NMF를 이용한 Motor Imagery 뇌파 분류

이혜경[O] Andrezej Cichocki 최승진
포항공과대학교 컴퓨터공학과
leehk@postech.ac.kr[O], cia@brain.riken.jp, seungjin@postech.ac.kr

NMF for Motor Imagery EEG Classification

Hyekyoung Lee[0], Andrezej Cichocki, Seungjin Choi
POSTECH Dept. of computer science and engineering

요 약

In this paper, we present a method of feature extraction for motor imagery single trial EEG classification, where we exploit nonnegative matrix factorization (NMF) to select discriminative features in the time-frequency representation of EEG. Experimental results with motor imagery EEG data in BCI competition 2003, show that the method indeed finds meaningful EEG features automatically, while some existing methods should undergo cross-validation to find them.

1. Introduction

Brain computer interface (BCI) is a system that is designed to translate a subject's intention or mind into a control signal for a device such as a computer, a wheelchair, or a neuroprosthesis. The most popular sensory signal used for BCI is EEG which is the multivariate time series data where electrical potentials induced by brain activities are recorded in a scalp. Exemplary spectral characteristics of EEG involving motor, might be μ rhythm (8-12 Hz) and β rhythm (18-25 Hz) which decrease during movement or in preparation for movement (ERD) and increase after movement and in relaxation (ERS) [1]. ERD and ERS could be used as relevant features for the task of motor imagery EEG classification. However those phenomena might happen in a different frequency band for some subjects, for instance, in 16-20 Hz, not in 8-12 Hz [2]. Moreover, it is not guaranteed that a subject always concentrates on imagination during experiments. Thus, it is desirable to determine appropriate activated frequencies and associated features for each subject, during motor imagery experiments. In this paper we present a method of discriminative feature extraction where we exploit the sparseness, L_1 norm, and nonnegative matrix factorization (NMF). Morlet wavelets are used to construct a nonnegative data matrix from the time-domain EEG data. We use the NMF with α-divergence that was recently proposed in [3,4]. Numerical experiments using Data Set III of BCI competition 2003 show that our NMF-based method learns basis vectors indicating discriminative frequencies and determine useful features for the task of single-trial online classification of imaginary left and right hand movements [5].

2. Nonnegative Matrix Factorization

NMF is one of widely-used multivariate analysis methods for nonnegative data, which has many potential applications in pattern recognition and machine learning [6,7]. Denote the data matrix by $X=[x(1),\cdots,x(N)]\in R^{m\times N}$, where $\{x(t)\}$ is N observed m-dimensional data points. NMF seeks a decomposition of the nonnegative data matrix X that is of the form:

$$X\approx AS,\ \ ([A]_{ij}\geq 0 \text{ and } [S]_{ij}\geq 0) \tag{1}$$

where $A\in R^{m\times n}$ contains basis vectors in its columns and $S\in R^{n\times N}$ is the associated encoding variable matrix. Amari's α-divergence and its multiplicative algorithm were proposed in [3,4]. The α-divergence between X and AS is given by

$$D_\alpha[X\,\|\,AS]=\frac{1}{\alpha(1-\alpha)}\sum_{i,j}\left[\alpha X_{ij}+(1-\alpha)[AS]_{ij}-X_{ij}^{\alpha}[AS]_{ij}^{1-\alpha}\right] \tag{2}$$

The parameter α is associated with the characteristics of a learning machine, in the sense that the model distribution is more inclusive (as α goes to ∞) more exclusive (as α goes to $-\infty$). The multiplicative algorithm regarding the minimization of the α-divergence of AS from X in Eq. (2), is given by [4]

$$S_{ij}\leftarrow S_{ij}\left[\frac{\Sigma_k[A_{ki}(X_{kj}/[AS]_{kj})^{\alpha}]}{\Sigma_l A_{li}}\right]^{\frac{1}{\alpha}}, \tag{3}$$

$$A_{ij}\leftarrow A_{ij}\left[\frac{\Sigma_k[S_{jk}(X_{ik}/[AS]_{ik})^{\alpha}]}{\Sigma_l S_{jl}}\right]^{\frac{1}{\alpha}}. \tag{4}$$

3. Proposed Method

The overall structure of our proposed single trial EEG classification is illustrated in Fig. ₩ref{fig:algorithm test}, where the method consists of three steps: (1) preprocessing involving wavelet transform; (2) NMF-based feature extraction; (3) probabilistic model-based classification. Each of these steps is described in detail.

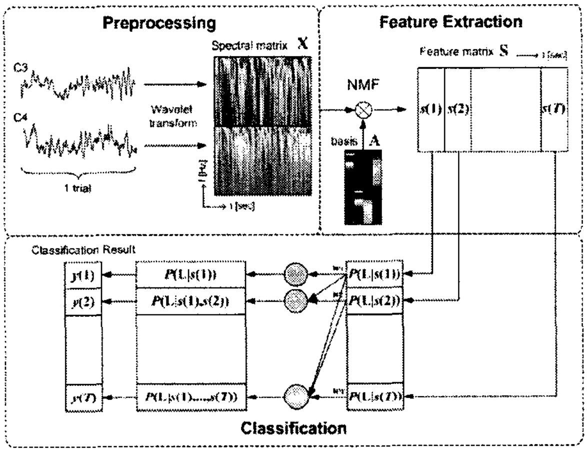

Fig.1. the overall structure of proposed method

Data Description : We used one of BCI competition 2003 data sets, which was provided by the Department of Medical Informatics, Institute for Biomedical Engineering, Graz University of Technology, Austria [8]. We use only, C_3 and C_4 channels because of the contralateral property of brain activity.

Preprocessing : We obtain the time-frequency representation of the EEG data, by filtering it with complex Morlet wavelets,

$$\Psi_{\tau,d(f)}(t)=\pi^{-1/4}e^{iw_0(t-\tau)/d(f)}e^{-\{(t-\tau)/d(f)\}^2/2},$$

where w_0 is the characteristic eigenfrequency, $d(f)=(w_0+\sqrt{2+w_0^2})/(4\pi f)$ and f is the main receptive frequency. The wavelet transform of $C_{i,k}(t)(i=3,4)$ at time τ and frequency f with C_i channel and the kth trial is their convolution with scaled and shifted wavelets. The amplitude of the wavelet transform, $x_{i,k}(f,\tau)$, is given by $x_{i,k}(f,\tau)=\| C_{i,k}(t)*\Psi_{\tau,d(f)}(t)\|$ for $i=3,4$ and $k=1,\ldots,K$ where K is the number of trials. Concatenating those amplitudes for $i=3,4$ and $(f_1,\ldots,f_{27})=[4,\ldots,30]$ Hz, leads to $x_k(t)\in R^{54}$ of the form $x_k(t)=[x_{3,k}(f_1,t),\cdots,x_{3,k}(f_{27},t)\ x_{4,k}(f_1,t),\cdots,x_{4,k}(f_{27},t)]^\top$. Incorporating with T data points in each trial, we construct $X_k=[x_k(1),\cdots,x_k(T)]\in R^{54\times T}$. Collecting K trials leads to the data matrix $X=[X_1,\cdots,X_K]\in R^{54\times KT}$. Labelled and unlabelled data are distinguished by X_{train} and X_{test}, respectively.

Feature Extraction : We extract feature vectors by applying NMF to the data matrix X constructed from the wavelet transform of EEG over the frequency range $f\in[4,\ldots,30]$ Hz. The data matrix $X\in R^{54\times KT}$ contains a large number of data vectors reflecting K trials and T data points of EEG. Instead of using the whole data vectors, we first select candidate vectors which are expected to be more discriminative, then use only those candidate vectors as inputs to NMF, in order to determine the basis matrix A. The power spectrum in the localized frequency range such as μ or β band of C_3 and C_4 channels, is activated during the imagination of movement. Thus, we investigate the power and sparseness of each data vector to select candidate vectors. We use the sparseness measure in [9], described by $\xi(x)=\{\sqrt{m}-(\Sigma|x_i|)/\sqrt{\Sigma x_i^2}\}/\{\sqrt{m}-1\}$, where x_i is the ith element of the m-dimensional vector x. The candidate vector selection is performed in the following way. First, we compute the power of each column of X, by summing its elements. The average power $\bar{\phi}$ is computed by the sum of all elements in x_i. For each column of X, the sparseness is calculated for C_3 and C_4 channels, by considering the first 27 rows and the last 27 rows of X, respectively. Averaged sparseness values for each channel are computed, then they are added, leading to the final average sparseness. We select candidate vectors from X if the data vector has the power greater than the average power and has the sparseness greater than 70% of the average sparseness. We apply the NMF algorithm in Eq. (3) and (4), to the candidate data matrix $\tilde{X}$, leading to $\tilde{X}=A\tilde{S}$. In our experiments, about 31% of data vectors were selected as candidate vectors.

Classification : We use the probabilistic model-based classifier proposed in [5], where Gaussian class-conditional densities for a single data point in time t are integrated temporally by taking the expectation of the class probabilities with respect to the discriminative power at each point in time.

4. Numerical Experiments

The time-domain EEG data is transformed into the time-frequency representation by complex Morlet

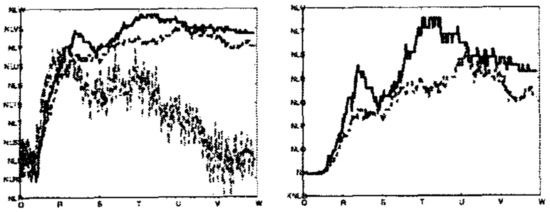

Fig.2. Online (a) classification accuracy and (b) MI

wavelets with $w_0 = 6$, $f = [4;...;30]$ Hz, and $\tau = [3;...;9]$ sec. We select candidate spectral vectors using the method described in Sec.3.3. Then we apply the NMF algorithm in (3) and (4) with $\alpha = 0.5;1;2$ and $n = 2;4;5;6$ (the number of basis vectors). As the number of basis vector increases, the spectral components such as μ rhythm (8-12 Hz), β rhythm (18-22 Hz), and sensori-motor rhythm (12-16 Hz), appear in the order of their importance. All rhythms have the property of contralateral dominance, so they are present in basis vectors associated with C_3 or C_4 channel, separately. In our empirical study, the best performance was achieved when $\alpha = 0.5;1$ and 5 basis vectors. The single trial on-line classification result, is shown in Fig.2, where the classification accuracy is shown in (a) and the mutual information (MI) between the true class label and the estimated class label is plotted in (b). The classification accuracy is suddenly raised from 3.43 sec. The maximal classification accuracy is 88.57% at 6.05 sec, which is higher than the result without the data selection step in the training phase (86.43% at 7.14 sec). MI hits the maximum, 0.6549 bit, which occurs at 6.05 sec. The result is better than the one achieved by the BCI competition 2003 winner (0.61 bit). Table 1 show the maximum mutual information in the time courses per a trial varying the value of α and the number of basis. The smaller the value of α, the better the mutual information, however, α is not critical of determining the performance.

5. Conclusion

We have presented an NMF-based method of feature extraction for on-line classification of motor imagery EEG data. We have also introduced a method of data selection where the power and the sparseness was exploited. Empirical results confirmed that the data selection scheme really improved the classification accuracy by 2.14 % and the mutual information by 0.1127 bit. Existing methods should undergo the cross-validation several times, in order to select discriminative frequency features. However, we have shown that our NMF-based method could find discriminative and representative basis vectors (which reflected appropriate spectral characteristics) without the cross-validation, which improved the on-line classification accuracy. Our method improved the mutual information achieved by BCI competition 2003 winner, by 0.0449 bit, where two frequencies (10 and 22 Hz) were selected using the leave-one-out cross validation. The value of α in the NMF algorithm, was not critical in our empirical study. However, it was confirmed that the parameter α is associated with the characteristics of a learning machine, showing that distributions of basis vectors become more smooth, as α goes to ∞.

α	number of basis				
	2	4	5	6	7
0.5	0.5545	0.5803	0.6549	0.6256	0.5875
1	0.5545	0.5803	0.6549	0.6256	0.5803
2	0.5408	0.5745	0.6404	0.6256	0.5803

Table 1. MI for various α and number of basis

References

[1] Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M., Brain-computer interfaces for communication and control, Clinical Neurophysiology, 113, p.767-791, 2002.

[2] Lal, T.N., Schroder, M., Hinterberger, T., Weston, J., Bogdan, M., Birbaumer, N., Scholkopf, B., Support vector channel selection in BCI, Technical Report 120, Max Planck Institute for Biological Cybernetics, 2003.

[3] Cichocki, A., Zdunek, R., Amari, S., New algorithms for non-negative matrix factorization in applications to blind source separation, ICASSP, 2006.

[4] Cichocki, A., Choi, S., Nonnegative matrix factorization with α-divergence: A multiplicative algorithm, IAPR Int'l Workshop on Statistical Techniques in Pattern Recognition, 2006.

[5] Lemm, S., Schafer, C., Curio, G., BCI competition 2003-data set III: Probabilistic modelling of sensorimotor μ rhythms for classification of imaginary hand movements, IEEE Trans. Bio. Eng., 51, 2004.

[6] Lee, D.D., Seung, H.S., Learning the parts of objects by non-negative matrix factorization, Nature, 401, p.788-791, 1999.

[7] Lee, D.D., Seung, H.S., Algorithms for non-negative matrix factorization, NIPS, 2001.

[8] Blankertz, B., Muller, K.R., Curio, G., Vaughan, T.M., Schalk, G., Wolpaw, J.R., Schlogl, A., Neuper, C., Pfurtscheller, G., Hinterberger, T., Schroder, M., Birbaumer, N., The BCI competition 2003: Progress and perspectives in detection and discrimination of EEG single trials, IEEE Trans. Bio. Eng.,51, 2004.

[9] Hoyer, P.O., Non-negative matrix factorization with sparseness constraints, JMLR, 5, p.1457-1469, 2004.