ECAI 2014
T. Schaub et al. (Eds.)

doi:10.3233/978-1-61499-419-0-531

Uncorrelated Multilinear Nearest Feature Line Analysis

Ye Liu[1] and **Liqing Zhang**[1]

Abstract. In this paper, we propose a new subspace learning method, called uncorrelated multilinear nearest feature line analysis (UMNFLA), for the recognition of multidimensional objects, known as tensor objects. Motivated by the fact that existing nearest feature line (NFL) can effectively characterize the geometrical information of limited samples, and uncorrelated features are desirable for many pattern analysis applications since they contain minimum redundancy and ensure independence of features, we propose using the NFL metric to seek a feature subspace such that the within-class feature line (FL) distances are minimized and between-class FL distances are maximized simultaneously in the reduced subspace, and impose an uncorrelated constraint to extract statistically uncorrelated features directly from tensorial data. UMNFLA seeks a tensor-to-vector projection (TVP) that captures most of the variation in the original tensorial input, and employs sequential iterative steps based on the alternating projection method. Experimental results on the task of single trial electroencephalography (EEG) recognition suggest that UMNFLA is particularly effective in determining the low-dimensional projection space needed in such recognition tasks.

1 Introduction

Increasingly large amount of multidimensional data are being generated in many applications, such as image sequences, 3D video sequences, and neuroimaging data. Tensors (i.e., multiway arrays) are efficient representations for such massive data whose elements can be accessed with two or more indices. Recently, tensor decompositions and factorizations are emerging as promising tools for exploratory analysis of multidimensional data [7, 4]. Tensor subspace learning methods that deal with data represented as higher order tensors were shown to outperform their corresponding vector subspace methods, especially for small sample problems [21], such as multilinear PCA [14], multilinear LDA [20] and tensor subspace analysis [5].

In many practical applications, it is difficult and expensive to collect enough training samples, and we need to fully exploit the discriminant information from the available samples for feature extraction and recognition tasks. Aiming at enhancing or generalizing the representative capacity of limited training samples, Li *et al.* [9] proposed a novel method, called nearest feature line (NFL), to virtually enlarge the training set for data classification and recognition. Multiple feature points are linearly combined, yielding a linear subspace to represent the class, and the class of the nearest subspace to the query point is chosen as the final classification. This leads to the concept of nearest feature subspace (NFS). The basic idea of NFL is to use a linear model to interpolate and extrapolate each pair of prototype samples belonging to the same class to model the possible variants of the training samples. An infinite number of pseudo prototypes for each class are generated by linear interpolation. The classification task is done by selecting the minimal distance between the input and the feature lines. Owing to the excellent generalization capacity, NFL has been successfully used in many practical applications. In the existing NFL variants, the attention focused mainly on extracting statistically orthogonal features, however, they did not take the correlations among features into account. Uncorrelated features contain minimum redundancy and ensure independence of features [22, 13]. Hence, they are highly desirable in practical applications since the subsequent classification task can be greatly simplified.

Motivated by the discussions above, this paper aims to extend 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, which is important and promising for multidimensional structured data classification. To this end, we propose a novel tensor-based NFL method (tensor-NFL) to learn an optimal multilinear subspace from a limited amount of multidimensional data, so as to minimize within-class FL distance and maximize between-class FL distance in low dimensional space, constrained by producing uncorrelated features.

2 Uncorrelated Multilinear Nearest Feature Line Analysis (UMNFLA)

2.1 Multilinear Algebra

For the development to follow, we first introduce the notation adopted in this paper. A tensor is a multiway array or multidimensional matrix and the order of a tensor $\underline{\mathbf{X}} \in R^{I_1 \times I_2 \times \cdots \times I_M}$ is *M*. We use underlined boldface capital letters to denote tensors while matrices are expressed by boldface capital letters and vectors by lower-case letters. The element of $(i_1, i_2, \ldots, i_M)$ of tensor $\underline{\mathbf{X}}^{I_1 \times I_2 \times \cdots \times I_M}$ is denoted by $(\underline{\mathbf{X}})_{i_1 i_2 \ldots i_M}$, element (i, j) of a matrix $\mathbf{X}$ is denoted by $\mathbf{X}_{ij}$ and element i of a vector x is denoted by x_i.

In [4], the contracted product of two tensors $\underline{\mathbf{X}} \in R^{I_1 \times I_2 \times \cdots \times I_M \times J_1 \times J_2 \times \cdots \times J_N}$ and $\underline{\mathbf{Y}} \in R^{I_1 \times I_2 \times \cdots \times I_M \times K_1 \times K_2 \times \cdots \times K_P}$ along the first M modes is denoted as $\underline{\mathbf{Z}} = [\![\underline{\mathbf{X}} \otimes \underline{\mathbf{Y}}; (1:M)(1:M)]\!] \in R^{J_1 \times \ldots \times J_N \times K_1 \times \ldots \times K_P}$, given by:

$$[\![\underline{\mathbf{X}} \otimes \underline{\mathbf{Y}}; (1:M)(1:M)]\!] = \sum_{i_1}^{I_1} \cdots \sum_{i_M}^{I_M} (\underline{\mathbf{X}})_{i_1 \ldots i_M j_1 \ldots j_N} (\underline{\mathbf{Y}})_{i_1 \ldots i_M k_1 \ldots k_P}$$

Especially, contracted product of $\underline{\mathbf{X}}$ and $\underline{\mathbf{Y}}$ on all indices except the k-th index is denoted as $[\![\underline{\mathbf{X}} \otimes \underline{\mathbf{Y}}; (\overline{k})(\overline{k})]\!]$, as described in [4].

The mode-d product of a tensor $\underline{\mathbf{X}} \in R^{I_1 \times I_2 \times \cdots \times I_M}$ and a matrix $\mathbf{P} \in R^{J_d \times I_d}$ is denoted as a tensor $\underline{\mathbf{Y}} = \underline{\mathbf{X}} \times_d \mathbf{P} \in$

[1] 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. Corresponding author email: zhang-lq@cs.sjtu.edu.cn

$R^{I_1\times\ldots\times I_{d-1}\times J_d\times I_{d+1}\times\ldots\times I_M}$, with elements

$$(\underline{\mathbf{Y}})_{i_1\ldots i_{d-1}j_d i_{d+1}\ldots i_M} = \sum_{j_d=1}^{J_d} (\underline{\mathbf{X}})_{i_1\ldots i_M}\mathbf{P}_{j_d i_d}$$

Besides, the mode-d matricization of a tensor $\underline{\mathbf{X}}$ [4] is denoted by $mat_d(\underline{\mathbf{X}}) \in R^{I_d\times I_1\ldots I_{d-1}I_{d+1}\ldots I_M}$.

2.2 Tensor-to-Vector Projection (TVP)

Lu *et al.* [13] has proposed a definition called tensor-to-vector projection (TVP), by which a tensor $\underline{\mathbf{X}}^{I_1\times I_2\times\ldots\times I_M}$ is projected to a *vector*. $\underline{\mathbf{X}}$ is firstly projected to a scalar by $\underline{\mathbf{X}}\prod_{m=1}^{M}\times_m w^{(m)^T}$ through M unit vectors $\{w^{(1)^T}, w^{(2)^T}, \ldots, w^{(M)^T}\}$ called an elementary multilinear projection (EMP), here $w^{(m)}$ is the mth component of the EMP. Then $\underline{\mathbf{X}}$ is projected to a vector by multiple projections through K EMPs $\{w_k^{(1)^T}, \ldots, w_k^{(M)^T}\}|_{k=1}^K$, written as $y = \underline{\mathbf{X}}\prod_{m=1}^{M}\times_m w_k^{(m)^T}|_{k=1}^K \in R^K$, where the kth *component* of y is obtained from the kth EMP as $y_k = \underline{\mathbf{X}}\prod_{m=1}^{M}\times_m w_k^{(m)^T}$.

2.3 Nearest Feature Line (NFL)

In the literature [16, 9], Feature Line (FL) is defined as a line passing through two 2-dimensional samples x_1 and x_2 with the same label. As a natural extension of the traditional FL, tensor-based Feature Line (TFL) is described as a line passing through two multidimensional samples $\underline{\mathbf{X}}_1 \in R^{I_1\times I_2\times\cdots\times I_M}$ and $\underline{\mathbf{X}}_2 \in R^{I_1\times I_2\times\cdots\times I_M}$ with the same label, denoted as $\widetilde{\underline{\mathbf{X}}_1\underline{\mathbf{X}}_2}$. As illustrated in Fig. 1(a), $\underline{\mathbf{X}}^*$ represents the projection of $\underline{\mathbf{X}}$ onto TFL $\widetilde{\underline{\mathbf{X}}_1\underline{\mathbf{X}}_2}$. The feature line distance between $\underline{\mathbf{X}}$ and $\widetilde{\underline{\mathbf{X}}_1\underline{\mathbf{X}}_2}$ is defined as the Euclidean distance between $\underline{\mathbf{X}}$ and $\underline{\mathbf{X}}^*$:

$$\begin{aligned} d(\underline{\mathbf{X}}, \widetilde{\underline{\mathbf{X}}_1\underline{\mathbf{X}}_2}) &= d(\underline{\mathbf{X}}, \underline{\mathbf{X}}^*) \\ &= \sqrt{\sum_{i_1=1}^{I_1}\cdots\sum_{i_M=1}^{I_M}((\underline{\mathbf{X}})_{i_1\ldots i_m} - (\underline{\mathbf{X}})^*_{i_1\ldots i_m})^2} \end{aligned} \quad (1)$$

where $\underline{\mathbf{X}}^* = \underline{\mathbf{X}}_1 + t(\underline{\mathbf{X}}_2 - \underline{\mathbf{X}}_1)$ and the position parameter t is

$$t = \frac{[\![(\underline{\mathbf{X}} - \underline{\mathbf{X}}_1)\otimes(\underline{\mathbf{X}}_2 - \underline{\mathbf{X}}_1); (1:M)(1:M)]\!]}{[\![(\underline{\mathbf{X}}_2 - \underline{\mathbf{X}}_1)\otimes(\underline{\mathbf{X}}_2 - \underline{\mathbf{X}}_1); (1:M)(1:M)]\!]} \quad (2)$$

Suppose that we have a training dataset $\Omega = \{\underline{\mathbf{X}}_n^{I_1\times I_2\times\ldots\times I_M}, c_n\}_{n=1}^N$, $\underline{\mathbf{X}}_n$ represents the n-th point in the dataset and c_n denotes the corresponding label. Assume that the number of points which has the same class with $\underline{\mathbf{X}}_n$ is l_n, then there are $P_n = l_n(l_n-1)/2$ TFLs formed by the prototypes having the same class with $\underline{\mathbf{X}}_n$; Let the set $\{\underline{\mathbf{X}}^*_{n1}, \underline{\mathbf{X}}^*_{n2}, \ldots, \underline{\mathbf{X}}^*_{nP_n}\}$ denotes all the projections of $\underline{\mathbf{X}}_n$ onto all the TFLs. We define ***the within-class FL distance*** between $\underline{\mathbf{X}}_n$ and all its projections as $\sum_{p=1}^{P_n} d^2(\underline{\mathbf{X}}_n, \underline{\mathbf{X}}^*_{np})$, here $\underline{\mathbf{X}}^*_{np}$ denotes the pth projection of $\underline{\mathbf{X}}_n$. Furthermore, we can calculate the within-class FL distance between all the points and their projections:

$$J_{within} = \sum_{n=1}^{N}\sum_{p=1}^{P_n} d^2(\underline{\mathbf{X}}_n, \underline{\mathbf{X}}^*_{np}) \quad (3)$$

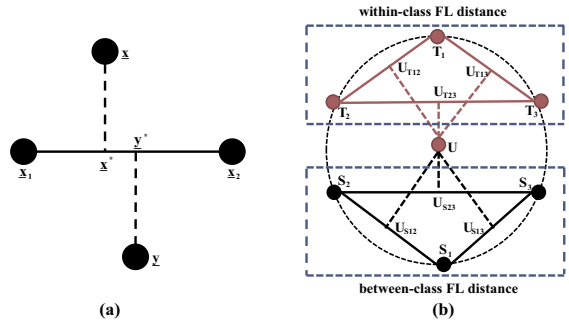

Figure 1. (a) A brief view of tensor-based feature line(TFL) $\widetilde{\underline{\mathbf{X}}_1\underline{\mathbf{X}}_2}$, which is defined as a line passing through two multidimensional samples $\underline{\mathbf{X}}_1$ and $\underline{\mathbf{X}}_2$ with the same label. $\underline{\mathbf{X}}^*$ and $\underline{\mathbf{Y}}^*$ are the projections of $\underline{\mathbf{X}}$ and $\underline{\mathbf{Y}}$. (b) Concept of within-class FL distance and between-class FL distance. Different circles represent different training data. Red circles have the same label while the black ones have different label with red ones. Within-class FL distance is defined as the sum of the distance between data U and all its projections U_{T12}, U_{T13} and U_{T23}. Between-class FL distance is defined as the sum of the distance between data U and all its projections U_{S12}, U_{S13} and U_{S23}.

Similarly, we can calculate the ***between-class FL distance*** between all the points and their projections:

$$J_{between} = \sum_{n=1}^{N}\sum_{q=1}^{Q_n} d^2(\underline{\mathbf{X}}_n, \underline{\mathbf{X}}^*_{nq}) \quad (4)$$

where Q_n denotes the number of the projections onto all the TFLs formed by the prototypes having the different class with $\underline{\mathbf{X}}_n$, and $\underline{\mathbf{X}}^*_{nq}$ represents the qth projection of $\underline{\mathbf{X}}_n$.

Fig. 1(b) shows the concept of within-class FL distance and between-class FL distance. It is obvious that $\widetilde{T_1T_2}$ provides not only the original data T_1 and T_2, but also more extra virtual variations like $U_{T_{12}}$ for training. Here we use a linear model to interpolate and extrapolate each pair of prototype samples belonging to the same class (or different class) to model the possible variants of the training samples. An infinite number of pseudo prototypes for each class are generated by linear interpolation, yielding an excellent generalization capacity of our method.

Hence, UMNFLA method aims to obtain K EMPs $\{w_k^{(1)^T}, w_k^{(2)^T}, \ldots, w_k^{(M)^T}\}|_{k=1}^K$ that project the original multiway data into a lower dimensional subspace, so as to minimize the within-class FL distance and maximize the between-class FL distance in a low dimensional space in each EMP direction, constrained by producing uncorrelated features. Note that UMNFLA employs the information of the within-class FL distance and the between-class FL distance constructed by the samples $y_n|_{n=1}^N$ in a low dimensional space which are the projections of $\underline{\mathbf{X}}_n|_{n=1}^N$ through the K EMPs. Based on the above discussion, we first formulate the proposed UMNFLA algorithm without obtaining uncorrelated features as the following optimization problem, for the kth EMP:

$$\begin{aligned} w_k^{(m)}|_{m=1}^M = \arg\min \frac{1}{NP_n}\sum_{n=1}^{N}\sum_{p=1}^{P_n} d^2(y_{nk}, y^*_{nkp}) / \\ \frac{1}{NQ_n}\sum_{n=1}^{N}\sum_{q=1}^{Q_n} d^2(y_{nk}, y^*_{nkq}) \end{aligned} \quad (5)$$

where y_{nk}, calculated by $y_{nk} = \underline{\mathbf{X}}_n \prod_{m=1}^{M} \times_m {w_k^{(m)}}^T$, is the kth scalar of the vector y_n which is the projection of $\underline{\mathbf{X}}_n$ by the K EMPs through TVP. P_n is the number of projections of y_n onto all the TFLs formed by the prototypes having the same class with y_n, and y^*_{nkp} stands for the kth component of y^*_{np} which is the pth projection of y_n onto the corresponding TFL. Similarly, Q_n and y^*_{nkq} denotes corresponding variables of the between-class FL distance.

Furthermore, the optimization problem is constrained by obtaining uncorrelated features. As Koren [8] said, suppose X and Y be vector observations of the variables x and y, x and y are uncorrelated iff $(X-\overline{x})^T(Y-\overline{y}) = 0$. For the sake of simplicity, training dataset is preprocessed to be zero-mean firstly. Let $h_i \in R^N$ denotes the ith coordinate vector with its nth component $h_i(n) = \underline{\mathbf{X}}_n \prod_{d=1}^{M} \times_d {w_i^{(d)}}^T, n = 1 \ldots N$ while $h_j \in R^N$ denotes the jth one, thus h_i and h_j are uncorrelated iff ${h_i}^T h_j = 0$.

Therefore, the core problem of the UMNFLA method – the optimization problem defined in (5) for obtaining the kth EMP with the constraint that features must be uncorrelated – is formulated as follows:

$$w_k^{(m)}|_{m=1}^{M} = \arg\min \frac{1}{NP_n} \sum_{n=1}^{N} \sum_{p=1}^{P_n} d^2(y_{nk}, y^*_{nkp}) / \frac{1}{NQ_n} \sum_{n=1}^{N} \sum_{q=1}^{Q_n} d^2(y_{nk}, y^*_{nkq}) \tag{6}$$

$$subject\ \ to: \quad {w_k^{(m)}}^T w_k^{(m)} = 1, \qquad m = 1, \ldots, M \quad and$$
$$\frac{h_k^T h_l}{\| h_k \| \, \| h_l \|} = \delta^{kl}, \quad k, l = 1, \ldots, K$$

where δ^{kl} is the Kronecker delta defined as:

$$\delta^{kl} = \begin{cases} 1, & if\ k = l \\ 0 & otherwise \end{cases}$$

However, we cannot obtain the K optimal EMPs simultaneously, as it is impossible to determine $M \times K$ parameters in one function. In this paper, a successful solution introduced in [6, 13] is applied to solve this problem. By this approach, the EMPs are obtained in K steps. First, the first EMP is obtained without any uncorrelated constraint. Then the kth ($k = 2 \ldots K$) EMP is determined in the kth step through minimizing the cost function constrained by $h_k^T h_l = 0$ for $l = 1, \ldots, k-1$. However, to obtain the kth EMP in the kth step, the problem defined in (6) is a complicated nonlinear problem without a closed form solution as we cannot determine M sets of parameters in each mode simultaneously in one function. Therefore, the alternative least square (ALS) method is derived to reduce the multilinear problem into several smaller suboptimal problems. We just calculate the optimal ${w_k^{(m)}}^*$ for the mth component of the kth EMP by fixing the other M-1 components $\left\{ {w_k^{(1)}}^*, \ldots, {w_k^{(m-1)}}^*, {w_k^{(m+1)}}^*, \ldots, {w_k^{(M)}}^* \right\}$.

To solve the mth subproblem for obtaining the optimal ${w_k^{(m)}}^*$ for the mth component of the kth EMP, the cost function defined in (6) without the constraints is derived by fixing the other $M-1$ $w_k^{(\bar{m})}$, where $\bar{m} = 1 \ldots m-1, m+1, \ldots M$, as follows

$$w_k^{(m)} = \arg\min tr\left({w_k^{(m)}}^T A_k^m w_k^{(m)} / {w_k^{(m)}}^T B_k^m w_k^{(m)} \right) \tag{7}$$

where A_k^m and B_k^m are defined as:

$$A_k^m = \frac{1}{NP_n} \sum_{n=1}^{N} \sum_{p=1}^{P_n} mat_m \left((\underline{\mathbf{X}}_n - \underline{\mathbf{X}}^*_{np}) \overline{\times}_m {w_k^{(m)}}^T \right) \times mat_m^T \left((\underline{\mathbf{X}}_n - \underline{\mathbf{X}}^*_{np}) \overline{\times}_m {w_k^{(m)}}^T \right) \tag{8}$$

$$B_k^m = \frac{1}{NQ_n} \sum_{n=1}^{N} \sum_{q=1}^{Q_n} mat_m \left((\underline{\mathbf{X}}_n - \underline{\mathbf{X}}^*_{nq}) \overline{\times}_m {w_k^{(m)}}^T \right) mat_m^T \left((\underline{\mathbf{X}}_n - \underline{\mathbf{X}}^*_{nq}) \overline{\times}_m {w_k^{(m)}}^T \right) \tag{9}$$

Similar derivation can be found in [20]. According to equation 7, we can clearly observe that the optimization function is only determined by one parameter $w_k^{(m)}$ for each subproblem. For $k = 1$, $w_1^m, m = 1, \ldots, M$ is obtained as the unit eigenvector of $(A_1^m)^{-1} B_1^m$ associated with the largest eigenvalue for a nonsingular A_1^m. For $k = 2, \ldots, K$, the optimization problem is constrained with obtaining uncorrelated features and the unit eigenvector, given the first $k-1$ optimal EMPs.

Another problem has to be pointed out: the uncorrelated constraint that $h_k^T h_l = 0$ also relies on several parameters in the kth step, thus for calculating $w_k^{(m)}$ in the kth EMP, h_k can be similarly rewritten by fixing the $\{w_k^{(1)}, \ldots, w_k^{(m-1)}, w_k^{(m+1)}, \ldots, w_k^{(M)}\}$ as :

$$\begin{aligned} h_k &= [\underline{\mathbf{X}}_1 \prod_{d=1}^{M} \times_d {w_k^{(d)}}^T, \ldots, \underline{\mathbf{X}}_N \prod_{d=1}^{M} \times_d {w_k^{(d)}}^T]^T \\ &= [\underline{\mathbf{X}}_1 \overline{\times}_m {w_k^{(m)}}^T \times_m {w_k^{(m)}}^T, \ldots, \underline{\mathbf{X}}_N \overline{\times}_m {w_k^{(m)}}^T \times_m {w_k^{(m)}}^T]^T \\ &= [\underline{\mathbf{X}}_1 \overline{\times}_m {w_k^{(m)}}^T, \ldots, \underline{\mathbf{X}}_N \overline{\times}_m {w_k^{(m)}}^T]^T w_k^{(m)} \\ &= {Y_k^m}^T w_k^{(m)} \end{aligned} \tag{10}$$

where $Y_k^m = [\underline{\mathbf{X}}_1 \overline{\times}_m {w_k^{(m)}}^T, \ldots, \underline{\mathbf{X}}_N \overline{\times}_m {w_k^{(m)}}^T]$. Therefore, the uncorrelated constraint can be written as $h_k^T h_l = {w_k^{(m)}}^T Y_k^m h_l = 0$, which also relies on one parameter w_k^m as the optimization function does.

Thus, in conclusion, the optimal $w_k^{(m)}$ as the mth component of the kth EMP in the kth step can be obtained by solving the following optimization problem:

$$w_k^{(m)} = \arg\min tr\left({w_k^{(m)}}^T A_k^m w_k^{(m)} / {w_k^{(m)}}^T B_k^m w_k^{(m)} \right) \tag{11}$$

subject to: ${w_k^{(m)}}^T w_k^{(m)} = 1, \quad {w_k^{(m)}}^T Y_k^m h_l = 0, \ \ l = 1, \ldots, k-1$

The above cost function is then transformed into another function without constraints by adding Lagrange multipliers to the original optimization function. Finally, the optimal $w_k^{(m)}$ is learned as the unit eigenvector corresponding to the largest eigenvalue of $(A_k^m)^{-1} R_k^m B_k^m$, where $R_k^m = I - Y_k^m H_{k-1} (H_{k-1}^T {Y_k^m}^T (A_k^m)^{-1} Y_k^m H_{k-1})^{-1} H_{k-1}^T {Y_k^m}^T (A_k^m)^{-1}$, and $H_{k-1} = [h_1\ h_2 \ldots\ h_{k-1}] \in R^{N \times (k-1)}$. Algorithm 1 details the alternating steps of obtaining K EMPs in K steps.

3 Experimental Configuration

To evaluate UMNFLA in terms of its predictive ability, robustness, and effectiveness under different conditions related to small number of samples and high dimensionality, a simulation study on two

EEG datasets, which were recorded from diverse populations including the healthy people and some special populations suffering from neurophysiological diseases (e.g., stroke), was undertaken. The task is to classify the type of the imagination for each trial in an offline fashion. EEG signals are multidimensional and contaminated with artifacts of Electromyography (EMG), Electrooculogram (EOG) and Electrocardiograph (ECG), resulting in a low signal-to-noise ratio. Moreover, EEG signal is not simply the time-invariant instantaneous linear mixture of neural activity in the brain, and EEG patterns across different subjects exist individual variability, yielding more difficult classification task of EEG data.

Algorithm 1 The Uncorrelated Multilinear Nearest Feature Line Analysis

Input: Training dataset $\Omega = \left\{ \underline{\mathbf{X}}_n^{I_1 \times I_2 \times \ldots \times I_M}, c_n \right\}_{n=1}^{N}$, $\underline{\mathbf{X}}_n^{I_1 \times I_2 \times \ldots \times I_M}$ denotes the n-th sample of the training dataset and c_n is the corresponding class label; the length of feature vector K; the threshold σ to test the convergence and the maximum number of iteration T.

Output: The set of the optimal K EMPs $\{w_k^{(1)}, w_k^{(2)}, \ldots, w_k^{(M)}\}|_{k=1}^{K}$.

Method:

for iteration $k = 1$ to K. Obtain the kth EMP in the kth step. **do**
  Set $w_{k(0)}^{(m)} = \mathbf{1}, \ \ m = 1 \ldots M$.
  for iteration $t = 1$ to T **do**
    for iteration $m = 1$ to M. Obtain the mth component $w_k^{(m)}$ of the kth EMP. **do**
      Calculate A_k^m, B_k^m, H_{k-1}, Y_k^m and R_k^m as defined in the above explanations (if $k = 1$, then set $R_k^m = I$).
      Obtain $w_{k(t)}^{(m)}$ as the unit eigenvector corresponding to the largest eigenvalue of $(A_k^m)^{-1} R_k^m B_k^m$.
    end for
    break if $t = T$ or $\sum_{m=1}^{M} ||w_{k(t)}^{(m)} - w_{k(t-1)}^{(m)}|| \leq \sigma$, set $w_k^{(m)} = w_{k(t)}^{(m)} \ \ m = 1 \ldots M$.
  end for
end for

3.1 Data Acquisition

Dataset I was collected from five healthy subjects (labeled 'aa', 'al', 'av', 'aw' and 'ay' respectively) performing right hand and foot motor imagery in a benchmark dataset of dataset IVa from the famous BCI competition III [2]. Two types of visual cues, a letters appearing behind a fixation cross and a randomly moving object, shown for 3.5 s were used to indicate the target class. The presentation of target cues was intermitted by periods of random length, 1.75 to 2.25 s, in which the subject could relax. The EEG signal was recorded from 118 Ag/AgCl electrodes, band-pass filtered between 0.05 and 200 Hz, and down-sampled to 100 Hz. We extracted a time segment located from 500 to 2500 ms after the cue instructing the subject to perform motor imagery. Each type of imagination was carried out 140 times. Thus 280 trials were available for each subject.

Dataset II was collected from five stroke patients performing left or right upper limbs movement in a BCI-FES rehabilitation system [10]. All the patients had to participate in BCI-FES rehabilitation training for 24 times in two months (three times per week). EEG was recorded by a 16-channel (FC3, FCZ, FC4, C1-C6, CZ, CP3, CPZ, CP4, P3, PZ and P4) g.USBamp amplifier at a sampling rate of 256 Hz. We selected 100 left and 100 right trials for each patient and divided the data into a training set (120 trials) and test set (80 trials). A time segment located from 500 to 2500 ms is used for analysis.

3.2 Data Preprocessing

Firstly, we employ FastICA to remove artifacts arising from eye and muscle movements. After that, EEG signals are digitally filtered in a specific band to contain the focused EEG spectrums. For healthy people, exemplary spectral characteristics of EEG in motor imagery tasks are α rhythm (8-13 Hz) and β rhythm (14-30 Hz) [17]. However, it is not available to obtain the spectral characteristics related to motor imagery of some special populations suffering from neurophysiological diseases (e.g., stroke) [12, 18]. Therefore, EEG signals in Dataset I are bandpass filtered between 8 and 30 Hz, which encompasses both the α and the β rhythm. Raw data in Dataset II is preprocessed by the band filter in a general range (4-45 Hz).

3.3 Feature Extraction and Classification

In order to evaluate our proposed algorithm, we apply power spectral density (PSD), common spatial pattern (CSP) [17], regularized CSP (RCSP) [11], iterative 2-dimensional nearest feature line (INFL) [16], wavelet transform method (WT) [3], nonnegative multiway factorization (NMWF) [15], uncorrelated multilinear discriminant analysis (UMLDA) [13] and our proposed UMNFLA on the two datasets for feature extraction. PSD features are calculated by a fast Fourier transform. Weighted Tikhonov Regularization is chosen for RCSP, as it reaches both the highest median and mean accuracy and has only one single hyperparameter to tune (α) [11]. Note that for the other four methods WT, NMWF, UMLDA and UMNFLA, features are extracted directly from structure preserved multiway EEG data in spatial-spectral-temporal domain reconstructed by wavelet transform method [3]. In this paper, a complex Morlet wavelet [15] is chosen as the wavelet transform method. i.e. $\phi(t) = \frac{1}{\sqrt{2\pi}} \exp(2i\pi t) \exp(-t^2/2)$ (frequencies: 8-30 Hz for Dataset I and 4-45Hz for Dataset II; center frequency: 1; bandwidth parameter: 2). For UMNFLA, σ is set to 0.01 and T is set to 50. Finally, we employ a Fisher score strategy [1] for feature selection, as more features cannot improve the training accuracy. Fisher score (a variant, $\frac{\|\mu_+ - \mu_-\|^2}{\sigma_+ + \sigma_-}$, is used in the actual computation), which measures the discriminability of individual feature for classification task, is computed for each individual feature in the feature vector. Then features with n-largest Fisher scores are retained as discriminative features.

A linear support vector machine (SVM), which achieves high-level performance in many applications, is utilized as the classifier. A 5-fold cross-validation is used to choose suitable SVM parameters to predict the labels of test datasets. In this paper, the classification threshold is set to 0.5.

4 Results

4.1 Results on Dataset I

Classification accuracy. For Dataset I, Table 1 gives a detailed offline classification results for all the competing algorithms and our proposed algorithm. The optimal feature dimensionality for each algorithm is determined according to the training performance, as more features cannot improve the training accuracy. One can clearly observe that UMNFLA yields superior recognition accuracies against

other algorithms for all the given subjects, e.g., the averaged classification rate for UMNFLA is 90.11%, for UMLDA 85.69%, and for INFL 76.96%, and for CSP 84.33%.

Table 1. Experimental results on the test accuracies (mean and standard deviation (Std) in %) obtained for each subject in Dataset I for all the competing algorithms and our proposed algorithm UMNFLA.

Subject	aa	al	av	aw	ay	Mean	Std
PSD	64.63	83.44	54.74	65.39	75.43	68.73	11.01
CSP	84.62	94.62	61.42	89.61	91.36	84.33	13.31
RCSP	84.89	93.83	66.01	89.72	90.56	85.00	11.09
INFL	76.32	84.26	58.29	81.33	84.58	76.96	10.95
WT	75.24	81.43	65.79	82.53	85.58	78.11	7.84
NMWF	82.44	89.71	67.19	89.26	89.27	83.57	9.64
UMLDA	87.33	94.28	66.71	90.21	89.96	85.69	10.90
UMNFLA	**92.65**	**97.93**	**73.33**	**92.44**	**94.18**	**90.11**	9.63

4.2 Results on Dataset II

Classification accuracy. For each patient in Dataset II, classification accuracy in each day is calculated under different feature dimensionalities. Then classification accuracies in the same week are averaged to represent the mean accuracy of the week. After that, those mean accuracies of the eight weeks are averaged to represent the mean accuracy of the corresponding patient. Fig. 2 shows the mean accuracies of all the methods under different feature dimensionalities. For all the patients, the accuracies of all the methods change greatly with the increase of feature dimensionality. Another important observation is that UMNFLA has the best performance among all the algorithms. Comparisons using a Mann-Whitney U test between UMNFLA and the other methods show that the accuracies provided by UMNFLA are significantly higher than the others (All UMNFLA vs. each in the competing methods: $p < 0.05$). The classification accuracies by UMNFLA for almost all the patients could even exceed 70%.

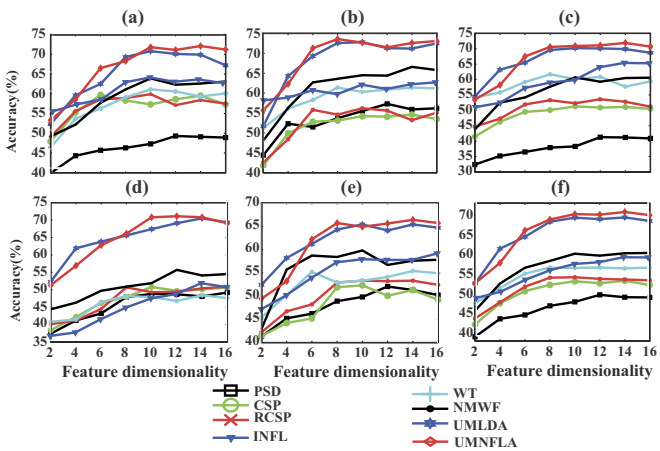

Figure 2. The mean accuracies obtained for each stroke patient in Dataset II for all the competing algorithms and our proposed algorithm UMNFLA under different feature dimensionalities. (a-e) Patient 1-5. (f) Group mean.

Spatial-spectral-temporal patterns. Apart from the superior classification performance, we try to observe the spatial-spectral-temporal patterns (weights) obtained by UMNFLA by visualizing them in 2-D graphs. Through feature selection by Fisher score strategy, the most discriminative EEG patterns can be learned from the corresponding projection matrices of the retained features. In order to observe the EEG patterns changes over time, we choose the raw EEG data of three days, day 1, 30 and 60, to represent the different phases during rehabilitation. For comparison, CSP, which has been proven to be very useful to extract subject-specific, discriminative spatial filters [17], is utilized to obtain spatial patterns of stroke patients.

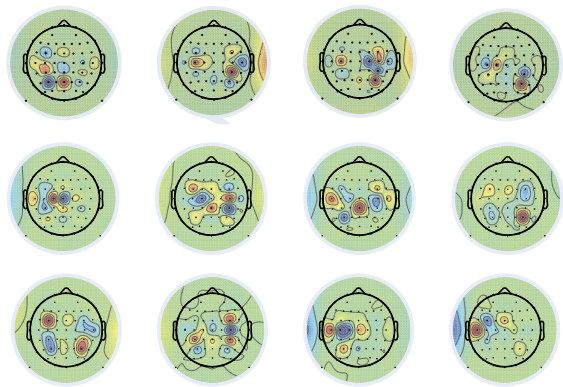

Figure 3. The four most discriminative spatial patterns extracted by CSP for Patient 1 on day 1, 30 and 60 (from top to bottom: day 1, 30 and 60).

Fig. 3 shows the results of the four most discriminant spatial patterns obtained by CSP and Fig. 4 illustrates the spatial-spectral-temporal patterns with the first two largest Fisher scores learned by UMNFLA on the three chosen days for Patient 1 (with lesion in right side), which may provide insights to the underlying cortical activity pattern. In general, the spatial filters obtained by CSP appear as messy, with large weights in several unexpected locations from a neurophysiological point of view. On the contrary, UMNFLA filters are physiologically more relevant. In detail, for the unaffected (left) hemisphere, the activated cortices are stable and mainly located in the left central lobe (C3). However, in the affected (right) hemisphere, larger cortical regions are activated and shifted during rehabilitation. These regions gradually migrate from right central and parietal lobes (C4+P4) to right central, frontal-central and parietal lobes (C4+FC4+P4), and finally to around central lobe (C4). Similar phenomena are also reported in [19]. In terms of spectral characteristics, the most active frequency bands concentrate on a lower band (6-12Hz) at the beginning, but gradually scatter at a wide-ranged band (6-30Hz). Similar observation is also reported from [18].

5 Discussion and Conclusion

In this paper, we propose a novel subspace learning method, called uncorrelated multilinear nearest feature line analysis (UMNFLA), for the recognition of multidimensional objects. UMNFLA extracts uncorrelated discriminative features directly from tensorial data using the TVP of tensor objects, and learns a subspace to project the samples to a low-dimensional feature space such that the within-class feature line distances in the subspace are minimized and between-class feature line distances in the subspace are maximized, simultaneously. Experimental results on EEG datasets recorded from diverse populations including the healthy people and some special popula-

tions suffering from neurophysiological diseases (e.g., stroke) have demonstrated the efficacy and robustness of the proposed method.

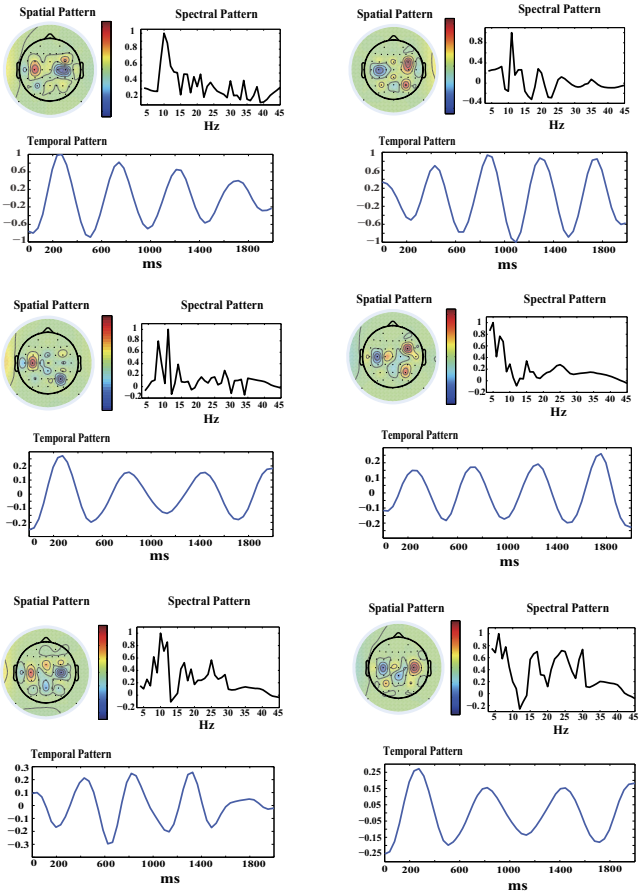

Figure 4. The first two most discriminative spatial, spectral and temporal patterns extracted by UMNFLA using Fisher score for Patient 1 on day 1, 30 and 60 respectively (from top to bottom: day 1, 30 and 60). Values of y axis in the spectral and temporal patterns are normalized. In spatial pattern, red dots represent higher power and blue ones represent lower power.

To provide some insights into the UMNFLA, we exploit some interesting advantages of UMNFLA: First, UMNFLA extracts multilinear discriminative features in multiway discriminative subspace, which greatly improves significant interpretations of multidimensional objects. Second, UMNFLA preserves structural information of tensor objects in the process of projection of the original data, and simultaneously takes class label information into consideration which makes UMNFLA a better classification method than the multiway decomposition methods (like NMWF). Third, UMNFLA never depends on some impossible prior knowledge, like pre-identified frequency band and channels configuration in EEG analysis, leading to a better performance than some traditional methods. Finally, UMNFLA expands the capacity of the available database by using a linear model to provide infinite feature points and their information, and then the small sample size problem is effectively solved.

ACKNOWLEDGEMENTS

The work was supported by the National Natural Science Foundation of China (Grant Nos. 91120305, 61272251).

REFERENCES

[1] Christopher M Bishop et al., *Pattern recognition and machine learning*, volume 1, springer New York, 2006.

[2] Benjamin Blankertz et al, 'The BCI competition III: Validating alternative approaches to actual BCI problems', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **14**(2), 153–159, (2006).

[3] 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), 1057–1061, (2004).

[4] A. Cichocki, R. Zdunek, A.H. Phan, and S. Amari, *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*, Wiley, 2009.

[5] Xiaofei He, Deng Cai, and Partha Niyogi, 'Tensor subspace analysis', in *NIPS*, volume 4, p. 1, (2005).

[6] Z. Jin, J.Y. Yang, Z.S. Hu, and Z. Lou, 'Face recognition based on the uncorrelated discriminant transformation', *Pattern recognition*, **34**(7), 1405–1416, (2001).

[7] Tamara G Kolda and Brett W Bader, 'Tensor decompositions and applications', *SIAM review*, **51**(3), 455–500, (2009).

[8] Y. Koren and L. Carmel, 'Robust linear dimensionality reduction', *IEEE Transactions on Visualization and Computer Graphics*, **10**(4), 459–470, (2004).

[9] S.Z. Li and J. Lu, 'Face recognition using the nearest feature line method', *IEEE Transactions on Neural Networks*, **10**(2), 439–443, (1999).

[10] Ye Liu et al, 'A Tensor-Based Scheme for Stroke Patients' Motor Imagery EEG Analysis in BCI-FES Rehabilitation Training', *Journal of neuroscience methods*, **222**, 238–249, (2014).

[11] F. Lotte and C. Guan, 'Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms', *IEEE Transactions on Biomedical Engineering*, **58**(2), 355–362, (2011).

[12] Isabelle Loubinoux et al, 'Correlation between cerebral reorganization and motor recovery after subcortical infarcts', *Neuroimage*, **20**(4), 2166–2180, (2003).

[13] H. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, 'Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition', *IEEE Transactions on Neural Networks*, **20**(1), 103–123, (2009).

[14] H. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, 'Uncorrelated multilinear principal component analysis for unsupervised multilinear subspace learning', *IEEE Transactions on Neural Networks*, **20**(11), 1820–1836, (2009).

[15] M. Mørup, L. Hansen, J. Parnas, and S.M. Arnfred, 'Decomposing the time-frequency representation of EEG using non-negative matrix and multi-way factorization', *Technical University of Denmark Technical Report*, (2006).

[16] Yanwei Pang, Yuan Yuan, and Xuelong Li, 'Iterative subspace analysis based on feature line distance', *IEEE Transactions on Image Processing*, **18**(4), 903–907, (2009).

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

[18] S. Shahid, R.K. Sinha, and G. Prasad, 'Mu and beta rhythm modulations in motor imagery related post-stroke EEG: a study under BCI framework for post-stroke rehabilitation', *BMC Neuroscience*, **11**, 1–2, (2010).

[19] Wing-Kin Tam, Kai-yu Tong, Fei Meng, and Shangkai Gao, 'A minimal set of electrodes for motor imagery BCI to control an assistive device in chronic stroke subjects: a multi-session study', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **19**(6), 617–627, (2011).

[20] D. Tao, X. Li, X. Wu, and S.J. Maybank, 'General tensor discriminant analysis and gabor features for gait recognition', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **29**(10), 1700–1715, (2007).

[21] Lior Wolf, Hueihan Jhuang, and Tamir Hazan, 'Modeling appearances with low-rank SVM', in *CVPR'07.*, pp. 1–6. IEEE, (2007).

[22] Jieping Ye, Ravi Janardan, Qi Li, and Haesun Park, 'Feature reduction via generalized uncorrelated linear discriminant analysis', *IEEE Transactions on Knowledge and Data Engineering*, **18**(10), 1312–1322, (2006).