Recognition Letters

1. Introduction

Brain computer interface (BCI) is a system that is designed to translate a subject's brain activities into sequences of commands for the computer. BCI provides a potentially powerful new communication channel for people to mentally control machines, and is valuable for those with severe motor disabilities. The most popular sensory signal used for BCI is the scalp-recorded electroencephalogram (EEG), because it is a noninvasive measurement of brain electrical activities and has a temporal resolution of milliseconds (Millán et al., 2004; Wolpaw et al., 2002). The fundamental of EEG based BCI is to detect changes of brain electrical activities in different mental states and utilize classification of EEG signals to transmit information (Blankertz et al., 2002; Wolpaw et al., 2000; Lotte et al., 2007). Therefore, the effectiveness and efficiency of the BCI based communication critically depend on extracting discriminative features from the recorded single trial EEG in different mental states.

The most commonly used mental control strategy in BCI is the motor imagery (Grosse-Wentrup et al., 2007; Guger et al., 2000; Song et al., 2006b; Song et al., 2006a), because it can be associated

with an enhancement (event-related synchronization, ERS) and suppression (event-related desynchronization, ERD) of brain rhythmic activity within specific frequency band over centro-parietal lobes (Pfurtscheller and Neuper, 1997). The most successful algorithm in this context, evidenced by the 2003 BCI Competition (Blanchard and Blankertz, 2004), is termed as common spatial patterns (CSP) (Ramoser et al., 2000). CSP is a decomposition method proposed by Koles (1991) that finds a set of spatial patterns that simultaneously diagonalize the covariance matrices corresponding to two classes of data, and then the eigenvectors with the largest and smallest eigenvalues correspond to the patterns with the maximum ratio of power between the two classes. For the classification of two types of motor imageries, CSP can achieve the accuracy above 90% on single trial EEG measurements (Ramoser et al., 2000).

Although the CSP algorithm proves to be highly successful for the motor imagery paradigm, it is not optimized for the EEG classification problem. There are three major drawbacks. Firstly, the selection of "good" CSP spatial patterns is done somewhat arbitrarily. A widely used heuristic is to choose several generalized eigenvectors from both ends of the spectrum. CSP is rather a decomposition technique than a classification technique (Koles, 1991), it is often observed that patterns corresponding to overwhelming strong power come to the top of the eigenvalue spectrum though they are not correlated to the class label so strongly (Tomioka et al., 2007). In fact, we cannot obtain an optimal pattern

^{*} Corresponding author. Tel./fax: +86 21 34204423.

E-mail addresses: lijie1216@sjtu.edu.cn (J. Li), zhang-lq@cs.sjtu.edu.cn (L. Zhang).

^{0167-8655/\$ -} see front matter \odot 2009 Elsevier B.V. All rights reserved. doi:10.1016/j.patrec.2009.11.012

2

for discrimination and the best classification accuracy of CSP is usually based on a combination of several patterns. Secondly, the performance of CSP severely depends on the preprocessing procedure of the temporal filtering, because CSP detects the changes of rhythmic activities based on the variances of signals. Only having the EEG signals bandpass filtered through the frequency domain of interest, high or low signal variances could reflect a strong or weak (attenuated) rhythmic activity respectively (Lemm et al., 2005). For improvements of the information rate and robustness of BCI, developing new paradigms in BCI is necessary (Lal et al., 2004). However, if the spectral characteristics in new paradigms are not well known in advance as in the motor imagery task, CSP does not work well. Finally, simultaneous diagonalization of covariance matrices makes CSP prone to overfitting, which is a typical problem especially when the number of channels is large, and when the number of available trials is small (Farguhar et al., 2006). In the extracted patterns of CSP, the most significant channels for classification are often masked by the channels unrelated to discrimination. Additionally, this drawback makes CSP require a large number of channels to be applied, whereas the practical BCI is expected to be implemented with a small number of channels. Therefore, it is demanding to design a powerful discriminative feature extraction method for single trial EEG classification.

In this paper, a tensor-based scheme is introduced for single trial EEG classification. EEG signals are represented as third order tensors in the spatial-spectral-temporal domain by wavelet transform, and then a regularized tensor discriminant analysis (RTDA) algorithm is proposed for a multi-way discriminative subspace extraction from tensor-represented EEG data. Different from other tensor models applied in the analysis of EEG signals recently, e.g., PARAFAC (Harshman, 1970; Martinez-Montes et al., 2004), Tucker model, and non-negative multi-way/ tensor factorization (NMWF/NTF) (Mørup et al., 2006a; Mørup et al., 2006; Lee et al., 2007), RTDA uses the class label information and certain regularity constraints on the EEG signal factorization in order to find a robust discriminative subspace. The high data dimensionality of EEG often lead to overfitting in discriminative feature extraction. In order

to improve generalization capability, RTDA takes the reasonable assumptions about EEG signals to impose regularization terms on each mode of the discriminative subspace: Sparseness terms are imposed on the spatial and spectral patterns to force the discriminative features to be based on a small number of channels and frequency bands, since sparse coding can find succinct representations and is believed to be a very useful approach in the analysis of EEG signals (Li et al., 2006). And because EEG signals are acquired in continuous recording, a smoothness term is applied on the temporal pattern to ensure that neighboring samples in time do not vary drastically.

Compared with the CSP algorithm in the applications to two classes of datasets, the proposed scheme has the following advantages: (1) an optimal multi-way discriminative subspace can be extracted, obtaining significant spatial–spectral–temporal patterns for EEG classification; (2) the proposed scheme can identify discriminative characteristics robustly, and works well without prior neurophysiologic knowledge. This is a valuable property for developing new paradigms in BCI whose discriminative neural correlates are not known and (3) the proposed scheme is able to find the most significant channels for classification, and can be applied to channel selection in BCI effectively. Computer simulations show that the number of used channels can be reduced to 2 in two datasets with very little loss in performance. Therefore, it has great potential for the practical application of BCI.

2. Methods

In this section, the tensor-based scheme for single trial EEG classification is introduced. Then the tensor representation of EEG data and RTDA algorithm are briefly described.

2.1. The tensor-based scheme

As illustrated in Fig. 1, the proposed scheme mainly contains four components: first, EEG signals acquired by an ESI-128 Channel



Fig. 1. The tensor-based scheme for single trial EEG classification. The black arrows stand for the procedure during the training stage, and the red ones stand for the testing procedure. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

High-Resolution EEG/EP Systems (SynAmps2, Neuroscan) are decomposed by the wavelet transform and represented in the spatial-spectral-temporal domain as high dimensional third order tensors; Second, RTDA is developed to extract a multi-way discriminative subspace from the third order tensors in the training dataset. Third, class features are obtained by projecting the tensors into the discriminative subspace. Finally, a SVM classifier with linear kernel function is trained to predict class labels for the testing data.

2.2. Tensor representation of EEG data

In the proposed scheme, EEG signals are decomposed by the wavelet transform and represented in the spatial-spectral-temporal domain as the third order tensors, that is, the tensor representation of multi-channel time-varying EEG spectrums.

For a 2-way (*channel* × *time*) EEG epoch sample $X_{(c,t)}$ at channel c and time t, the third order (*channel* × *frequency* × *time*) tensor $\mathscr{X}_{(c,t)}$ is given by the amplitude of the convolution with a wavelet function $W_{(f,t)}$,

$$\mathscr{X}_{(c,f,t)} = \|W_{(t,f)} * X_{(c,t)}\|,\tag{1}$$

In this work, we select the complex Morlet wavelet, $\psi_{(t)} = \frac{1}{\sqrt{\pi\sigma}} exp(2i\pi\Omega t) exp(-\frac{t^2}{\sigma})$ with the center frequency $\Omega = 1$ and the bandwidth parameter $\sigma = 2$ as the wavelet mother, since it has been successfully applied in the analysis of the temporal development of the frequency of EEG signals (Mørup et al., 2006a,).

2.3. The RTDA algorithm

In this scheme, the RTDA algorithm is applied for discriminative multi-linear subspace extraction. Given a set of sample data $\Phi = \{\mathscr{X}_n, y_n\}_{n=1}^N, \mathscr{X}_n \in \mathbb{R}^{D_1 \times D_2 \times \cdots \times D_m}$ denotes the nth $(1 \leq n \leq N)$ sample (*m*th order tensor), and $y_n \in \{1, -1\}$ represents the corresponding class label. The discriminative model is proposed based on logistic regression, and the log-odds ratio of the posterior class probability is defined to be a multi-way linear function with respect to the multi-way data \mathscr{X} :

$$\log \frac{P(y=+1|\mathscr{X})}{P(y=-1|\mathscr{X})} = f(\mathscr{X},\theta) = \mathscr{X} \prod_{d=1}^{m} \times_{d} w_{d} + b,$$
(2)

where $\theta := (w_1 \cdots w_m, b)$, $w_d|_{d=1}^m$ are vectors of size D_d and $b \in R$ is the bias term. w_d is assumed to be the discriminative pattern on the *d*th mode, $\mathscr{X} \times_d w_d$ is the mode-*d* product of tensor \mathscr{X} and vector w_d , therefore, $\mathscr{X} \prod_{d=1}^m \times_d w_d$ calculates the coefficient of \mathscr{X} being projected into the discriminative subspace spanned by $w_d|_{d=1}^m$, and then the parameter *b* would be derived as a offset in the data. More details about tensor and its operation can be found in (Tao et al., 2007).

To introduce a probability density function family to model the class posterior, we use the logistic regression model. The class posterior probability is modeled as:

$$P(\mathbf{y}|\mathcal{X},\theta) = \frac{1}{1 + e^{-\mathbf{y}f(\mathcal{X},\theta)}}.$$
(3)

The negative log-likelihood of (3) is minimized with additional regularization terms, which is written as follows:

$$\min_{\substack{w_{d}|_{d=1}^{m} \in \mathbb{R}^{D_{d}}, b \in \mathbb{R} \\ n=1}} \sum_{n=1}^{N} \log(1 + e^{-y_{n}f(\mathscr{X}_{n},\theta)}) + \sum_{d=1}^{m} \lambda_{d}^{1} ||w_{d}||^{2} + \lambda_{d}^{2} |w_{d}| \\
- \lambda_{d}^{3}(w_{d}^{T}K_{d}w_{d}),$$
(4)

here, three types of regularization terms are applied to limit overfitting in learning the discriminative subspace caused by the high data dimensionality of EEG. $\lambda_d^1 ||w_d||^2$ has a shrinking effect on the optimal solution of w_d . $\lambda_d^2 ||w_d||$ is a popular regularization term acting as the sparseness constraint on w_d . $\lambda_d^3 (w_d^T K_d w_d)$ declares a prior belief of smoothness on w_d , in which K_d is a symmetric $D^d \times D^d$ matrix. Let $r_{d(i,j)}$ be a distance measure between sample *i* and *j* on the *d*th mode, $K_{d(i,j)}$ is defined as $1/r_{d(i,j)}$, so $K_{d(i,j)}$ evaluates the degree of correlation between two samples according to their distance. $\lambda_d^1, \lambda_d^2, \lambda_d^3$ are the tuning parameters for the pattern on the *d*th mode, which balance a tradeoff between matching the training data and generalizing to the testing data.

With the multi-way linear probability function and proper regularization terms, it is particularly convenient to incorporate the reasonable assumptions about EEG signals into the discriminative model. Since single trial EEG data are represented as a third order (*channel* × *frequency* × *time*) tensor in this work, the sparseness of spatial and spectral patterns are enforced by $\lambda_d^2 |w_d||_{d=1,2}$ and the temporal smoothness is imposed by $\lambda_d^3 (w_d^T K_d w_d)|_{d=3}$, i.e., $\lambda_1^1, \lambda_2^1, \lambda_3^1 > 0, \lambda_1^2, \lambda_2^2 > 0, \lambda_3^2 = 0, \lambda_3^3 > 0, \lambda_1^3, \lambda_2^3 = 0.$ $r_{3(ij)}$ is defined as the absolute time difference between the sample *i* and *j*. The specific values of those parameters can be automatically selected by the performance during the training procedure.

It is easy to see that the optimal discriminative patterns, can be obtained by arg-minimizing (4), as:

$$\begin{split} w_{d}|_{d=1}^{m} &= \arg\min_{w_{d}|_{d=1}^{m} \in \mathbb{R}^{D_{d}}, b \in \mathbb{R}} \left\{ \sum_{n=1}^{N} \log\left(1 + e^{-y_{n}(\mathcal{X}_{n} \prod_{d=1}^{m} \times_{d} w_{d} + b)}\right) \\ &+ \sum_{d=1}^{m} \lambda_{d}^{1} ||w_{d}||^{2} + \lambda_{d}^{2} |w_{d}| - \lambda_{d}^{3} (w_{d}^{T} K_{d} w_{d}) \right\}, \end{split}$$
(5)

However, the problem defined in (5) does not have a closed form solution, so the alternating projection method is used to obtain a numerical solution. Therefore, the solution to (5) is decomposed into *m* different optimization sub-problems, as follows,

$$\begin{split} w_{d}|_{d=1}^{m} &= \arg\min_{w_{d}\in R^{D_{d}}, b\in R} \left\{ \sum_{n=1}^{N} \log\left(1 + e^{-y_{n}(\mathcal{X}_{n}} \prod_{l=1,l \neq d}^{m} \times_{l} w_{l}) \times_{d} w_{d} + b} \right) \\ &+ \lambda_{d}^{1} \|w_{d}\|^{2} + \lambda_{d}^{2} |w_{d}| - \lambda_{d}^{3} (w_{d}^{T} K_{d} w_{d}) \right\}. \end{split}$$
(6)

The w_d in (6) can be solved by many approximation methods, and in this paper, the gradient descent technique is applied. Algorithm 1 describes the alternating projection optimization procedure for RTDA with pre-defined tuning parameter $\lambda_d^1, \lambda_d^2, \lambda_d^3$. The key step is step 3, which involves extracting the most discriminative pattern w_d^t on mode d and in the tth iteration using $w_k^{t-1}|_{k=1}^m$ found in the (t-1)th iteration. By iterating the step t and Algorithm 1, a solution to the optimization problem, that is, the optimal discriminative patterns on each mode would be obtained.

Algorithm 1. Regularized tensor discriminant analysis algorithm
Input : The training dataset $\Phi = \{\mathscr{X}_n, y_n\}_{n=1}^N, \mathscr{X}_n \in \mathbb{R}^{D_1 \times D_2 \times \dots \times D_m}$ denotes the <i>n</i> th $(1 \le n \le N)$ sample (<i>m</i> th order tensor), $y_n \in \{1, -1\}$ represents the corresponding class label, the tuning parameters $\lambda_d^1, \lambda_d^2, \lambda_d^3$, and the maximum number of iteration <i>c</i> .
Output : The set of discriminative patterns on each mode $\{w_d _{d=1}^m\}$.
Initialization : Set $w_d^0 = 1_{D_d} _{d=1}^m$.
Method:
1: For iteration number $t = 1$ to c {
2: For mode <i>d</i> = 1 to <i>m</i> {
3: $w_d^t = argmin_{w_d \in \mathbb{R}^{D_d}, b \in \mathbb{R}} \left\{ \sum_{n=1}^N \log \left(1 + e^{-y_n(\mathscr{X}_n \prod_{l=1, l \neq d}^m X = W_l^{l-1}) \times dw_d^t + b} \right) + \right\}$
$\lambda_{t}^{1} \ w_{t-1}^{t-1} \ ^{2} + \lambda_{t}^{2} \ w_{t}^{t} \ - \lambda_{t}^{3} (w_{t}^{t} K_{d} w_{t}^{t}) \}$

4

3. Experiment setup and data acquisition

Two datasets, collected during our BCI experiments, were used to verify the effectiveness and robustness of the proposed tensorbased scheme. Totally, six healthy male subjects, aged from 21 to 30, participated in data collection. They all gave informed consent as approved by the Ethics Committee. Sixty-two channels of EEG were recorded by an ESI-128 Channel High-Resolution EEG/EP System (SynAmps2, Neuroscan at Lab for Brian-like Computing and Machine Intelligence, Shanghai Jiao Tong University, China. EEG electrode positioning follows the 10–20 International System of Electrode Placement). In both data collection stages, each subject was asked to seat in an armchair, keeping arms on the chair arms with hands relaxing, and two eyes were requested to look at a computer monitor placed approximately 1 m in front of the subject at eyes level.

Dataset 1 were collected from four subjects in the motor imagery task, which has been extensively used in BCI systems. The subject was instructed to imagine a movement of the right or left hand for about 2s to control a cursor movement on the computer screen. EEG signals were recorded, sampled at 500 Hz and bandpass filtered between 0.1 Hz and 100 Hz. For each subject, 100 left and 100 right trials were acquired, and divided into a training dataset (60 trials, 30 trials for each class) and a testing dataset (140 trials, 70 trials for each class).

Dataset 2 were obtained from the other two subjects performing a cognitive task, i.e., the mental arithmetic. The subject was requested to look at a computer monitor, and calculate in mind when an arithmetic formula containing three integers was shown on the screen. The formula displayed for 2 s, and the subject was required to take relaxation when the formula disappeared from the screen. EEG signals were recorded, sampled at 1000 Hz and bandpass filtered between 0.1 Hz and 100 Hz. According to the change of visual cue, EEG data were segmented into two classes of epochs, i.e., mental arithmetic and resting EEG trials. At last, for each subject, 100 mental arithmetic and 100 resting trials were acquired and divided into a training dataset and a testing dataset in the same fashion as for dataset 1.

4. Data analysis and performance evaluation

In this section, the proposed tensor-based scheme is applied to the two aforementioned datasets, and the experimental results are presented respectively.

4.1. Results on dataset 1

Exemplary spectral characteristics of EEG in motor imagery tasks are α rhythm (8–13 Hz) and β rhythm (14–30 Hz) which decrease during movement or in preparation for movement and increase after movement and during relaxation, and those phenomena happen in sensorimotor area (centro-parietal lobes) (Pfurtscheller and Neuper, 1997; Ramoser et al., 2000). In details, imaging left or right hand movement causes ERD over sensorimotor area on the contralateral hemisphere, whereas ERS on the ipsilateral hemisphere, and there are also some slight differences in the most discriminative frequency bands and channels depending on individual subjects. In the subsection, the effectiveness and robustness of the proposed scheme in dataset 1 are presented, and the application to channel selection is also demonstrated. For comparison, the highly successful method in this context, CSP, is applied to the dataset. Since neural correlates for the motor imagery paradigm are well known, it would be shown that the signifi-



Fig. 2. The example of the multi-channel's 2-way EEG spectrums maps for Sub. 3 (in order to give more significant illustration, this figure shows the assemble difference between two classes of data instead of single trial data). The 62 channels are shown in the top of figure with their respective time-frequency plots from 0 to 2 s and 8 to 30 Hz, and C3 and C4 channels' 2-way maps are magnified specially in the bottom of the figure.

cant results agreeing well with the expectation could be obtained by the proposed scheme without considering prior knowledge.

It is worth mention that wavelet approach has been widely used as a feature extraction method for EEG classification, in which the multi-channel time-varying EEG spectrums have been computed as feature vectors directly (Lemm et al., 2004; Bostanov, 2004; Lotte et al., 2007). In order to demonstrate the benefit of the multi-way analysis of the proposed scheme, the conventional wavelet transform method (WT) is also applied in the dataset 1, by unfolding the tensor representation multi-channel time-varying spectrum into feature vectors.

4.1.1. Classification performance

The raw EEG data were preprocessed by a given frequency band filter (8–30 Hz, which contains all α and β rhythms related to motor imagery). Visual inspection showed that artifacts had been mostly filtered out, and then filtered signals were segmented into epochs (1 ms to +2000 ms).

The third order tensor $\mathscr{X}(channel \times frequency \times time)$ was constructed in the given spatial–spectral–temporal range (62 channel; 8–30 Hz; 1–2000 ms, step by 20 ms). In detail, by wavelet transform, 2-way EEG spectrums maps (*frequency* × *time*) were obtained from 1-way data (*time*) acquired in every channel, and then folding those multi-channel's 2-way maps yielded the third order tensor. Fig. 2 gives an example of the multi-channel's 2-way maps for Sub. 3 (in order to give more significant illustration, this figure shows the assemble difference between two classes of training data instead of single trial data). In the Fig. 2, it is obviously observed that the high absolute amplitudes with opposite directions appears on the centro-parietal region of two respective brain hemispheres in α and β rhythm, especially around C3 and C4 channel at 12–13 Hz after 400 ms. This matches the well-known findings in

Table 1

The classification accuracies of the proposed scheme, WT and CSP on dataset 1 (with data filtered by 8–30 Hz, tensors constructed in the given spatial–spectral–temporal range (62 channel; 8–30 Hz; 1–2000 ms, step by 20 ms). For CSP, the optimal number of spatial patterns is also listed.

	Subject	Sub. 1	Sub. 2	Sub. 3	Sub. 4
CSP WT RTDA	Patterns number Accuracy Accuracy Accuracy	2 91.4% 79.3% 83.6%	8 56.4% 62.9% 75.7%	4 90.0% 90.0% 94.3%	6 62.1% 55.0% 61.4%

this paradigm, and confirms that the multi-channel 2-way EEG spectrums maps calculating by wavelet transform, i.e., the third order tensors, can capture the ERD and ERS phenomena from the spatial-spectral-temporal domains effectively.

In the case of CSP, the optimal accuracies based on combinations of 2-8 spatial patterns were calculated (more patterns cannot improve the training accuracy and seem to lead to overfitting), and the specific number of patterns was decided by the performance in the training dataset. Classification results listed in the Table 1 present that the performance of the proposed scheme compares favorably against CSP. Especially for Sub. 2, the accuracy is up to 75.7%, while CSP is 56.4%. It also can be seen that the optimal classification accuracy of CSP is usually based on a combination of several patterns, while the proposed scheme can acquire high accuracies by one multi-way discriminative subspace, which greatly improves significant interpretations of EEG classification in the spatial-spectral-temporal domain. For the WT method, fisher score is further applied to select features to improve the classification performance. As listed in Table 1, although the WT method proves to be an effective method for EEG classification, it cannot achieve higher accuracies than the proposed scheme for each subject, since the structural information has been ignored in unfolding the multichannel time-varying EEG spectrums into vectors.

Fig. 3 shows the two most important spatial patterns of CSP for each subject, and Fig. 4 illustrates the spatial patterns extracted by RTDA. Note that in the patterns of CSP in Fig. 3, some channels not over or close to the sensorimotor cortex also have large absolute weights because of the CSP algorithm's tendency to overfitting, and it is difficult to understand which part of the brain generates the class relevant activity and identify the discriminative channels. In contrast with CSP, the patterns extracted by RTDA have great significance in discrimination (except for Sub. 4. For him, neither of the methods can acquire a high classification accuracy). As illustrated in Fig. 4, channels over centro-parietal on the two brain hemisphere show the highest absolute weights with opposite directions for discrimination respectively, which agrees well with the expected underlying cortical activity patterns during two classes of motor imagery tasks, e.g., for Sub. 3, C3 and C4 are recognized as the two most important channels with opposite weights, which means C3 is a focus for one task and at the same time C4 is another focus for the other task. Significant spectral and temporal patterns also can be extracted by RTDA as illustrated in Figs. 5 and 6 respectively. The discriminative spectral-temporal characteristics are demonstrated to concentrate on in α rhythm and after



J. Li, L. Zhang/Pattern Recognition Letters xxx (2009) xxx-xxx



Fig. 4. The spatial patterns extracted by RTDA for the four subjects, respectively (with tensors constructed in the given spatial-spectral-temporal range (62 channel; 8– 30 Hz; 1–2000 ms, step by 20 ms));



Fig. 5. The spectral patterns extracted by RTDA for the four subjects, respectively (with tensors constructed in the given spatial-spectral-temporal range (62 channel; 8–30 Hz; 1–2000 ms, step by 20 ms));



Fig. 6. The temporal patterns extracted by RTDA for the four subjects, respectively (with tensors constructed in the given spatial-spectral-temporal range (62 channel; 8– 30 Hz; 1–2000 ms, step by 20 ms));

Table 2

The classification accuracies of the proposed scheme, WT and CSP on dataset 1 (with data filtered by 4–45 Hz, tensors constructed in the general spatial–spectral-temporal range (62 channel; 4–45 Hz, step by 2 Hz; 1–2000 ms, step by 20 ms). For CSP, the optimal number of spatial patterns is also listed.

Subject Sub. 1	Sub. 2	Sub. 3	Sub. 4
CSP Patterns number 2	8	4	6
Accuracy 52.9%	52.9%	45.0%	51.4%
WT Accuracy 77.1%	60.0%	91.4%	52.9%
RTDA Accuracy 83.6%	75.0%	93.6%	57.1%

stimulus apparently. There are some quantitative differences among subjects, and evident β rhythm activity is also observed for Sub. 3.

4.1.2. Robustness investigation

To investigate the robustness of the proposed scheme, the feature extraction without prior neurophysiologic knowledge was also considered. The raw EEG data were just preprocessed by filtering through the general EEG wave spectral range (4–45 Hz) and the third order tensor $\mathscr{X}(channel \times frequency \times time)$ was constructed in the general spatial–spectral–temporal range (62 channel; 4– 45 Hz, step by 2 Hz; 1–2000 ms, step by 20 ms;). Classification accuracies of CSP, WT and the proposed scheme are all listed in Table 2. For all subjects, classification accuracies of CSP drop below 60%. As wavelet transform is a robust algorithm to obtain the EEG spectrums, the WT method keeps up high performance, but then with the multi-way analysis of the tensor representation of EEG spectrums, the proposed scheme always achieves higher accuracies than the WT method.

Fig. 7 shows the two most important spatial patterns for each subject extracted by CSP. Compared with the previous patterns with data filtered by 8–30 Hz in Fig. 3, they have no significance in discrimination at all. However, significant patterns can still be extracted by RTDA. The spatial patterns in Fig. 8 and the spectral patterns in Fig. 9 are highly consistent with the previous patterns as shown in Figs. 4 and 5 respectively.

Therefore, CSP achieves good performance only in classification of properly preprocessed EEG data and is ineffective for spectral properties lacking tasks (i.e., the prior neurophysiologic knowledge of those tasks is not available), however, the proposed scheme is more powerful than CSP to extract discriminative features robustly without prior neurophysiologic knowledge.

4.1.3. Application to channel selection

Although CSP is highly successful for the motor imagery paradigm, it is also known for its tendency to overfitting (Farquhar



Fig. 7. The two most important spatial patterns extracted by CSP for the four subjects, respectively (with data filtered by 4-45 Hz).



Fig. 8. The spatial patterns extracted by RTDA for the four subjects, respectively (with tensors constructed in the general spatial–spectral–temporal range (62 channel; 4–45 Hz, step by 2 Hz; 1–2000 ms, step by 20 ms));



Fig. 9. The spectral patterns extracted by RTDA for the four subjects, respectively (with tensors constructed in the general spatial–spectral–temporal range (62 channel; 4–45 Hz, step by 2 Hz; 1–2000 ms, step by 20 ms));



Fig. 10. A procedure that the extracted spatial pattern is sparsified as the sparseness is enhanced in training stage with data from Sub. 3.

et al., 2006). As illustrated in Figs. 3 and 7, the extracted patterns of CSP cannot be counted onto exactly identify the most significant channels for classification. Additionally, this drawback makes it

typically require a large number of channels to be mounted, whereas the practical application of BCI is expected to be implemented with a small number of channels.

Table 3

The classification accuracies of the proposed scheme on dataset 1 only using two selected channels' data.

Subject	Sub. 1	Sub. 2	Sub. 3	Sub. 4
Selected channels	CP3,CP4	CP1,CP4	C3,C4	C4,T8
Accuracy	83.6%	70.7%	93.6%	56.4%

The proposed scheme can be applied to channel selection in BCI effectively. In RTDA, the sparseness regularization term imposed on the spatial pattern lowers the weights of the channels little related to discrimination and heightens those of the most significant channels for classification. By searching the maximum weights in the extracted spatial pattern, the most significant channels for classification are identified automatically.

Fig. 10 illustrates a procedure that the spatial pattern is sparsified as the sparseness is enhanced in the training stage with data from Sub. 3. Finally, C3 and C4 are identified as the most significant channels. For all subjects, the most significant channels can be selected out in the spatial patterns shown in Fig. 4 or Fig. 8, and they agree well with the expected underlying cortical activity (except for Sub. 4). For each subject, we only held the two selected channels' data to calculate classification accuracies according to the proposed scheme. The results listed in Table 3 show that the optimal performance can be closely achieved.

4.2. Results on Dataset 2

Compared with the motor imagery, the cognitive task has rarely been applied in BCI community, and the spectral characteristics in this paradigm are not well known in advance as in the motor imagery task. Therefore, the simulation on dataset 2 involving the cognitive task was used to explore the efficiency of the proposed scheme in the case of lacking prior neurophysiologic knowledge.

The raw EEG data were preprocessed by the band filter in the general EEG wave spectral range (4-45 Hz), and then the EEG data were segmented into epochs (1 ms to +2000 ms). For each trial, the third order tensor \mathscr{X} (*channel* × *frequency* × *time*) was constructed in the general EEG spectral range (62 channel; 4–45 Hz, the general EEG wave frequency range, step by 2 Hz; 1–2000 ms, step by 20 ms;). Fig. 11 gives an example of the multi-channel's 2-way EEG spectrums maps for Sub. 5 (in order to give more significant illustration, this figure shows the assemble difference between two classes of training data instead of single trial data). The multi-channel 2-way maps present that the discriminative characteristics are concentrated on the low frequency especially the θ band (4–7 Hz) activity in the frontal area. For comparison, CSP was also used in those epoch data. As listed in Table 4, the proposed scheme can achieve high classification accuracies. The spatial-spectraltemporal patterns of the discriminative subspace extracted by RTDA are illustrated for each subject in Figs. 12 and 13 respectively. The results indicate that the discriminative characteristics

adie 4		
he classification accuracies of the proposed scheme	and CSP o	n dataset 2.

Subject	Sub.5	Sub.6
RTDA	85.0%	75.1%
CSP	79.3%	69.3%



т

Fig. 11. The example of the multi-channel's 2-way EEG spectrums maps for Sub.5 (in order to give more significant illustration, this figure shows the assemble difference between two classes of data instead of single trial data). The 62 channels are shown in the top of figure with their respective time–frequency plots from 0 to 2 s and 4 to 45 step by 2 Hz, and AF3 and AF4 channels' 2-way maps are magnified specially in the bottom of the figure.

J. Li, L. Zhang/Pattern Recognition Letters xxx (2009) xxx-xxx



Fig. 12. The spatial-spectral-temporal patterns extracted by RTDA for Sub. 5.



Fig. 13. The spatial-spectral-temporal patterns extracted by RTDA for Sub. 6.

are not concentrated on band of 8–30Hz over centro-parietal that is important for the motor imagery tasks, while the low frequency especially the θ band (4–7 Hz) activity in the frontal area is closely related to the discrimination. Furthermore, by the previously introduced channel selection method, the number of used channels can also be reduced to 2 and there is no loss in the classification performance.

5. Conclusion

In this paper, a tensor-based scheme is introduced for single trial EEG classification.

First, benefiting from the ability of wavelet transform to construct a time-frequency representation, the 2-way (channel× time) EEG signals can be converted into the third order $(channel \times frequency \times time)$ tensors, revealing the spatial-spectral-temporal characteristic of EEG signals directly. Then, the RTDA algorithm is proposed for a discriminative subspace extraction from tensor-represented EEG data By multi-way discriminative analysis and regularization terms incorporating reasonable assumptions about EEG signals, RTDA overcomes the difficulties in extracting class features from the EEG signal due to its low signal-to-noise ratio and high data dimensionality. Unlike the conventional wavelet transform method, the proposed scheme includes the structural information in multi-channel time-varying EEG spectrums endorsed by tensor representation, and improves the performance for EEG classification. Evaluations on two datasets confirm the effectiveness and robustness of the proposed scheme. For the motor imagery task, the proposed scheme achieves some better performance than CSP and works well without prior neurophysiologic knowledge. In the cognitive task, although the discriminative characteristics are not specifically known as the motor imagery task, the scheme still extracts the discriminative patterns and features in the general EEG wave range and acquires very high classification accuracies. It is worth mentioning that the proposed scheme can be applied to channel selection in BCI effectively. It is demonstrated that the number of used channels can be reduced to 2 in two datasets with very little loss in performance.

This study shows that the proposed tensor-based scheme is efficient for single trial EEG classification in BCI and RTDA algorithm is a promising data exploratory tool for developing BCI system.

Acknowledgements

The authors thank the editors and anonymous reviewers who have given many valuable comments. The work was supported by the Science and Technology Commission of Shanghai Municipality, China (Grant No. 08511501701) the National Natural Science Foundation of China (Grant No. 60775007) and the National Basic Research Program of China (Grant No. 2005CB724301).

References

- Blanchard, G., Blankertz, B., 2004. BCI competition 2003-data set IIa: Spatial patterns of self-controlled brain rhythm modulations. IEEE Trans. Biomed. Eng. 51 (6), 1062–1066.
- Blankertz, B., Curio, G., Müller, K.R., 2002. Classifying Single Trial EEG: Towards Brain Computer Interfacing, vol. 14. MIT Press.
- Bostanov, V., 2004. BCI competition 2003-data sets Ib and Ilb: Feature extraction from event-related brain potentials with the continuous wavelet transform and the *t*-value scalogram. IEEE Trans. Biomed. Eng. 51 (6), 1057–1061.
- Farquhar, J., Hill, N.J., Lal, T.N., Schölkopf, B., 2006. Regularised CSP for sensor selection in BCI. In: Proceedings of the Third International Brain–Computer Interface Workshop and Training Course, pp. 14–15.
- Grosse-Wentrup, M., Gramann, K., Buss, M., 2007. Adaptive Spatial Filters with predefined Region of Interest for EEG based Brain–Computer-Interfaces, vol. 19. MIT Press.
- Guger, C., Ramoser, H., Pfurtscheller, G., 2000. Real-time EEG analysis with subjectspecific spatial patterns for a brain-computer interface (BCI). IEEE Trans. Rehabil. Eng. 8 (4), 447–456.
- Harshman, R.A., 1970. Foundations of the PARAFAC Procedure: Models and Conditions for an "explanatory" Multi-modal Factor Analysis. s. n.
- Koles, Z.J., 1991. The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG. Electroencephalogr. Clin. Neurophysiol. 79 (6), 440–447.
- Lal, T.N., Schröder, M., Hinterberger, T., Weston, J., Bogdan, M., Birbaumer, N., Schölkopf, B., 2004. Support vector channel selection in BCI. IEEE Trans. Biomed. Eng. 51 (6), 1003–1010.
- Lee, H., Kim, Y.D., Cichocki, A., Choi, S., 2007. Nonnegative tensor factorization for continuous EEG classification. Int. J. Neural Systems 17 (4), 1–13.
- Lemm, S., Blankertz, B., Curio, G., Müller, K.R., 2005. Spatio-spectral filters for improving the classification of single trial EEG. IEEE Trans. Biomed. Eng. 52 (9), 1541–1548.
- Lemm, S., Schäfer, C., Curio, G., 2004. BCI competition 2003Cdata set iii: Probabilistic modeling of sensorimotor mu rhythms for classification of imaginary hand movements. IEEE Trans. Biomed. Eng. 51 (6), 1077–1080.
- Li, Y.Q., Cichocki, A., Amari, S.I., 2006. Blind estimation of channel parameters and source components for EEG signals: A sparse factorization approach. IEEE Trans. Neural Networks 17 (2), 419–431.
- Lotte, F., Congedo, M., Lecuyer, A., Lamarche, F., Arnaldi, B., 2007. A review of classification algorithms for EEG-based brain-computer interfaces. J. Neural Eng. 4, 1–13.

10

J. Li, L. Zhang/Pattern Recognition Letters xxx (2009) xxx-xxx

- Martinez-Montes, E., Valdés-Sosa, P.A., Miwakeichi, F., Goldman, R.I., Cohen, M.S., 2004. Concurrent EEG/fMRI analysis by multiway Partial Least Squares. Neuroimage 22 (3), 1023–1034.
- Millán, J.R., Renkens, F., Mourino, J., Gerstner, W., 2004. Noninvasive brain-actuated control of a mobile robot by human EEG. IEEE Trans. Biomed. Eng. 51 (6), 1026– 1033.
- Mørup, M., Hansen, L.K., Herrmann, C.S., Parnas, J., Arnfred, S.M., 2006a. Parallel Factor Analysis as an exploratory tool for wavelet transformed event-related EEG. Neuroimage 29 (3), 938–947.
- Mørup, M., Hansen, L.K., Parnas, J., Arnfred, S.M., 2006b. Decomposing the timefrequency representation of EEG using non-negative matrix and multi-way factorization. Tech. rep., Technical report, Informatics and Mathematical Modeling, Technical University of Denmark, 2006.
- Pfurtscheller, G., Neuper, C., 1997. Motor imagery activates primary sensorimotor area in humans. Neurosci. Lett. 239 (2–3), 65–68.
- Ramoser, H., Müller-Gerking, J., Pfurtscheller, G., 2000. Optimal spatial filtering of single trial EEG during imagined handmovement. IEEE Trans. Rehabil. Eng. 8 (4), 441–446.

- Song, L., Gysels, E., Gordon, E., 2006a. Phase Synchrony Rate for the Recognition of Motor Imagery in BCI, vol. 19. MIT Press.
- Song, L., Gysels, E., Gordon, E., 2006b. Phase Synchrony Rate for the Recognition of Motor Imagery in Brain–Computer Interface, vol. 18. MIT Press.
- Tao, D., Li, X., Wu, X., Maybank, S.J., 2007. General tensor discriminant analysis and gabor features for gait recognition. IEEE Trans. Pattern Anal. Machine Intell. 29 (10), 1700–1715.
- Tomioka, R., Aihara, K., Müller, K.R., 2007. Logistic Regression for Single Trial EEG Classification, vol. 19. MIT Press.
- Wolpaw, J.R., Birbaumer, N., Heetderks, W.J., McFarland, D.J., Peckham, P.H., Schalk, G., Donchin, E., Quatrano, L.A., Robinson, C.J., Vaughan, T.M., 2000. Braincomputer interface technology: A review of the first international meeting. IEEE Trans. Rehabil. Eng. 8 (2), 164–173.
- Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M., 2002. Brain-computer interfaces for communication and control. Clin. Neurophysiol. 113 (6), 767–791.