Uncorrelated Multiway Discriminant Analysis for Motor Imagery EEG Classification

Ye Liu
Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering
Department of Computer Science and Engineering
Shanghai Jiao Tong University, Shanghai, 200240, China
lyyx000@sjtu.edu.cn

Qibin Zhao
Laboratory for Advanced Brain Signal Processing
Brain Science Institute, RIKEN, Saitama, Japan
qbzhao@brain.riken.jp

Liqing Zhang*
Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering
Department of Computer Science and Engineering
Shanghai Jiao Tong University, Shanghai, 200240, China
zhang-lqcs.sjtu.edu.cn

Accepted 21 February 2015
Published Online 18 May 2015

Motor imagery-based brain–computer interfaces (BCIs) training has been proved to be an effective communication system between human brain and external devices. A practical problem in BCI-based systems is how to correctly and efficiently identify and extract subject-specific features from the blurred scalp electroencephalography (EEG) and translate those features into device commands in order to control external devices. In real BCI-based applications, we usually define frequency bands and channels configuration that related to brain activities beforehand. However, a steady configuration usually loses effects due to individual variability among different subjects in practical applications. In this study, a robust tensor-based method is proposed for a multiway discriminative subspace extraction from tensor-represented EEG data, which performs well in motor imagery EEG classification without the prior neurophysiologic knowledge like channels configuration and active frequency bands. Motor imagery EEG patterns in spatial-spectral-temporal domain are detected directly from the multidimensional EEG, which may provide insights to the underlying cortical activity patterns. Extensive experiment comparisons have been performed on a benchmark dataset from the famous BCI competition III as well as self-acquired data from healthy subjects and stroke patients. The experimental results demonstrate the superior performance of the proposed method over the contemporary methods.

Keywords: Brain computer interface (BCI); electroencephalography; classification; tensor Factorization.

1. Introduction

Brain–computer interface (BCI) provides a communication system between human brain and external devices.\textsuperscript{15,21,22,25,49} Recently, BCI technology has attracted great attention in prediction and diagnosis of neurological disorders.\textsuperscript{2,6,35,36} Among asserts of brain diffused signals, electroencephalogram (EEG) is an effective noninvasive method to be studied in BCI researches.\textsuperscript{35-36} Feature extraction and
classification are the key practical problem in BCI-based systems.\textsuperscript{5,13,16,40,44}

In BCI-related studies, common spatial patterns (CSP)\textsuperscript{38} has been demonstrated as the most effective way to extract spatial filters in terms of subject-specific EEG patterns.\textsuperscript{4} CSP is used for discriminating two classes of EEG data by maximizing the variance of one class while minimizing the variance of other class.\textsuperscript{38} However, there are mainly two issues in CSP when applied in the real BCI-based applications. One issue is that CSP method is known to be very sensitive to frequency bands related to brain activity.\textsuperscript{38} So far we usually set the frequency filter to a broad band beforehand in real applications.\textsuperscript{38} Another issue is that CSP is prone to overfitting when large number of channels are used,\textsuperscript{14} which may be solved by removing some redundant electrodes. In a word, in terms of subject-specific variability among subjects, it is highly desirable to optimize both frequency band and channels configuration.

Recently, several novel approaches, namely, common spatio-spectral pattern (CSSP),\textsuperscript{20} common sparse spectral spatial pattern (CSSSP),\textsuperscript{11} iterative spatio-spectral pattern learning (ISSPL),\textsuperscript{47} and Filter bank common spatial pattern (FB-CSP)\textsuperscript{1} have been proposed, in which the classification performance is improved by simultaneously optimizing the spatial and spectral filter. Despite various studies and recent advances, the flexibility of the spatial filters and spectral filters is still very limited. Learning optimum spatial-spectral filters is still a challenging and open issue. Moreover, most of these algorithms are developed only for healthy people, but do not evidence their effectiveness and robustness when applied on EEG collected from stroke patients who are also the potential audience of BCI.\textsuperscript{14} Some previous studies\textsuperscript{23,31,43} have proved that EEG patterns in stroke patients differ from those of healthy people in both spectral and spatial domains. In this case, these traditional algorithms cannot work well when directly applied on stroke patients.

Motivated by the problems mentioned above, in this study, we propose a novel tensor-based method for motor imagery EEG classification, which also takes correlations among multimodal features into account to find the optimal uncorrelated features using multilinear transformation. Uncorrelated features contain minimum redundancy and ensure independence of features. They are highly desirable in practical recognition tasks since they contain more information than the correlated ones in the same dimension and the subsequent classification task can be greatly simplified. Main contributions of our method are as follows:

1. Our method extends the vector-based subspace learning method to a tensor variate input space while producing uncorrelated features in order to consider the multiway structure of inputs into the model learning and predictions, which is important and promising for multidimensional structured EEG data classification when lacking of the prior knowledge like channels configuration and active frequency bands.

2. The discriminative spatial patterns and spectral patterns can be simultaneously identified from limited training dataset, which may provide insights to the underlying cortical activity patterns, e.g. for brain source localization and physiological knowledge exploration.

3. Extra virtual training data can be generated by the proposed method, thus most of the variations in the original data can be captured and the capacity of the available database is expanded. Therefore, the method is suitable for the small sample size problem and has better generalization performance for the overfitting problem.

The rest of the paper is organized as follows: a detailed formulation of the tensor-based method is illuminated in Sec. 2. Section 3 describes the experimental arrangement and data acquisition. Section 4 displays comparison results among some state-of-the-art methods and our method, followed by a conclusion in Sec. 5.

2. Method

In this section, a tensor-based method, called uncorrelated multilinear nearest feature line analysis (UMNFLA), is introduced in detail. It aims to learn the optimal multilinear subspace directly from limited multidimensional EEG, so as to minimize the within-class feature line (FL) distance and maximize the between-class FL distance in a low-dimensional space, constrained by producing uncorrelated features.
2.1. Multilinear algebra

Some tensor-based notation and algebra are firstly introduced in this paper. A tensor is a multilinear array or multidimensional matrix and the order of a tensor $\mathbf{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_M}$ is $M$. Tensors are denoted by underlined boldface capital letters, matrices by boldface capital letters and vectors by lower-case letters. The element of $(i_1, i_2, \ldots , i_M)$ of tensor $\mathbf{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_M}$ is denoted by $\mathbf{X}_{i_1, i_2, \ldots , i_M}$.

In Ref. 10, the contracted product of two tensors $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_M}$ and $\mathbf{P} \in \mathbb{R}^{J_1 \times J_2 \times \ldots \times J_M}$ along the first $M$ modes is denoted as $\mathbf{Z} = [\mathbf{X} \otimes \mathbf{Y} ](1 : M)(1 : M) \in \mathbb{R}^{I_1 \times \ldots \times I_M \times J_1 \times \ldots \times J_M}$, given by:

$$\mathbf{Z} = \mathbf{X} \otimes \mathbf{Y} = \sum_{i_1}^{I_1} \cdots \sum_{i_M}^{I_M} \mathbf{X}_{i_1 \ldots i_M} \mathbf{Y}_{i_1 \ldots i_M}(1 : M)(1 : M).$$

Especially, contracted product of $\mathbf{X}$ and $\mathbf{Y}$ on all indices except the 4th index is denoted as $[\mathbf{X} \otimes \mathbf{Y} ](1 : M)$, as described in Ref. 10.

The mode-$d$ product of a tensor $\mathbf{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_M}$ and a matrix $\mathbf{P} \in \mathbb{R}^{I_1 \times J_d}$ is denoted as a tensor $\mathbf{X} \hat{\otimes} \mathbf{P} \in \mathbb{R}^{I_2 \times \ldots \times I_M \times J_d}$, with elements

$$\mathbf{X}_{i_1 \ldots i_{d-1}, i_{d+1} \ldots i_M} \mathbf{P}_{i_d, j_d}.$$

Besides, the mode-$d$ matricization of a tensor $\mathbf{X} \in \mathbb{R}^{I_1 \times \ldots \times I_M}$ is denoted by $\mathbf{X}_{\hat{\otimes} d} \mathbf{P} \in \mathbb{R}^{I_1 \times \ldots \times I_M \times J_d}$, as detailed in the study 21.

2.2. Tensor-to-vector projection

The UMNFLA framework developed in this paper takes a multilinear subspace (or tensor subspace) approach of feature extraction, where tensorial data is projected into a subspace for better discrimination. Tensor-to-tensor projection (TTP)22,23 and tensor-to-vector projection (TVP)23 are commonly used in multilinear projection. In this paper, we choose to develop the UMNFLA method by TVP rather than TTP.

The TVP projects a tensor to a vector and it can be viewed as multiple projections from a tensor to a scalar, as illustrated in Fig. 1. A tensor $\mathbf{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_M}$ is first projected to a scalar by $\mathbf{X}_{\hat{\otimes} d} \mathbf{P} \in \mathbb{R}^{I_1 \times \ldots \times I_M \times J_d}$ through $M$ unit vectors $\{w_1, w_2, \ldots , w_M\}$ called an elementary multilinear projection (EMP), here $w_m$ is the $m$th component of EMP. Then $\mathbf{X}$ is projected to a vector by multiple projections through $K$ EMPs $\{w_1, \ldots , w_M\}$, written as $y = \sum_{k=1}^{M} \mathbf{X}_{\hat{\otimes} d} \mathbf{P} \in \mathbb{R}^K$, where the $k$th component of $y$ is obtained from the $k$th EMP as $y(k) = \mathbf{X}_{\hat{\otimes} d} \mathbf{P} \in \mathbb{R}^K$.

2.3. Uncorrelated multilinear nearest feature line analysis

Motivated by the discussions above, we aim to extend nearest feature line (NFL)-based subspace learning method to a tensor variate input space while producing uncorrelated features in order to consider the multiway structure of inputs into the model learning and predictions. As a natural extension of the traditional FL, tensor-based feature line (TFL) is described as a line passing through two multidimensional samples $\mathbf{X}_t \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_M}$ and $\mathbf{X}_s \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_M}$ with the same label, denoted as $\mathbf{X}_t \mathbf{X}_s$. We denote $\mathbf{X}_t$ as the projection of $\mathbf{X}$ onto TFL $\mathbf{X}_t \mathbf{X}_s$. The FL distance between $\mathbf{X}_t$ and $\mathbf{X}_s$ is defined as the Euclidean distance between $\mathbf{X}_t$ and $\mathbf{X}_s$.

$$d(\mathbf{X}_t \mathbf{X}_s) = \sqrt{\sum_{i_1=1}^{I_1} \cdots \sum_{i_M=1}^{I_M} (\mathbf{X}_t_{i_1 \ldots i_M} - \mathbf{X}_s_{i_1 \ldots i_M})^2},$$

where $\mathbf{X}_t = \mathbf{X}_t + (\mathbf{X}_s - \mathbf{X}_t)$ and the position parameter $t$ is

$$t = \frac{(\mathbf{X}_t - \mathbf{X}_s) \otimes (\mathbf{X}_t - \mathbf{X}_s)(1 : M)(1 : M)}{\mathbf{X}_t \mathbf{X}_s(1 : M)(1 : M)}.$$

Denote $\Omega = \{\mathbf{X}_t \mathbf{X}_s \mid \mathbf{X}_t, \mathbf{X}_s \}$ as the training dataset, here $\mathbf{X}_t$ stands for the $k$th point and $c_k$ is the corresponding label. Assume that the number of points having the same class with $\mathbf{X}_t$ is $L_{t}$, then there are $P_{t} = \frac{L_t}{4}$ TFLs formed by the prototypes having the same class with $\mathbf{X}_t$. Let the set $\{\mathbf{X}_t \mathbf{X}_s \mid \mathbf{X}_t, \mathbf{X}_s \}$ denotes all the projections of $\mathbf{X}_t$ onto the corresponding TFL. We define the within-class FL distance between $\mathbf{X}_t$ and all its projections as $\sum_{i=1}^{P_{t}} d(\mathbf{X}_t \mathbf{X}_s)$, here $\mathbf{X}_s$ denotes the $i$th projection of $\mathbf{X}_t$ onto the corresponding TFL. 1550013-3
The within-class FL distance between all the points and their projections is formulated as:

\[ J_{\text{within}} = \frac{1}{N} \sum_{n=1}^{N} \sum_{p=1}^{P_n} \beta^2(\mathbf{x}_n, \mathbf{u}_n^p). \]  

Similarly, the between-class FL distance between all the points and their projections:

\[ J_{\text{between}} = \sum_{n=1}^{N} \sum_{m=1}^{M_n} \beta^2(\mathbf{x}_n, \mathbf{u}_n^m), \]

where \( Q_n \) is the number of the projections onto all the TFLs in terms of the samples having different class with \( \mathbf{x}_n \) and \( \mathbf{u}_n^m \) is the \( m \)th projection of \( \mathbf{x}_n \).

Hence, UMNFLA method aims to minimize the within-class FL distance and maximize the between-class FL distance in a low-dimensional space while producing uncorrelated features by learning K EMPs \( \{w_k^1, w_k^2, \ldots, w_k^M\} \) \( k=1, \ldots, K \). Thus UMNFLA can be formulated without producing uncorrelated features as:

\[ w_k^m = \arg \min \frac{1}{N} \sum_{n=1}^{N} \sum_{p=1}^{P_n} \beta^2(y_n(k), y_n^m(p)) \]

\[ - \frac{1}{N Q_n} \sum_{n=1}^{N} \sum_{m=1}^{M_n} \beta^2(y_n(k), y_n^m(k)), \]

(7)

where \( y_n(k) \), which is the projection of \( \mathbf{x}_n \) by the \( K \)th EMP calculated by \( y_n(k) = \sum_{m=1}^{M} w_k^m x_n^m \).

is the \( m \)th component of feature vector \( y_n \). \( P_n \) is the number of projections of \( y_n \) onto the TFLs formed by the prototypes having the same class with \( y_n \), and \( y_n^m(k) \) stands for the \( m \)th component of the \( p \)th projection of \( y_n \) onto the corresponding TFL. Similarly, \( Q_n \) and \( y_n^m(k) \) denotes corresponding variables of the between-class FL distance.

Then let us consider the constraint that producing uncorrelated features. Assume \( X \) and \( Y \) are vector observations of the variables \( x \) and \( y \), and \( y \) are uncorrelated if \( \langle X - \bar{X}, Y - \bar{Y} \rangle = 0 \). For the convenience of formulation, the training samples are assumed to be zero-mean. Let \( h_i \in \mathbb{R}^N \) denotes the \( i \)th coordinate vector with its \( n \)th component \( h_i(n) = \sum_{n=1}^{N} w_n^m x_n^m \), \( n = 1, \ldots, N \) while \( h_j \) denotes the \( j \)th one, thus \( h_i \) and \( h_j \) are uncorrelated iff \( h_i^T h_j = 0 \), as shown in Fig. 2.

Thus, the objective function for the \( K \)th EMP, which is to determine the \( K \)th EMP that minimize the scatter ratio in Eq. (7) while producing uncorrelated features, is formulated as follows:

\[ w_k^m = \arg \min \frac{1}{N} \sum_{n=1}^{N} \sum_{p=1}^{P_n} \beta^2(y_n(k), y_n^m(p)) \]

\[ - \frac{1}{N Q_n} \sum_{n=1}^{N} \sum_{m=1}^{M_n} \beta^2(y_n(k), y_n^m(k)) \]

(8)
is formulated as follows:

$$w_m^k = \arg \min_{w_m^k} \text{tr}(w_m^{kT} A_m^k w_m^k / w_m^{lT} B_m^l w_m^l),$$

(9)

where $A_m^k$ and $B_m^l$ are defined as:

$$A_m^k = \frac{1}{N N_p} \sum_{n=1}^N \sum_{p=1}^{P_n} \text{mat}_{m,n}(\langle x_n - \bar{x}_m^k \rangle \langle x_n \rangle^T),$$

(10)

$$B_m^l = \frac{1}{N N_p} \sum_{n=1}^N \sum_{p=1}^{P_n} \text{mat}_{m,n}(\langle x_n - \bar{x}_m^l \rangle \langle x_n \rangle^T),$$

(11)

Detailed derivation can be found in some other study.\(^{43}\) From Eq. (9), it is obvious that the objective function is determined by only one parameter $w_m^k$ for each mode, which is linear and much simpler. Considering the constraint that producing uncorrelated features in the objective function defined in Eq. (8), which also relies on several parameters in one function, we again formulate it into another form that is determined only by one parameter $w_m^k$ by fixing the other $M - 1 \{w_m^d, d \neq m\}$ in the $m$th step. The constraint can be derived as follows:

$$\begin{aligned}
&b_i^T h_i = \left[ \sum_{m=1}^M \prod_{k=1}^K \sum_{n=1}^{X_m} \langle x_n \rangle^T \prod_{l=1}^L \sum_{m=1}^{X_m} w_m^{lT} \right] h_i \\
&= w_m^k \left[ \sum_{n=1}^{X_m} \langle x_n \rangle^T \sum_{m=1}^{X_m} w_m^{lT} \right] h_i \\
&= w_m^k Y_m^k h_i,
\end{aligned}$$

(12)

where $Y_m^k = \left[ \sum_{n=1}^{X_m} \langle x_n \rangle^T \sum_{m=1}^{X_m} w_m^{lT} \right]$. In a word, through the ALS strategy, the new objective function, which determines the optimal $w_m^k$ as the $m$th component of the $k$th EMP in the $k$th step, is formulated as:

$$w_m^k = \arg \min_{w_m^k} \text{tr}(w_m^{kT} A_m^k w_m^k / w_m^{lT} B_m^l w_m^l)$$

subject to: $w_m^d = 1, w_m^l Y_m^k h_l = 0, l = 1, \ldots, K - 1$.

By adding Lagrange multipliers to the original optimization function, the above objective function is transformed into another function without constraints. Finally, the optimal $w_m^k$ is obtained as the
The optimal $c$ value of $(\Omega = \{X_{s} = \{x_{i,j,k}\}, c_{s}\})$, the training dataset, where $x_{i,j,k}$ represents the $s$th sample and $c_{s}$ is the class label; $K$: the number of the EMPs; $T$: the maximum number of iteration; $\sigma$: the threshold.

Output: The optimal $K$ EMPs\{$w_{1}^{k}, w_{2}^{k}, \ldots, w_{M}^{k}\}|\mathbb{R}^{n}$.\)

Method:
1. for iteration $k = 1$ to $K$, do
2. Set $w_{1}^{(0)} = 1, m = 1, \ldots, M$.
3. for iteration $t = 1$ to $T$ do
4. for iteration $m = 1$ to $M$, do
5. Calculate $A_{m}^{k}, B_{m}^{k}, H_{k-1}$, and $R_{m}^{k}$ (if $k = 1$, then set $R_{m}^{k} = I$).
6. Obtain $w_{1}^{(t)}$ as the unit eigenvector corresponding to the smallest eigenvalue of $(A_{m}^{k})^{-1}R_{m}^{k}B_{m}^{k}$.\)
7. end for
8. break if $t = T$ or $\sum_{m=1}^{M}||w_{m}^{(t)} - w_{m}^{(t-1)}|| \leq \sigma$, set $w_{m}^{(t)} = w_{m}^{(t-1)} m = 1, \ldots, M$.
9. end for
10. end for

Algorithm 1 details the alternating steps of obtaining the $K$ EMPs in $K$ steps, by which feature vector is finally obtained by projection of the original EEG through the multiple projection strategy TVP.

3. Experimental Configuration

3.1. Data acquisition

Three different datasets (Datasets I–III), collected during the BCI-based motor imagery experiments, were assembled to verify the effectiveness and robustness of our proposed algorithm. The classification problem involved in our experiment was to classify the type of the imagination for each trial. Dataset I was recorded from our own online motor imagery experiments, where we compared the performance of UMNFLA against some state-of-the-art approaches.

The objective was to see if UMNFLA could surpass these state-of-the-art approaches in classification performance. Dataset II we used was one publicly available dataset IVa from BCI Competition III, where we benchmarked UMNFLA against the same competing approaches in the first dataset. The objective was to confirm if UMNFLA could indeed automatically find spatial-spectral filters besides the encouraging classification performance. Dataset III was collected from five stroke patients performing left and right upper limbs movement in our own designed motor imagery experiments. The objective was to assess the feasibility and robustness of UMNFLA when decoding the specific neurophysiological information of stroke patients’ EEG without prior knowledge like the active motor cortex regions and frequency bands. The three datasets are described as follows.

Dataset I was collected from eight healthy subjects performing the motor imagery task on two different scales (index finger versus arm) with our designed BCI system. The electrodes used in our experiments were C1–C6, CZ, CP1–CP6, CPZ. EEG data was collected at the sampling rate of 256Hz. In the whole experiment, each subject had to complete five sessions for each of different scales (index finger and arm). Each session was consisted of 12 trials separated by intervals of 2s and each trial lasted for 4s. EEG collected from each subject was split into two parts: The first four sessions for training and the last one for testing.

Dataset II was the famous publicly available dataset IVa from BCI competition III in which five healthy subjects (labeled ‘aa’, ‘al’, ‘av’, ‘aw’ and ‘ay’ respectively) participated in the right-hand and right-foot motor imagery experiments. We used 118
electrodes for EEG recordings, and then band-pass filter all the samples to a band 0.05–200Hz. Finally, 280 trials were recorded for each subject. According to the study, we divided all the samples into a training dataset and a test dataset, in which 168/224/84/56/28 formed the training set for the five subjects and the remaining trials were the test set. For each trial, a segment of EEG, from 500 to 2500 ms, was captured for analysis.

Dataset III was composed of EEG data recorded by 16 channels (namely, FC3, FCZ, FC4, CP1–CP6, CZ, CP3, CPZ, CP4, P3, PZ and P4) with a sampling rate of 256 Hz in five stroke patients. All the patients had to participate in the left and right upper limbs motor imagery experiments designed in a BCI-based training system for 24 times in two months. Patients performed motor imagery for five sessions and each session composed of 40 trials. We extracted a time segment starting from 0.5 to 4.5s after the visual cue. All the trials were divided into a training set with 120 trials and a testing set with 80 trials.

3.2. Data preprocessing

Raw data was first band-pass filtered into a specific band. The spectral information of healthy people were mainly concentrated in an rhythm (8–13Hz) and \( \beta \) rhythm (14–30Hz). However, we did not know any spectral information related to motor imagery of stroke patients in advance. Therefore, in this paper, we filtered EEG signals in Datasets I and II in 8–30Hz, while Dataset III was filtered in a broad band from 4 to 45Hz.

3.3. Feature extraction

For comparison, we also employed seven state-of-the-art algorithms of power spectral density (PSD),\textsuperscript{46} CSP, regularized CSP (RCSP),\textsuperscript{30} iterative two-dimensional nearest feature line (INFL),\textsuperscript{37} wavelet transform method (WT), non-negative multiway factorization (NMWF),\textsuperscript{34,35} and uncorrelated multilinear discriminant analysis (UMLDA).\textsuperscript{33} to extract discriminative features from the three datasets. For PSD, spectral features were calculated by a fast Fourier transform (length of data: 1000, 200, 1024 for Datasets I–III, respectively; sample rate: 250Hz, 100Hz, 256Hz for Datasets I–III, respectively; frequency band: 8–30 Hz, 8–30Hz, 4–45Hz for Datasets I–III, respectively). We employed CSP and RCSP to extract spatial features by using all the channels. The feature dimensionality will be detailed later. Regularization of RCSP was Weighted Tikhonov, which was similar in the study.\textsuperscript{30} Considering the other four methods WT, NMWF, UMLDA and UMLFLA, the multiway EEG data was reconstructed by a complex Morlet wavelet\textsuperscript{34} i.e. \( \phi(t) = \frac{1}{\sqrt[2]{2}} \exp(2\pi\text{i}t) \exp(-t^2/2) \) (frequencies: 8–30Hz for Datasets I and II and 4–45Hz for Dataset III; center frequency: 1; bandwidth parameter: 2). For NMWF, we first cast the multidimensional EEG data into a non-negative tensor, and then determined discriminative basis vectors which well reflected meaningful spectral characteristics. In terms of UMLDA, it aimed at maximizing the traditional Fisher discriminant criterion (FDC) defined in traditional linear discriminant analysis (LDA) in each elementary projection, while the features extracted were constrained to be uncorrelated. For the tensor-based methods, each original multiway EEG data was projected to a vector \( y_n \), and \( y_n \) was used for representation and classification instead of \( X_n \).

3.4. Feature selection

After feature extraction, we employed a Fisher score strategy\textsuperscript{29} for feature selection, as more features cannot improve the training accuracy. Fisher score is defined as

\[
\text{Fisher score} = \frac{|\mu_1 - \mu_2|^2}{\sigma_1 + \sigma_2},
\]

where \( \mu_1 \) and \( \sigma_1 \) denote the mean and variance of one class over an individual feature, while \( \mu_2 \) and \( \sigma_2 \) represent the mean and variance of the other class.

We first calculated Fisher score of each individual feature and then all the features were ranked by Fisher score in descending order. The first \( n \) ranked features with larger Fisher scores were retained as the most discriminative features for training, and other redundant ones were discarded. In this step, the most discriminative patterns were obtained from the corresponding projection matrices of the retained features. Thus Fisher score strategy was also used to select the discriminative motor imagery EEG patterns from the multilinear subspace in spatial-spectral-temporal domain.
3.5. SVM for classification

A linear support vector machine (SVM), which was proven to be successful in many applications, was used for classification. A 5-fold cross-validation was used for evaluation.

4. Results

4.1. Results on Dataset I

In the first experiment, we evaluated the classification performance of our proposed algorithm based on one self-acquired dataset (Dataset I) collected from eight normal persons. For CSP-based algorithms (CSP and RCSP), we used four patterns per class which led to eight-dimensional output signal. Feature dimensionality for the other methods was also set to eight. The classification results are shown in Table 1, in which the best classification accuracy for each subject is highlighted in bold. As can be seen, UMNFLA and UMLDA in general outperformed the other approaches. UMNFLA in particular consistently performed the best for all the six subjects, especially for S7, whose performance was relatively poor. It was also interesting to note that the AMI accuracies of all subjects were relatively higher than IFMI accuracies, indicating the motor imagery performance in AMI form was better than the performance by the IFMI form.

In terms of statistical significance, we applied a Mann–Whitney U test to test the differences in classification performance between our proposed algorithm and CSP. The results showed that the better recognition accuracy for UMNFLA was significant at the significance level of 0.05. The averaged classification accuracy by UMNFLA among all the subjects could reach 75.79 ± 5.73% for IFMI and 80.88 ± 6.92% for AMI, and outperformed CSP by about 10% in mean classification accuracy and by almost 8% in median classification accuracy.

4.2. Results on Dataset II

In the second experiment, besides the comparison between UMNFLA and the state-of-the-art methods in classification performance, we tried to confirm if UMNFLA could indeed automatically find spatial-spectral filters when applied on the widely used benchmark dataset (Dataset II).

(1) Classification Accuracy. Figure 3 gives a detailed offline classification results on the test datasets. Figure 3(a) gives a comparison in classification accuracies obtained for each subject between UMNFLA and the other seven algorithms, while Fig. 3(b) presents the boxplots with median, minimum and maximum values. We set the feature dimensionality for each method according to the training performance, since the training performance improvement cannot be achieved by more features. As can be seen, encouraging results were obtained. UMNFLA was the best all around performer, producing the best classification rates in all the five subjects, e.g., the averaged classification rate for UMNFLA is 92.29%, for UMLDA 85.69%, and for INFL 76.96%, and for CSP 84.33%. With a closer look at the UMNFLA results, it was realized that bigger improvements were achieved by subject S7 with relatively poor performances.

Table 1. Classification accuracies (%) on different scales (IFMI and AMI) obtained by all the competing algorithms and our proposed UMNFLA for each subject in Dataset I. The best classification accuracy for each subject is highlighted in bold. Abbreviations: IFMI = index finger motor imagery, AMI = arm motor imagery.

<table>
<thead>
<tr>
<th>Subject</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
<th>S6</th>
<th>S7</th>
<th>S8</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFMI</td>
<td>52.36</td>
<td>55.61</td>
<td>51.47</td>
<td>58.91</td>
<td>63.43</td>
<td>65.65</td>
<td>47.28</td>
<td>55.51</td>
</tr>
<tr>
<td>AMI</td>
<td>64.74</td>
<td>71.25</td>
<td>63.35</td>
<td>69.21</td>
<td>66.32</td>
<td>72.92</td>
<td>61.32</td>
<td>69.08</td>
</tr>
<tr>
<td>IFMI</td>
<td>64.74</td>
<td>71.25</td>
<td>63.35</td>
<td>69.21</td>
<td>66.32</td>
<td>72.92</td>
<td>61.32</td>
<td>69.08</td>
</tr>
<tr>
<td>AMI</td>
<td>64.74</td>
<td>71.25</td>
<td>63.35</td>
<td>69.21</td>
<td>66.32</td>
<td>72.92</td>
<td>61.32</td>
<td>69.08</td>
</tr>
<tr>
<td>IFMI</td>
<td>52.36</td>
<td>55.61</td>
<td>51.47</td>
<td>58.91</td>
<td>63.43</td>
<td>65.65</td>
<td>47.28</td>
<td>55.51</td>
</tr>
<tr>
<td>AMI</td>
<td>64.74</td>
<td>71.25</td>
<td>63.35</td>
<td>69.21</td>
<td>66.32</td>
<td>72.92</td>
<td>61.32</td>
<td>69.08</td>
</tr>
<tr>
<td>IFMI</td>
<td>52.36</td>
<td>55.61</td>
<td>51.47</td>
<td>58.91</td>
<td>63.43</td>
<td>65.65</td>
<td>47.28</td>
<td>55.51</td>
</tr>
<tr>
<td>AMI</td>
<td>64.74</td>
<td>71.25</td>
<td>63.35</td>
<td>69.21</td>
<td>66.32</td>
<td>72.92</td>
<td>61.32</td>
<td>69.08</td>
</tr>
<tr>
<td>IFMI</td>
<td>52.36</td>
<td>55.61</td>
<td>51.47</td>
<td>58.91</td>
<td>63.43</td>
<td>65.65</td>
<td>47.28</td>
<td>55.51</td>
</tr>
<tr>
<td>AMI</td>
<td>64.74</td>
<td>71.25</td>
<td>63.35</td>
<td>69.21</td>
<td>66.32</td>
<td>72.92</td>
<td>61.32</td>
<td>69.08</td>
</tr>
</tbody>
</table>

UMLDA 68.42 71.17 67.46 72.84 65.35 73.81 66.72 78.94 77.91 84.15 81.49 84.72 59.64 65.35 76.73 72.56 74.72

UMNFLA 74.16 78.59 71.49 76.18 71.47 76.34 71.27 78.94 82.38 89.73 85.76 92.51 71.37 72.59 78.49 82.17

If this paper was translated from English, please provide the original English text.
Uncorrelated Multiway Discriminant Analysis

Fig. 3. (a) Experimental results on the test accuracies obtained for each subject in Dataset II for all the competing algorithms PSD, CSP, RCSP, INFL, WT, NMWF, UMLDA and our proposed algorithm UMNFLA. (b) Boxplots of all the eight algorithms with median, minimum and maximum values.

(2) Spatial-Spectral Filters. Besides the improved classification performance, UMNFLA was also employed to learn the spatial-spectral filters and then visualized them in 2D graphs for observation. We took subject ‘a1’ as an example to illustrate this point. For comparison, CSP, which has been proven to be very useful to extract subject-specific, discriminative spatial filters, was utilized to learn the spatial filters. PSD was employed to obtain the spectral information. Figure 4 illustrates the spatial filters by CSP and spectral filters by PSD corresponded to right-hand movement imagery and right-foot movement imagery, respectively. Note that the most important channels were centered on the sensorimotor areas, and the exemplary spectral characteristic was α rhythm (8–13Hz) which decreased during movement and increased after movement, and those phenomena, termed event-related desynchronization (ERD) and event-related synchronization (ERS), happened in sensorimotor area. Different from the one-way PSD and the two-way CSP, spatial, spectral and temporal filters were simultaneously learned by UMNFLA directly from the structural preserved multiway EEG. Figure 5 gives an illustration of spatial and spectral filters learned by NMWF and UMNFLA, respectively. In general, these pictures showed larger coefficients of the spectral filters learned by both NMWF and UMNFLA mainly concentrated around α rhythm, which was almost similar with the results shown in Fig. 4. However, it was obvious that spatial filters learned by NMWF were chaos, with large coefficients in several more scattered channels from a neurophysiological point of view. In contrast, UMNFLA filters appeared as physiologically significant, with strong weights over the sensorimotor areas. This might be due to the fact that UMNFLA takes class information into consideration, yielding a more discriminative method.

4.3. Results on dataset III

In the above experiments, we showed quantitative evidences indicating UMNFLA could achieve an improvement in classification accuracy for normal persons whose related spatial and spectral characteristics were available. Besides, UMNFLA can learn the spatial-spectral filters from the EEG data. In this experiment, without any prior neurophysiologic knowledge like active motor cortex
regions and frequency bands, we tried to verify the feasibility and robustness of UMNFLA when decoding unknown neurophysiologic information of stroke patients' EEG. Experimental results were displayed in two aspects: (1) Classification accuracies when classifying 2-class motor imagery EEG. (2) Observations about the gradual changes of spatial-spectral patterns over time.

(1) Classification accuracy. We calculated the mean classification accuracy of one week by averaging the accuracies obtained in each day. Then we averaged the accuracies of all patients over weeks for each method, and the results are shown in Fig. 6. The classification accuracies were calculated using the leave-one-subject-out cross-validation method. The results show that UMNFLA has the highest classification accuracy among all methods.

Fig. 5. Spatial-spectral filters for subject ‘al’. Spatial pattern 1 and spectral filter 1 correspond to right-hand movement imagery. Spatial pattern 2 and spectral filter 2 correspond to right-foot movement imagery. (a) By NMWF. (b) By UMNFLA.

Fig. 6. Comparison of classification accuracies for all the patients in Dataset III between UMNFLA and the other methods when varying the feature dimensionality. (a)–(e) Patients 1–5. (f) Group mean.
Uncorrelated Multiway Discriminant Analysis

the mean accuracy of all the weeks to observe
the changes of the accuracy of each patient, which
was listed in Fig. 6. It is obvious that UMNFLA
outperformed the other competing algorithms in
classification performance. In terms of statistical sig-
nificance, the Mann–Whitney U test results showed
that the superior performance by UMNFLA was sig-
ificant. After two months, the UMNFLA accuracies
of all the patients could even reach 70%.

(2) Gradual changes of EEG patterns over time. Due
to the fact that brain plastics occurred in motor
areas of stroke patients, EEG patterns may deviate
from those of healthy people, as supported by the
study. To illustrate this point, we took Patient 1
(with lesion in right side) as an example. In order
to explore the gradual changes of EEG patterns dur-
ing rehabilitation, three days of day 1, 30 and 60
were selected for observation. We employed CSP to
extract spatial filters of stroke patients.

Figure 7 illustrates the CSP spatial filters while
Fig. 8 illustrates UMNFLA spatial-spectral filters
on day 1, 30 and 60 for Patient 1. The results
shows that CSP spatial filters scattered in several
neurophysiological irrelevant channels. In contrast,
the spatial-spectral filters obtained by UMNFLA

![Figure 7: CSP spatial patterns for Patient 1 on day 1, 30 and 60](image1)

![Figure 8: UMNFLA spatial-spectral filters of day 1, 30 and 60 for Patient 1](image2)
were clustered in the neurophysiological relevant locations. In detail, for the unaffected (left) hemisphere, the most active cortices changed little and mainly concentrated around C3. However, in the affected (right) hemisphere, the activated cortical regions were changing and shifted from relatively smaller areas to larger lobes, and finally concentrated to C4. Some similar phenomena were also reported by some other studies. In terms of spectral characteristics, active frequency band was updated from a lower band (6–12 Hz) at the beginning to a wide-ranged band (6–30 Hz) after two months. This dynamic band accentuation implied that active rhythms might be modulated during rehabilitation which was also reported in Ref. 41.

5. Conclusion

In this study, a novel tensor-based method, namely UMNFLA, was proposed for the classification of single-trial multidimensional EEG data during motor imagery. We aimed to find a subspace in which the within-class FL distances were minimized and between-class FL distances were maximized, and simultaneously extracted statistically uncorrelated features. We evaluated the proposed UMNFLA methods using three different EEG datasets collected from the healthy people and stroke patients. The experimental results showed that UMNFLA outperformed the other competing methods in classification performance. In particular, compared with the results in Datasets I and II, the classification performance difference between the widely used CSP and the proposed UMNFLA in the neuro-rehabilitation Dataset III was more salient.

Our studies show that UMNFLA is a robust and promising data exploratory tool for EEG analysis when lacking of prior neurophysiologic knowledge.

Acknowledgments

The work was supported by the National Key Basic Research Program of China (Grant No. 2015CB55004), JSPS KAKENHI (Grant No. 15K15855) and the National Natural Science Foundation of China (Grant Nos. 91220305, 61272251, 61202155). This work is also sponsored by China Scholarship Council (CSC).

References

13. L. J. Herrera, C. M. Fernadez, A. M. Mora, D. Migotina, R. Largo, A. Guillén and A. C. Rosa, Combination of heterogeneous EEG feature extraction methods and stacked sequential learning for sleep stage classification, Int. J. Neural Syst. 23(3) (2013) 1350012.


