2009 International Symposium on Nonlinear Theory and its Applications
NOLTA'09, Sapporo, Japan, October 18-21, 2009

Common Tensor Discriminant Analysis for EEG Classification

Qibin Zhao†‡, Liqing Zhang‡ and Andrzej Cichocki†

†Lab for Advanced Brain Signal Processing, RIKEN Brain Science Institute
Saitama, Japan
‡MOE-Microsoft Laboratory for Intelligent Computing and Intelligent Systems and
Department of Computer Science and Engineering, Shanghai Jiao Tong University
Shanghai, China
Email: qbzhao@brain.riken.jp, zhang-lq@cs.sjtu.edu.cn, a.cichocki@riken.jp

Abstract—In order to explore underlying multi-mode discriminative information simultaneously from high noise EEG signals, multi-way tensor analysis is more suitable for EEG feature extraction. In this paper, we propose a novel algorithm, called common tensor discriminant analysis (CTDA), to solve the supervised subspace learning by encoding each EEG epoch as a M-order tensor. A tensor-based discriminant analysis framework is presented for simultaneous optimization of a series of projection matrices based on tensor analysis theory and CSP criteria. Furthermore, CTDA has been extended to multiclass case. Experimental results demonstrate the effectiveness and superiority of our proposed algorithms.

1. Introduction

Recently, multilinear algebra, the algebra of high-order tensors, was applied for analyzing the multifactor structure image ensembles, EEG signals [1] and etc. These methods, such as tensor PCA [2], tensor LDA [3, 4], tensor subspace analysis [5, 6, 7], treat original data as second- or high-order tensors. For supervised feature classification [8], the tensorization can lead to structured dimensionality reduction by learning multiple interrelated subspaces. Discriminant analysis using tensor representation [9] can avoid the curse of dimensionality dilemma and overcome the small sample size problem .

As we know, EEG classification [10], especially motor imagery based EEG classification, attracts much attention recently as a result of the increasing demand for developing BCI systems. Previous studies demonstrated that CSP algorithm has been successfully used in EEG classification [11, 12] and BCI by optimizing discriminative spatial filters for two classes data. However, most non-motor imagery tasks are still difficult to be classified because of unsufficient discriminant information. There are several factors e.g., spatial, frequency, and time, that mostly affect the recognition accuracy. In order to preserve more discriminative information from original EEG signals, high order tensor is more suitable for EEG representation.

In this paper, we propose a novel tensor subspace learning algorithms, termed common tensor discriminant analysis (CTDA), as an extension of CSP method for high order tensor data. We develop a new general framework of simultaneous optimization of projection matrices on multifactors for M-order EEG tensor representation. Furthermore, CTDA has been extended to multi-class case.

To demonstrate the proposed method, instead of focusing on motor imagery EEG experiments, we are interested in comprehensive coginitive mental tasks such as visual and auditory imagery [13], which is difficult to be classified by CSP. Experimental results demonstrate that CTDA benefits from its encouraging properties and achieves competitive EEG recognition performance.

2. Common tensor discriminant analysis

As a supervised learning method, CTDA algorithm is trained on labeled tensor data, i.e., a set of M-order tensor samples $\mathcal{X}_i \in \mathbb{R}^{N_1\times\ldots\times N_M}, i = 1,\ldots,n$ which belong to several different classes. Firstly, we can define the mode-M covariance tensor (i.e., high order covariance) as

$$\mathcal{R} = \frac{1}{n}\sum_{i=1}^{n}[\![\mathcal{X}_i \circ \mathcal{X}_i; (M)(M)]\!], \tag{1}$$

where $\mathcal{R} \in \mathbb{R}^{N_1\times\ldots\times N_{M-1}\times N_{M-1}\ldots\times N_1}$ is a $2(M-1)$ order tensor which has a symmetric length on first $M-1$ and last $M-1$ mode. $\circ$ and $[\![\bullet]\!]$ are tensor outer product and tensor contraction operations defined in[3].

According to CSP objective functions[11], let $\mathbf{W}^{(c)}$ represents the maximal discriminative pattern for the c-th class, and $\mathbf{X}^{(c)}$ denotes connected training EEG epoch which belongs to the c-th class. Then, we have

$$\mathbf{W}^{(c)^T}\mathbf{R}^{(c)}\mathbf{W}^{(c)} = \mathbf{D}, \quad \mathbf{W}^{(c)^T}\left(\sum_{c=1}^{C}\mathbf{R}^{(c)}\right)\mathbf{W}^{(c)} = \mathbf{I}. \tag{2}$$

Based on $\mathbf{R}^{(c)} = \mathbf{X}^{(c)}\mathbf{X}^{(c)^T}$ and tensor contraction theory, we can further rewrite Eq.(2) as,

$$\left[\!\!\left[(\mathbf{X}^{(c)} \times_1 \mathbf{W}^{(c)^T}) \circ (\mathbf{X}^{(c)} \times_1 \mathbf{W}^{(c)^T}); (2)(2)\right]\!\!\right] = \mathbf{D},$$

and (3)

$$\sum_{c=1}^{C}\left[\!\!\left[(\mathbf{X}^{(c)} \times_1 \mathbf{W}^{(c)^T}) \circ (\mathbf{X}^{(c)} \times_1 \mathbf{W}^{(c)^T}); (2)(2)\right]\!\!\right] = \mathbf{I}.$$

Let $\mathcal{X}_i^{(c)} \in \mathbb{R}^{N_1 \times \ldots \times N_M}$ denote the i-th training sample (i.e., i-th epoch), which is a M-order tensor. Similarly, let $\mathcal{X}^{(c)}$ denotes connected EEG tensors which belong to c-th class, then we define multi-modal projections as,

$$\mathcal{Z}^{(c)} = \mathcal{X}^{(c)} \prod_{k=1}^{M-1} \times_k \mathbf{W}_k{}^T. \tag{4}$$

where $\mathbf{W}_k|_{k=1}^{M-1}$ denote k-th projection matrix on $M-1$ modes respectively, and each of projecting matrix only retains H_k projection directions corresponding to $H_k/2$ largest variance and $H_k/2$ smallest variance directions.

Based on analogy with Eq.(3), we define CTDA by replacing $\mathbf{X}^{(c)}$, $\mathbf{D}$ and $\mathbf{I}$ with $\mathcal{X}^{(c)}$, $\mathcal{D} \in \mathbb{R}^{H_1 \times \ldots \times H_{M-1} \times H_{M-1} \times \ldots \times H_1}$ and $\mathcal{I} \in \mathbb{R}^{H_1 \times \ldots \times H_{M-1} \times H_{M-1} \times \ldots \times H_1}$ respectively, then we obtain

$$\begin{aligned} &\left[\!\left[\mathcal{Z}^{(c)} \circ \mathcal{Z}^{(c)}; (M)(M) \right]\!\right] = \mathcal{D}, \\ &\sum_{c=1}^{C} \left[\!\left[\mathcal{Z}^{(c)} \circ \mathcal{Z}^{(c)}; (M)(M) \right]\!\right] = \mathcal{I}. \end{aligned} \tag{5}$$

The problem defined in Eq.(5) does not have a closed form solution, so we choose to use the alternating projection method, which is an iterative procedure, to obtain a numerical solution. Therefore, Eq.(5) is decomposed into $M-1$ different optimization sub-problems, as follows,

$$\begin{aligned} &\left[\!\left[\mathcal{Z}^{(c)} \circ \mathcal{Z}^{(c)}; (\bar{l})(\bar{l}) \right]\!\right] = \mathbf{D}_l, \quad l = 1, \ldots, M-1; \\ &\sum_{c=1}^{C} \left[\!\left[\mathcal{Z}^{(c)} \circ \mathcal{Z}^{(c)}; (\bar{l})(\bar{l}) \right]\!\right] = \mathbf{I}_l, \quad l = 1, \ldots, M-1. \end{aligned} \tag{6}$$

This can be interpreted as the tensor data $\mathcal{X}^{(c)}$ are filtered on $M-1$ modes by matrices $\mathbf{W}_k^{(c)}|_{k=1}^{M-1}$ respectively and matricized on each l-th mode would be diagonal matrices of $\mathbf{D}_l, \mathbf{I}_l \in \mathbb{R}^{H_l \times H_l}$.

To further explore each sub-problem in Eq.(6), we define

$$\mathcal{Y}^{(c)} = \mathcal{X}^{(c)} \prod_{\substack{k=1 \\ k \neq l}}^{M-1} \times_k \mathbf{W}_k{}^T. \tag{7}$$

then we can obtain

$$\begin{aligned} &\mathbf{W}_l^{(c)T} \left[\!\left[\mathcal{Y}^{(c)} \circ \mathcal{Y}^{(c)}; (\bar{l})(\bar{l}) \right]\!\right] \mathbf{W}_l^{(c)} = \mathbf{D}_l, \\ &\mathbf{W}_l^{(c)T} \left\{ \sum_{c=1}^{C} \left[\!\left[\mathcal{Y}^{(c)} \circ \mathcal{Y}^{(c)}; (\bar{l})(\bar{l}) \right]\!\right] \right\} \mathbf{W}_l^{(c)} = \mathbf{I}_l. \end{aligned} \tag{8}$$

To simplify Eq.(8), we define

$$\mathbf{U}_l^{(c)} = \left[\!\left[\mathcal{Y}^{(c)} \circ \mathcal{Y}^{(c)}; (\bar{l})(\bar{l}) \right]\!\right] \quad \text{and} \quad \mathbf{T}_l = \sum_{c=1}^{C} \mathbf{U}_l^{(c)}. \tag{9}$$

Thus, Eq.(8) can be written as

$$\mathbf{W}_l^{(c)T} \mathbf{U}_l^{(c)} \mathbf{W}_l^{(c)} = \mathbf{D}_l, \quad \mathbf{W}_l^{(c)T} \mathbf{T}_l \mathbf{W}_l^{(c)} = \mathbf{I}_l, \tag{10}$$

where $l \in [1 : M-1], c \in [1 : C]$, and $\mathbf{W}_l^{(c)}$ denotes projection matrix on l-th mode for c-th class. Hence, CDTA optimization problem is equivalent to $M-1$ sub-problems in Eq.(10) which can be solved by two step PCA method. Finally, we can combine the $\mathbf{W}_l^{(c)}$ corresponding to each c class as:

$$\mathbf{W}_l = [\mathbf{W}_l^{(1)}, \ldots, \mathbf{W}_l^{(C)}], \quad l = 1, \ldots, M-1. \tag{11}$$

Therefore, for the M order training tensors $\mathcal{X}$, CTDA would obtain $M-1$ optimal projection matrices $\mathbf{W}_l|_{l=1}^{M-1}$ by solving the $M-1$ alternative sub-problems defined in Eq.(10).

Once we obtain the projection matrices based on CTDA, the projection of M-order tensor data $\mathcal{X}$ is given as

$$\mathcal{Z} = \mathcal{X} \prod_{l=1}^{M-1} \times_l \mathbf{W}_l{}^T. \tag{12}$$

The feature vector of tensor data $\mathcal{X} \in \mathbb{R}^{N_1 \times N_2 \times \ldots \times N_M}$ used for classification is composed of the $H_1 \times \ldots \times H_{M-1}$ variances normalized by the total variance of the projections retained, and log-transformed,

$$\mathbf{f} = \log \left\{ \frac{\operatorname{diag}\left[\mathcal{Z}_{(M)}^T \mathcal{Z}_{(M)} \right]}{\operatorname{tr}\left[\mathcal{Z}_{(M)}^T \mathcal{Z}_{(M)} \right]} \right\}. \tag{13}$$

3. EEG classification during visual and auditory imagery tasks

The motivation for our experiments is that even though motor imagery tasks are effectively used in BCI systems, they may be difficult to perform for a person who has been paralyzed for several years. Also it has been shown that patients with spinal chord injuries do not perform motor imagery tasks as well as able-bodied persons. Furthermore, it's very promising to find more comprehensive coginitive based imagery tasks which can be classified from EEG signals. Therefore, we choose two cognitive tasks of visual imgery (VI) and auditory imagery (AI) to evaluate the effectiveness of CTDA. To classify two imagery tasks, EEG signals are firstly tansformed to a tensor representation, then we apply CTDA algorithm for feature extraction based on tensor training samples and linear support vector machines (SVM) for classification.

In the experimental sessions used for the present study, labeled trials of EEG signals were recorded in the following way: the subjects were sitting in a comfortable chair with arms lying relaxed on the armrests. Each trial consists of 10s for relaxation and 10s for cognitive tasks following visual cue stimulus. During the VI period, the subject was instructed to imagine some familiar visual scenes one by one. The familiar visual scenes can be any rooms or any furnitures in their home. During the AI period, the subject was asked to think of a favorite song or a familiar tune that they enjoyed. They were instructed to " listen " to it in their head, without mouthing the words or moving any part of their body.

EEG signals were recorded from 62 electodes on scalp according to 10-20 system with a sampling rate of 250Hz. In order to consider multi-factors information simultaneously, the EEG are transformed using a Morlet continuous wavelet transform (CWT) with center frequency $\omega_c = 1$ and bandwidth parameter $\omega_b = 2$. The frequency band of 5-30Hz is thus adopted for establishing time-frequency transform in CWT. Then, we obtain EEG tensor representation $\mathcal{X} \in \mathbb{R}^{N_d \times N_f \times N_t}$ which is a three-way time-varying EEG wavelet coefficients array, where N_d, N_f, N_t are number of channels, number of frequency bins, and time points, respectively.

As expected, by simultaneous optimization on multi-way tensor, CTDA can obtain optimal spatial filters and frequency patterns which contain the most discriminative information. For further illustrations of the proposed method, we will pick one specific dataset of one subject to visualize the projection patterns on each mode which are obtained by CTDA method. Fig.1 represents spatial projection matrices $\mathbf{W}_1^{(c)}, c = 1, 2$. There are two groups which denote the spatial projection matrices for two classes, i.e., VI and AI respectively. It's obvious that VI focuses on occipital area while AI focuses on temporal area which is close to ears. Meanwhile, CTDA can learn optimal discriminative frequency patterns for the specific imagery task (see fig.2).

(a) First spatial filter for VI (b) First spatial filter for AI

Figure 1: The largest spatial filters on scalp map for two imagery tasks.

Hence, we can conclude that after spatial projection on occipital area and frequency projection about 24-30Hz, the variance along time dimension are largest during VI task, while the variance are smallest during AI task. On the other hand, AI task has maximal variance for spatial projection on temporal area and frequency projection around 18Hz, meanwhile VI task has minimal variance for the same projection patterns. This illustrates that the different coginitave tasks not only have distinct spatial distribution but also have distince frequency distribution. The results indicate that the frequency information are not only subject-dependent but also class-dependent. Therefore, these frequency patterns can provide further discriminative information which can not be obtained by only spatial filters.

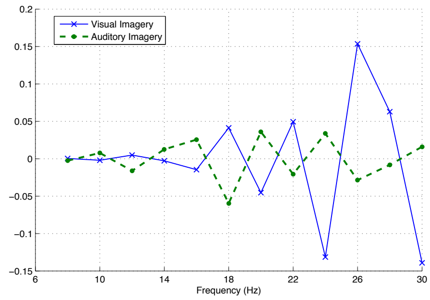

Figure 2: The largest filters in frequency domain for two imagery tasks. VI mainly focuses on 24-30Hz (high β rhythm), while AI mainly focuses around 18Hz (low β rhythm.

After we obtain optimal projections from CTDA, the EEG epoches can be projected using $\mathbf{W}_l$ and the feature vectors are calculted according to Eq.(13). Fig.3 illustrates the averaged feature vectors for each of two classes. It's clearly shown that two imagery tasks can be seperated by these optimal features with the number of features are 8 (2 maximal patterns on spatial and frequency modes for each of two classes). In the end, the 5×5-fold cross validation results for two subjects with two cognitive tasks (i.e., VI vs. AI) are presented in Table.1.

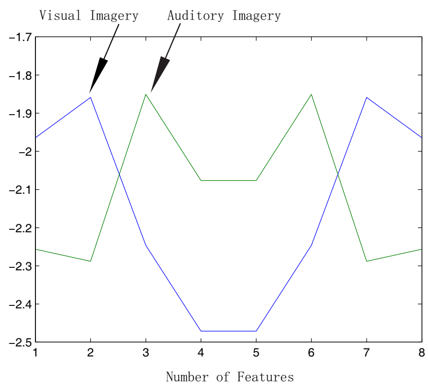

Figure 3: Averaged feature vectors for two imagery tasks.

Table 1: Classification accuracies[%] and iterative numbers for CTDA.

Subject	Iterative Number	Accuracy
SA	9	89.0 ± 1.9
SB	10	82.3 ± 2.2

So instead of having a spatial projection onto a broad frequency band signal as a solution given by CSP, CTDA can split information furthermore by projecting onto multi-frequency signals of the same local origin, stemming from different sub-bands, such that each projection fulfills the optimization criterion of maximizing the variance for one class, while having minimal variance for the sum of all classes. Summarizing, this yields an improved spatio-frequency resolution of the discriminative signals. Therefore, by combining the tensor representations of EEG and CTDA, more discriminative information hidden in raw signals can be obtained automatically by learning optimal projections on multi-dimensions simultaneously.

As compared with CSP, CTDA helps to reduce the number of parameters needed to model the data. For example, when a tensor $\mathcal{X}$ has the size $N_1 \times \ldots \times N_M$, we need to estimate the projection matrix $\mathbf{W}$ with the size $N_1 \ldots N_{M-1} \times H$ by vectorization operation and CSP, but we only need to estimate the projection matrices $\mathbf{W}_k|_{k=1}^{M-1}$ with the corresponding size $N_k \times H_k, k = 1, \ldots, M-1$ in CTDA.

4. Conclusions

In this paper, we have developed a supervised tensor-based learing framework so that they accept M-order tensors as inputs. Experimental analysis for EEG classification during two cognitive tasks (i.e., visual and auditory imagery) demonstrates the good performance and the advantages of proposed algorithm.

Acknowledgments

The work was supported in part by the Science and Technology Commission of Shanghai Municipality (Grant No. 08511501701), and the Okawa Foundation Research Grant Program, Japan.

References

[1] M. Mørup, L.K. Hansen, and S.M. Arnfred. ERP-WAVELAB A toolbox for multi-channel analysis of time–frequency transformed event related potentials. *Journal of Neuroscience Methods*, 161(2):361–368, 2007.

[2] J. Yang, D. Zhang, A.F. Frangi, and Yang J.Y. Two-dimensional pca: A new approach to appearance-based face representation and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(1):131–137, 2004.

[3] 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.

[4] J. Ye, R. Janardan, and Q. Li. Two-dimensional linear discriminant analysis. *NIPS*, 2004.

[5] X. Wang and X. Tang. A unified framework for subspace face recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(9):1222–1228, 2004.

[6] Y. Fu and T.S. Huang. Image classification using correlation tensor analysis. *IEEE Transaction on Image Processing*, 17(2):226–234, 2008.

[7] X. He, D. Cai, and P. Niyogi. Tensor subspace analysis. *NIPS06*, 2006.

[8] D. Tao, X. Li, W. Hu, S.J. Maybank, and X. Wu. Supervised tensor learning. *Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM'05)*, 2007.

[9] S. Yan, D. Xu, Yang Q., L. Zhang, Tang X., and H. Zhang. Discriminant analysis with tensor representation. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005.

[10] B. Blankertz, G. Curio, and K.R. Müller. Classifying single trial EEG: Towards brain computer interfacing. *Advances in Neural Information Processing Systems*, 14:157–164, 2002.

[11] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller. Optimal spatial filtering of single trial EEG during imagined handmovement. *Rehabilitation Engineering, IEEE Transactions on [see also IEEE Trans. on Neural Systems and Rehabilitation]*, 8(4):441–446, 2000.

[12] J.Müller-Gerking, G.Pfurtscheller, and H.Flyvbjerg. Designing optimal spatial filters for single-trial eeg classification in a movement tast. *Clinical Neurophysiology*, 110:787–798, 1999.

[13] A.F. Cabrera and K. Dremstrup. Auditory and spatial navigation imagery in brain-computer interface using optimized wavelets. *Journal of Neuroscience Methods*, 174:135–146, 2008.