Nonnegative Matrix Factorization for EEG Signal Classification

Weixiang Liu, Nanning Zheng, and Xi Li

Institute of Artificial Intelligence and Robotics
Xi'an Jiaotong University
Xi'an,Shaanxi Province 710049 P.R.China
{wxliu,xli}@aiar.xjtu.edu.cn, nnzheng@mail.xjtu.edu.cn

Abstract. Nonnegative matrix factorization (NMF) is a powerful feature extraction method for nonnegative data. This paper applies NMF to feature extraction for Electroencephalogram (EEG) signal classification. The basic idea is to decompose the magnitude spectra of EEG signals from six channels via NMF. Primary experiments on signals from one subject performing two tasks show high classification accuracy rate based on linear discriminant analysis. Our best results are close to 98% when training data and testing data from the same day, and 82% when training data and testing data from different days.

1 Introduction

Recognition and classification of Electroencephalogram (EEG) signals for Brain-Machine Interfaces (BCI) bring out many open problems from both theory and application; see [15] for a brief survey on the topic and [14] for more information. From the viewpoint of pattern recognition [1], one key problem is how to represent the recorded EEG signals for further analysis such as classification. In other words, it is, firstly, important to extract useful features from the EEG signals. There are many available representations from both time domain and frequency domain, such as AR model coefficients, Karhunen-Loéve transform [13] and maximum noise fraction (MNF) transform [15].

Recently nonnegative matrix factorization (NMF) [4], as a powerful feature extraction method for nonnegative data, has been successfully applied to music analysis [6,10,18,19], document clustering and information retrieval [7,8,20,21, 11,17], gene expression [9,22] and molecular analysis [23]. Inspired by the idea from above music analysis with magnitude spectra, we try to apply NMF to feature extraction for EEG signal classification. The basic idea is to decompose the magnitude spectra of EEG signals from six channels via NMF. We made our experiments on signals from one subject performing two tasks in two days and the results show high classification rate based on linear discriminant analysis. Our best results are approximately 98% when training data and testing data from the same day, and 82% when training data and testing data from different days.

F. Yin, J. Wang, and C. Guo (Eds.): ISNN 2004, LNCS 3174, pp. 470–475, 2004.

The rest of this paper is organized as follows. In section 2 we recall some basic theory of NMF. In section 3 we make our experiments and we give our conclusions and discuss some future work in section 4.

2 Nonnegative Matrix Factorization

NMF is a multivariate data analysis method. Given a nonnegative data set with m samples in R^d, denoted as $X_{d\times m}$, NMF finds two nonnegative matrix factors B and C (i.e. each element of B and C is nonnegative, denoted by $B \geq 0, C \geq 0$) such that [5]

$$X \approx BC \tag{1}$$

where B is a $d \times r$ basis matrix, each column of which is a basis vector, and C a $r \times m$ coefficient matrix, each column of which is a new feature vector. It leads to dimension reduction by choosing r smaller than d although it is an open problem to decide the optimal r. Two kinds of cost functions have been investigated [5] with multiplicative update rules which naturally preserve nonnegativity, i.e. the generalized Kullback-Leibler divergence

$$D_{KL}(B,C) = \sum_{i=1}^{d}\sum_{j=1}^{m}[X_{ij}\log\frac{X_{ij}}{(BC)_{ij}} - X_{ij} + (BC)_{ij}] \tag{2}$$

and the square Euclidean distance

$$D_2(B,C) = \sum_{i=1}^{d}\sum_{j=1}^{m}[X_{ij} - (BC)_{ij}]^2. \tag{3}$$

In this paper we adopt the multiplicative rules for minimizing eq. (2) as below [4]

$$C_{kj} \leftarrow C_{kj}\sum_i B_{ik}\frac{X_{ij}}{(BC)_{ij}} \tag{4}$$

$$B_{ik} \leftarrow B_{ik}\sum_j \frac{X_{ij}}{(BC)_{ij}}C_{kj} \tag{5}$$

$$B_{ik} \leftarrow \frac{B_{ik}}{\sum_l B_{lk}}. \tag{6}$$

For testing samples, it is convenient to get new features according to eq. (4) while fixing the learned basis matrix from training data.

With contrast to traditional principal component analysis (PCA) [3] and recent independent component analysis (ICA) [2], NMF has parts- based representation property because of nonnegativity [4].

NMF restricts its data with only nonnegativity. However, when we apply NMF for data analysis in magnitude spectra, it is not necessary to require that the source data is nonnegative. Some cases of NMF for music analysis [6,10,18, 19] provide promising results which inspire us to classify EEG signals via this method for representation.

As a new developed feature extraction method, NMF still has many open problems for investigation. See our technical report [12] for more information.

3 Experimental Results

3.1 EEG Data

The EEG data used here is available online[1], which has been discussed in detail [15,16]. We used the EEG signals from subject 1 performing two mental tasks, i.e. math multiplication and letter composing, in two days. Each tasks contains 5 trials on one day and another 5 trials on second day. Each trial is a 6×2500 matrix from six channels.

3.2 EEG Data Representation

Following [15], we first segmented each serial with 2500 times samples into 38 augmented data samples according to the procedure as below: each window has 128 time samples that overlap by 64 samples. Then we implemented discrete Fourier transform on each augmented sample and got the first 65 absolute versions of magnitude spectra. Finally each trial becomes a 38 data samples with 65×6 dimensions which can be reduced by NMF.

3.3 Classification Results

For classification, we adopted the discriminant analysis function *classify* in MATLAB[2] with linear method which fits a multivariate normal density to each group based on training data with a pooled estimate of covariance. We made three different initialization cases for dimension reduction by NMF and set $r = 6, 12, 18, \ldots, 66$ for comparison when testing the method.

Firstly we used the EEG signals from subject 1 on the first day. We used the first trial for training and the rest 4 trials for testing. The final results are shown in Table 1. We can see that from above table, different initializations and rs for NMF leads to different classification rate; the average accuracy of two tasks can get 98%. Our results are higher than those in [15] (90%) although we used 128 time samples per segmentation with overlap.

We also used the EEG signals from subject 1 from different days. We used the first trial on the first day for training and all 5 trials on the second day for testing. The final results are shown in Table 2. According to the average accuracy rate, the proposed method can get 82% or so as the best . Our results are higher than 75% reported in [15].

[1] http://www.cs.colostate.edu/eeg/index.html#Data
[2] http://www.mathworks.com/

Table 1. EEG signal classification results when training data and testing data from one day.

r	math	letter	average	r	math	letter	average	r	math	letter	average
6	0.770	0.875	0.822	6	0.829	0.868	0.849	6	0.770	0.855	0.813
12	0.822	0.914	0.868	12	0.921	0.928	0.924	12	0.895	0.928	0.911
18	0.914	0.954	0.934	18	0.947	0.928	0.938	18	0.908	0.895	0.901
24	0.947	0.895	0.921	24	0.895	0.934	0.914	24	0.908	0.895	0.901
30	0.954	0.947	0.951	30	0.974	0.947	0.961	30	0.987	0.934	0.961
36	0.928	0.941	0.934	36	0.928	0.908	0.918	36	0.947	0.954	0.951
42	0.961	0.961	0.961	42	0.993	0.941	0.967	42	0.934	0.928	0.931
48	0.947	0.914	0.931	48	**0.980**	**0.961**	**0.970**	48	0.934	0.941	0.938
54	0.921	0.967	0.944	54	0.947	0.961	0.954	54	0.921	0.928	0.924
60	0.987	0.954	0.970	60	0.974	0.947	0.961	60	0.947	0.914	0.931
66	**0.987**	**0.967**	**0.977**	66	0.908	0.868	0.888	66	**0.987**	**0.941**	**0.964**

Table 2. EEG signal classification results when training data and testing data from different days.

r	math	letter	average	r	math	letter	average	r	math	letter	average
6	0.647	0.642	0.645	6	0.674	0.489	0.582	6	0.695	0.516	0.605
12	0.905	0.542	0.724	12	0.884	0.695	0.789	12	0.863	0.616	0.739
18	0.784	0.737	0.761	18	0.911	0.647	0.779	18	0.932	0.595	0.763
24	0.858	0.637	0.747	24	0.789	0.784	0.787	24	0.889	0.674	0.782
30	**0.874**	**0.732**	**0.803**	30	**0.911**	**0.689**	**0.800**	30	0.900	0.579	0.739
36	0.895	0.653	0.774	36	0.842	0.716	0.779	36	0.847	0.574	0.711
42	0.868	0.711	0.789	42	0.889	0.647	0.768	42	0.895	0.674	0.784
48	0.921	0.563	0.742	48	0.853	0.679	0.766	48	**0.884**	**0.758**	**0.821**
54	0.879	0.632	0.755	54	0.916	0.537	0.726	54	0.884	0.653	0.768
60	0.895	0.511	0.703	60	0.821	0.753	0.787	60	0.853	0.716	0.784
66	0.916	0.621	0.768	66	0.847	0.526	0.687	66	0.900	0.668	0.784

Our results also indicate that it is difficult to select optimal r while reducing the dimension via NMF, which is an open problem for NMF [12].

4 Conclusions and Future Work

In this paper we first apply NMF to feature extraction for EEG signal classification. The basic idea is to decompose the magnitude spectra of EEG signals from six channels via NMF. We made our experiments on signals from one subject performing two tasks in the same day and the primary results show high classification rate based on linear discriminant analysis. Our best results are close to 98% when training data and testing data from the same day, and 82% when training data and testing data from different days.

Our results for EEG signal classification are promising. There are some directions for future work. Firstly we will further analyze the EEG signals based on NMF for several tasks. Secondly it is necessary to compare the proposed method with other methods such as MNF in [15], or with advanced classifiers such as support vector machines in [16]. Thirdly, it is possible to consider other time-frequency domain such as via Discrete Cosine Transform and Discrete Wavelet Transform. And finally, or the most important, it is potential to implement the method for real time analysis.

Acknowledgments. This work was supported by the NSF of China (No. 60205001). We would like to thank Lianyi Zhang for his introduction of the EEG data set and fruitful discussions.

References

1. Duda, R.O., Hart, P.E., Stork, D. G.: Pattern Classification. 2nd edn. John Wiley & Sons (2001)
2. Hyvärinen,A., Karhunen, J., Oja, E.: Independent Component Analysis. John Wiley & Sons (2001)
3. Jolliffe, I.T.: Principal component analysis. 2nd edn. Springer-Verlag, New York (2002)
4. Lee, D.D., Seung, H.S.: Learning the parts of objects with nonnegative matrix factorization. Nature **401** (1999) 788-791
5. Lee, D.D., Seung, H.S.: Algorithms for nonnegative matrix factorization.In: Leen, T., Dietterich, T., Tresp, V. (eds.): Advances in Neural Information Processing Systems 13. MIT Press, Cambridge, MA (2000)
6. Kawamoto, T., Hotta,K., Mishima,T., Fujiki, J., Tanaka, M., Kurita, T.: Estimation of Single Tones from Chord Sounds Using Non-Negative Matrix Factorization. Neural Network World **3** (2000) 429-436
7. Vinokourov, A.: Why Nonnegative Matrix Factorization Works Well For Text Information Retrieval. http://citeseer.nj.nec.com/458322.html.
8. Tsuge, S., Shishibori,M., Kuroiwa, S., Kita, K.: Dimensionality reduction using non-negative matrix factorization for information retrieval. In: IEEE International Conference on Systems, Man, and Cybernetics, Vol. 2. (2001) 960 -965
9. Seppänen,J. K. , Hollmén, J., Bingham, E., Mannila, H.: Nonnegative matrix factorization on gene expression data. Bioinformatics 2002, poster 49. (2002)
10. Smaragdis,P., Brown, J.C.: Non-negative matrix factorization for polyphonic music transcription. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. (2003) 177-180
11. Pauca,P., Shahnaz, F., Berry, M., Plemmons, R.: Text Mining using Nonnegative Matrix Factorizations. In: Proc. SIAM Inter. Conf. on Data Mining. (2003)
12. Liu, W.X., Zheng, N.N., Li, X.: Review on Nonnegative Matrix Factorization. Technical report, Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University. (2004)
13. Anderson, C., Devulapalli, S., Stolz, E.: EEG Signal Classification with Different Signal Representations. In: Girosi, F., Makhoul, J., Manolakos, E., Wilson, E. (des.): Neural Networks for Signal Processing V. IEEE Service Center, Piscataway, NJ. (1995) 475–483

14. Vaughan, T.M., Heetderks, W.J., Trejo, L.J., Rymer, W.Z., Weinrich, M., Moore, M.M., Kübler, A., Dobkin, B.H., Birbaumer, N., Donchin, E., Wolpaw, E.W. and Wolpaw, J.R.: Brain-computer interface technology: A review of the Second International Meeting. IEEE Transactions on Neural Systems & Rehabilitation Engineering **11** (2003) 94-109
15. Anderson, C.W., Kirby, M.: EEG Subspace Representations and Feature Selection for Brain-Computer Interfaces. In: Proceedings of the 1st IEEE Workshop on Computer Vision and Pattern Recognition for Human Computer Interaction (CVPRHCI). (2003)
16. Garrett, D., Peterson, D.A., Anderson, C.W., Thaut, M.H.: Comparison of Linear and Nonlinear Methods for EEG Signal Classification. IEEE Transactions on Neural Systems and Rehabilitative Engineering **11** (2003) 141-144
17. Xu,W., Liu,X., Gong, Y.H.: Document clustering based on non-negative matrix factorization. In: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval. (2003) 267 - 273
18. Plumbley,M. D. ,Abdallah,S. A., Bello,J. P., Davies, M. E., Monti,G., Sandler,M. B.: Automatic music transcription and audio source separation. Cybernetics and System **33** (2002) 603-627
19. Plumbley,M. D.: Algorithms for non-negative independent component analysis. IEEE Transactions on Neural Networks **14** (2003) 534- 543
20. Lu,J.J., Xu,B.W., Yang, H.J.: Matrix dimensionality reduction for mining web logs. In: Proceedings of IEEEWIC International Conference on Web Intelligence. (2003) 405 - 408
21. Xu,B.W., Lu,J.J.,Huang, G.S.: A constrained non-negative matrix factorization in information retrieval. In: IEEE International Conference on Information Reuse and Integration. (2003) 273 - 277
22. Kim,P.: Understanding Subsystems in Biology through Dimensionality Reduction, Graph Partitioning and Analytical Modeling. Phd thesis . (2003)
23. Brunet, J.P., Tamayo, P., Golub, T.R., Mesirov, J.P.: Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci U. S. A. **101** (2004) 4164-4169