

Classification of Motor Imagery EEG Based on Sparsification and Non-negative Matrix Factorization

Jing Su , Zuyuan Yang, Haiping Wang and Wei Han

School of Automation, Guangdong University of Technology

Abstract. The analysis of EEG is a hot topic in the area of biomedical signal processing. In this paper, the EEG signals with Mu (μ) rhythm and Beta (β) rhythm are used to solve the motor imagery problem, i.e., the imagery of the left hand and the right hand. The collected raw data is first filtered by FIR band-pass filter, followed by using the maximization of feature difference to increase the sparsity of the matrix. Then, to reduce the redundant information of these features, a non-negative matrix factorization (NMF) method is employed. Due to the usage of the NMF scheme, the obtained factorizations has been better class property. Simulations show that our method achieves higher classification accuracy (more than 91%) than existing results (about 86%).

1 Introduction

Biomedical signal is a weaker low frequency signal in the background of strong noise. It is a natural unstable signal which is emitted by complex living organisms. From the characteristics of the signal, it's detection methods and processing technology are different from the general signal. From the nature of electricity, it can be divided into electrical signals and non-electrical signals, such as ECG, EMG, EEG and other electrical signals. Body temperature, blood pressure, respiration, blood flow, pulse, and heart sounds belong to non-electrical signals[1]. EEG is a kind of bio-electric activity with rhythm and spontaneity. It is closely linked with life. Once the organism dies, the electricity phenomenon disappears[2].

In the practical application of brain-computer interface, there are three kinds of experimental models: motion imagination, steady-state visual evoked(SSV-EP)[3] and P300[4], which SSVEP and P300 are based on the induced pattern of EEG. The essence of the movement imagination of EEG is a spontaneous EEG, in this dynamic state can not feel the muscle movement. Some studies have shown that the area of sensory motor cortex activated by motor imagination is the same as the cortex area of sensory motor brain activated by actual physical exercise [5]. If one does not actively participate in motor activities, feel the processing or imagine such movements and processing[6], one will show significant oscillations in the sensory area of the 8Hz-12Hz and 18Hz-25Hz frequency band EEG. This oscillation, commonly referred to as Mu (μ) rhythms (8Hz-12Hz) and Beta(β) rhythms (18Hz-25Hz), that is generated by the circuit of the thalamic cortex [7]. In the exercise of imagination, ipsilateral hemispheric μ rhythm and β

rhythm are inhibited, while the countermeasures of half-brain μ rhythm and β rhythm increases. This phenomenon is event-related desynchronization (ERD) or event-related synchronization (ERS) [8,9].

In recent years, feature extraction methods based on EEG signals have been rapidly developed. These studies mainly use different feature extraction techniques to evaluate different classifiers. Recently, Zhao et al. calculated the energy difference between the μ rhythm and the β rhythm through power spectrum analysis to extract the signal features[10]. This method has the advantages of clear objective and simple calculation and is conducive to the realization of online BCI system. Sun et al. presents a method of online classification for BCI based on common spatial pattern(CSP) for feature extraction and support vector machine as a classifier[11]. Zhang et al. proposed a semi-supervised feature learning method based on deep stack network [12]. This method combines unsupervised learning of restricted boltzmann machine with gradient descent algorithm based on batch mode, which need two stages of pre-training and fine-tuning. However, these algorithms all have some limitations, such as poor classification accuracy or need labels as supervisory information.

Recently, non-negative matrix factorization(NMF) is commonly used to classify EEG signals. Lee et al. exploited non-negative matrix factorization to select discriminative features in the time-frequency representation of EEG[13]. The non-negative matrix factorization method was used to extract attention-related EEG features[14]. In this paper, we propose an unsupervised feature extraction method called sparsification and non-negative matrix factorization. The central idea of the method is that since the characteristics of motor imaging EEG focus on the μ rhythm and β rhythm, the signal is

first filtered by a 8-25 Hz band-pass filter and then the sparsity of the matrix, increases by maximizing the feature difference, extracting distinguishable features with matrix decomposition to extract, and finally clustering these features. The main advantage of this paper is that non-negative matrix factorization is an unsupervised learning process that is widely used and does not require known signals as oversight. After the feature difference is maximized, the sparsity of the matrix is increased, the computational complexity of non-negative matrix factorization is reduced, and the efficiency of the algorithm and the accuracy of the classification are improved. Experiments show that we method based on non-negative matrix decomposition to learn the discriminable basis vectors, which is very suitable for offline analysis of motor imaging EEG in many experiments.

The structure of this paper is as follows. In the second part, we introduces how to increase the matrix sparsity, and the background of non-negative matrix factorization, and explain the main idea of the algorithm. The real data experiment results are shown in Section 3, and the conclusions are presented in Section 4.

2 Method

2.1 Sparsification

In the experiment, we distinguish the left-right movement imagination by ERD and ERS. In order to improve the efficiency and classification accuracy of the algorithm, we first increase the signal sparsity by maximization of feature difference.

Suppose the k -th experimentally measured EEG signal is E_k ,

$$E_k = \begin{bmatrix} x_{1k}(1) & x_{1k}(2) & \dots & x_{1k}(t) \\ x_{2k}(1) & x_{2k}(2) & \dots & x_{2k}(t) \\ \vdots & \vdots & \ddots & \vdots \\ x_{nk}(1) & x_{nk}(2) & \dots & x_{nk}(t) \end{bmatrix} \quad (1)$$

It consists of the signal with n channels, the length of time is t , and the sparsification process is as follows:

1. Find the covariance matrix of left and right hand motion imagination respectively C_l and C_r [15]:

$$C_l = \sum_{K \in I_l} E_k^T E_k \quad C_r = \sum_{K \in I_r} E_k^T E_k \quad (2)$$

If the k -th example is a left-hand imagination movement, $k \in I_l$, otherwise, $k \in I_r$.

2. The eigenvalue decomposition of the covariance matrix is represented as

$$C = C_l + C_r = O_c \Sigma O_c^T \quad (3)$$

where Σ is the eigenvalue diagonal matrix, O_c is the corresponding eigenvector matrix.

3. Construct a whitening matrix, convert the covariance matrix C_l and C_r do the corresponding eigenvalued ecomposition, described by

$$P = \sqrt{\Sigma^{-1}} O_c^T \quad (4)$$

$$S_l = P C_l P^T = O \Sigma_l O^T \quad (5)$$

$$S_r = P C_r P^T = O \Sigma_r O^T$$

According to the principle of simultaneous diagonalization of the matrix, S_l and S_r have the same eigenvector, it's O , and their sum of corresponding eigenvalues is 1, which is:

$$S_l + S_r = I \quad (6)$$

where, I is the identity matrix. And so, $M = O^T P$ is the space filter that we want. Since the sum of the eigenvalues is 1, if S_l has the maximum eigenvalue, then the corresponding S_r has the smallest eigenvalue, and vice versa. So distinguishing between these two sets of data can be implemented with the feature vector M . The first and last eigenvectors of the EEG projection into M will give the least squares meaning to distinguish the best features of the two types of EEG. According to the projection matrix, single trial EEG experiment can be decomposed into:

$$V = M E \quad (7)$$

In this way, the characteristics used for classification can be obtained from the V variance of the first and last line.

For example, suppose there are two matrices E_1, E_2 :

$$E_1 = \begin{pmatrix} 1.5 & 1.3 \\ 1.7 & 1.6 \\ 2.0 & 2.0 \\ 1.9 & 1.5 \\ 1.6 & 1.4 \end{pmatrix} \quad E_2 = \begin{pmatrix} 1.4 & 1 \\ 1.8 & 1 \\ 1.3 & 1 \\ 1.6 & 2 \\ 1.8 & 1 \end{pmatrix}$$

After the sparsification, the projection matrix is obtained

$$|M| = \begin{pmatrix} 5.8402 & 5.0347 \\ 5.1576 & 5.7689 \end{pmatrix}$$

Since the non-negative matrix ensures that all data is non-negative, so add absolute value here. Therefore, the matrix E_1, E_2 after the maximization of feature difference are shown as V_1, V_2 :

$$V_1 = \begin{pmatrix} 2.0554 & 0.0525 \\ 1.6726 & 0.6712 \\ 1.3652 & 1.4684 \\ 3.3600 & 0.9126 \\ 2.1237 & 0.0209 \end{pmatrix} \quad V_2 = \begin{pmatrix} 0.0759 & 2.1 \\ 1.2287 & 1.3 \\ 0.3716 & 1.1 \\ 0.9709 & 3.4 \\ 0.7129 & 1.8 \end{pmatrix}$$

It can be seen that, after the maximization of feature difference, the sparsity of the matrix is well increased.

After sparsification of motor imagery EEG signals, the signals of different channels in different experiments were respectively classified into different matrices, and the features were extracted by NMF. Finally we conducted clustering analysis of the features of matrix.

2.2 NMF

Non-negative matrix factorization(NMF) has been introduced as a matrix factorization tool that produces a useful decomposition in the dimensionality reduction of data set. The non-negative matrix factorization can be described as follows:

$$V \approx W \times H \quad (8)$$

$V \in R^{m \times n}$ consists of column vectors each of which corresponds to a different sample. Find out basis matrix $W \in R^{m \times r}$ and weight matrix $H \in R^{r \times n}$. Each column vector in V has a corresponding representation vector in the H . Each matrix can be represented as a composition of multiple vectors, $V = [v_1 \ v_2 \ v_3 \ \dots \ v_n]$ and $H = [h_1 \ h_2 \ h_3 \ \dots \ h_n]$.

We consider the squared Euclidean distance as a discrepancy measure between the V and the $W \times H$, leading to the following cost function:

$$D(V \| WH) = \|V - WH\|^2 = \left(\sum_{i=1}^m \sum_{j=1}^n V_{ij} - [WH]_{ij} \right)^2 \quad (9)$$

It is shown that the cost function defined by Eq. (9) is a convex function of W and H separately. Therefore, multiplicative and alternating update rules for H and W are derived by gradient decent optimization [16]. The update rules is used to gradient descent algorithm. It is proven by LEE and SEUNG[16] that convergence is guaranteed when step size is chosen as η_H for the update H_{aj} and chosen as η_W are given by Eq.(10).

$$\eta_H = \frac{H_{aj}}{(W^T WH)_{aj}}, \quad \eta_W = \frac{W_{aj}}{(WHH^T)_{aj}} \quad (10)$$

The update rule of matrix H is given by Eq.(11).

$$H_{aj} \leftarrow H_{aj} \frac{(W^T V)_{aj}}{(W^T WH)_{aj}} \quad (11)$$

The update rule of matrix W is given by Eq.(12).

$$W_{ia} \leftarrow W_{ia} \frac{(VH^T)_{ia}}{(VHH^T)_{ia}} \quad (12)$$

Our signal classification process is divided into the following steps:

Step 1: Filtering the original signal X through an 8-25 band-pass filter to obtain the signal of the frequency band related to the characteristic, denoted as E .

Step 2: Sparsification of E by Eqs. (1)-(7), and the matrix V after sparsification was obtained.

Step 3: Divided the V into two matrices, denoted as V_{C3} and V_{C4} . The basis matrix (W_{C3} and W_{C4}) and the weight matrix (H_{C3} and H_{C4}) of the signal are obtained by Eqs. (11) and (12).

Step 4: Clustering of H_{C3} and H_{C4} .

3 Experimental results

In this section, we investigate the use of our proposed NMF algorithm for data clustering. Several experiments are carried out to show the effectiveness of our algorithm for motor imaging EEG classification. For our empirical study, 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 [17]. The data set involves left-right imagery hand movements and consists of 140 labeled trials for training and 140 unlabeled trials for test. Each trial has a duration of 9 seconds, where a visual cue (arrow) is presented pointing to the left or the right after 3-second preparation period and imagination task is carried out for 6 seconds. It contains EEG acquired from three different channels (with sampling frequency 128 Hz) C3, Cz and C4. In our study we use only two channels, C3 and C4, because ERD has contralateral dominance and Cz channel contains little information for discriminant analysis.

3.1 Performance evaluation and comparisons

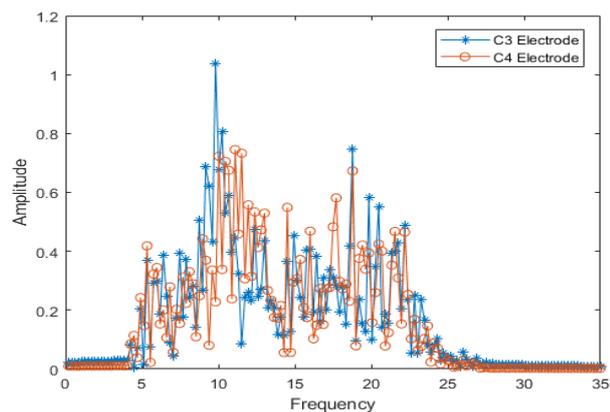
In this paper, we use a sparsification measure based on the relationship between the C3 and the C4 channel. We compare the motor imaging EEG with the maximization of feature difference to those without the maximization of feature difference, as shown in the Fig.1.

Divide the data with the maximization of feature difference into two matrices, V_{C3} contains the signal of the C3 channel of the left and right hand motion imaging signal, V_{C4} contains the signal for C4 channel, V_{C3} and V_{C4} , respectively, through the NMF feature extraction and classification, inferred accuracy of their classification.

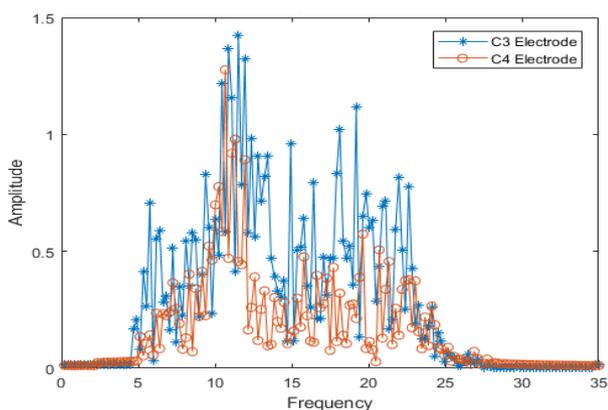
The test set in the dataset was extracted and classified separately by CSP and NMF, and compared with the method we proposed, as shown in Table 1.

Table 1. Classification accuracy.

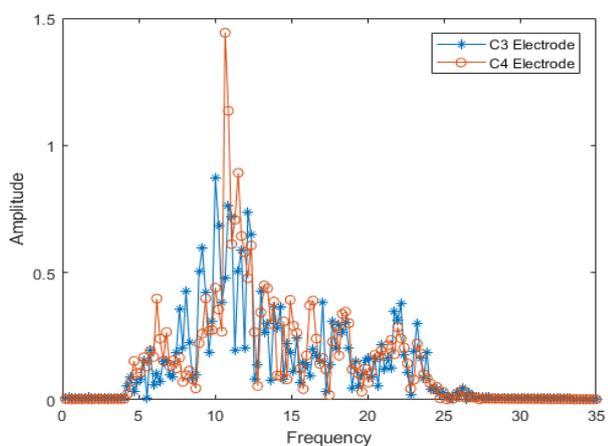
Method Accuracy	Left-Hand	Right-Hand	Average
CSP	85.71%	82.86%	84.29%
NMF	87.14%	85.71%	86.42%
S+NMF	94.28%	88.57%	91.43%



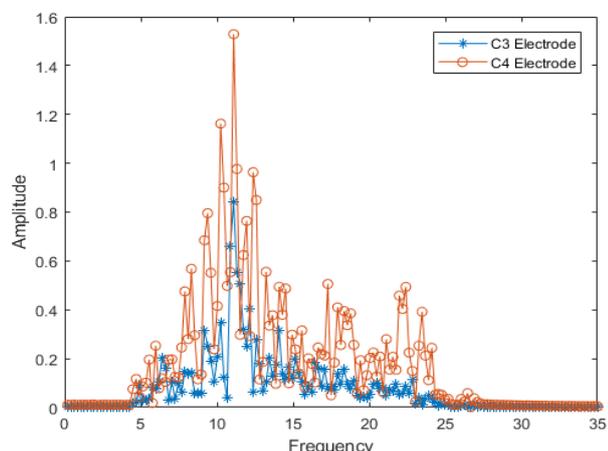
(a) Motor imagery on left-hand



(b) Motor imagery on left-hand after sparsification



(c) Motor imagery on right-hand



(d) Motor imagery on right-hand after sparsification

Figure 1. The signal after sparsification comparison.

To show the data classification performance, we compare our algorithm with other related methods on BCI competition 2003 data sets. The detailed classification accuracy and comparison shown in Table 1. As we can see, when only CSP and NMF were used, the average classification accuracy of the NMF was 2.13% higher than the CSP. And the accuracy of NMF classification was improved by 5.01% after the sparsification of data.

4 Conclusion

In this paper, we propose a non-negative matrix factorization method to classify motor imagination EEG. Based on the sparse degree of matrix and the selection of dimension will affect the efficiency of the algorithm and the accuracy of the classification. Firstly, the method of maximization of feature difference is used to increase the sparsity of the matrix; Secondly, the data of training set and test set of BCI competition in 2003 data are respectively classified by our method. Experiment shows: the average classification accuracy of test set was 91.43%, the highest single class motion imaging accuracy reached 94.28%.

Acknowledgments

This work was supported in part by the Guangdong program for support of top-notch young professionals under grant 2014TQ01X144, the National Natural Science Foundation of China under grants 61722304, 61333013, the Pearl River S&T Nova Program of Guangzhou under grant 201610010196, the Guangdong Natural Science Funds for Distinguished Young Scholar under grant 2014A030306037.

References

1. A. Karagiannis, P. Constantinou, "Noise-Assisted Data Processing With Empirical Mode Decomposition in Biomedical Signals." IEEE Engineering in Medicine & Biology Society, 2011, 15(1):11-8.
2. M. Boersma, D. Smit, H. Bie, et al. "Network analysis of resting state EEG in the developing young brain: structure comes with maturation." Human Brain Mapping, 2015, 32(3):413-425.
3. E. Kalunga, S. Chevallier, Q. Barthélemy, et al. "Online SSVEP-based BCI using Riemannian geometry". Neurocomputing, 2016, 191:55-68.
4. E. Lukhanina, N. Mel'Nik, N. Berezetskaya, et al. "Co-relations between indices of P300 EEG potential, cognitive tests, and variational pulsometry in Parkinsonian patients." Neurophysiology, 2008, 40(1):39-47.
5. S. Pittaccio, F. Zappasodi, S. Viscuso, et al. "Primary sensory and motor cortex activities during voluntary and passive ankle mobilization by the SHADE orthosis." Human Brain Mapping, 2011, 32(1):60-70.

6. A. Bamdadian, C. Guan, K. Kai, et al. "Towards improvement of MI-BCI performance of subjects with BCI deficiency." International Ieee/embs Conference on Neural Engineering. IEEE, 2015:17-20.
7. S. Huang, X. Wu. "Feature Extraction of EEG Signals Based on Mu/Beta Rhythm Imaging." Guangzhou: School of Bioscience and Engineering, South China University of Technology, 2010.
8. Çınar, Salim, N. Acır . "A novel system for automatic removal of ocular artefacts in EEG by using outlier detection methods and independent component analysis". Pergamon Press, Inc. 2017.
9. D. Toledo, G. Manzano, J. Barela, et al. "Cortical correlates of response time slowing in older adults: ERP and ERD/ERS analyses during passive ankle movement." 2016, 127(1):655-663.
10. L. Zhao, H. Shen, S. Cui, et al. "Research of Brain-Computer Interface Technology Based on Labview." IEEE Computer Society, 2008:72-80.
11. H. Sun, Y. Xiang, Y. Sun, et al. "On-line EEG classification for brain-computer interface based on CSP and SVM." International Congress on Image and Signal Processing. IEEE, 2010:4105-4108.
12. N. Zhang, X. Tang, Q. Liu. "Feature Extraction and Recognition of EEG Based on Semi-supervised Learning." Engineering Science Edition, 2017(s2): 230-237.
13. H. Lee, A. Cichocki, S. Choi. "Nonnegative Matrix Factorization for Motor Imagery EEG Classification." Springer Berlin Heidelberg, 2006:250-259.
14. M. Liu, J. Wang, C. Zheng, et al. Attention-related EEG features extracted by nonnegative matrix factorization [J]. Chinese Journal of Biophysics, 2006,22 (1): 67-72.
15. S. Duan, Y. Shang, L. Pan. "Feature Extraction and Classification of EEG Signals in Multi-Kinematics Imaging." Computer Measurement and Control, 2016,24 (2): 283-287.
16. D. D. Lee, H. S. Seung. "Algorithms for nonnegative matrix factorization." Advances in Neural Information Processing Systems, 2001, 13(6):556--562.
17. B. Blankertz, K. Müller, G. Curio, et al. "The BCI Competition 2003: progress and perspectives in detection and discrimination of EEG single trials." IEEE Trans. biomedical engineering, 2004, 51(6):1044-51.