1

Mixture factor analysis based probabilistic model for learning CSP and its application on EEG signal

Chao Huang, Haixian Wang, *Member, IEEE,*

***Abstract*—Common spatial pattern (CSP) is a prevalent method applied in brain-computer interfaces (BCI) system due to its high efficiency and accuracy. The probabilistic framework of CSP has been proposed for electroencephalogram (EEG) single-trial classification in recent years. Based on some criterions, CSP could be recognized in a probabilistic point of view. In this paper, we propose a promoted probabilistic CSP method, named Mixture Factor Analysis based Probabilistic model for learning CSP (MFAPCSP). With the help of expectation maximization (EM) method, we solve MFAPCAP and conduct binary classification especially for a small size training dataset. The experimental result on BCI competition dataset demonstrates the effectiveness of the proposed method.**

***Index Terms*—CSP, Possibility framework, Mixture component, EM.**

I. Introduction

NONINVASIVE brain-computer interface (BCI) is a prevalent machine learning technique which provides a new way for human beings especially the patient who suffers some kind of motor neuron disease to improve their daily lives. The objective of BCI system is to record and analyze the signals generated by our brain directly, such as electroencephalogram (EEG). One of the most typical BCI systems is the motor imagery system (MIs). MIs focuses on the discrimination between left and right hand motor imagery via single-trial EEG data analysis. In the mean while, it should be noted that there are some special features for the EEG motor imagery task. Processing of motor imagery commands or some intended movements in motor imaginary causes a decrease of the EEG activity named event-related desynchronization (ERD) while the increase one termed event related synchronization (ERS)[1].

In past decades, lots of algorithms have contributed to perfect BCI system, such as event-related brain potentials (ERP)[2], spatio-spectral filters [3][4] and principal component analysis [5]. Based on ERD and ERS, common spatial patterns (CSP) has been proposed [1], and its relevant algorithms become a prevalent technique which could solve the BCI decoding problem efficiently. CSP, known as Fukunaga-Koontz transformation, was created as an extension of PCA for feature extraction, and had shown its effectiveness in discriminating two classes of EEG data through maximizing the variance of one class and minimizing the other side [6]. Recently, many algorithms promoted traditional CSP and made it more efficient and robust [7].

With the help of factor analysis (FA) which is an extension of PCA [5], CSP can also be achieved in a probabilistic viewpoint. Wu proposed and demonstrated that FA could get a general result compared to classic CSP [7]. Similar to Gaussian mixture model (GMM), mixture factor analysis (MFA) contributes a more accurate and robust framework than a single component FA. Therefore, the MFA based probabilistic framework for CSP is expected to be superior to CSP in decoding brain activities.

In this paper, we propose Mixture Factor Analysis based Probabilistic model for learning CSP (MFAPCSP). Compared to the traditional CSP, MFAPCSP is expected to lead to better classification accuracy especially in the case of very small sample size. In next section, we will briefly review the classical method of CSP. And then, we will focus on the mixture probabilistic model for learning CSP (detailed procedure will be shown in Appendix A). Finally, the experimental results will be presented and discussed.

II. Method

A. Classicall CSP

Two classes of EEG dataset were recorded as $Y_k^{(i)} \in R^{C\times M_i} (i = 1, 2)$ ($Y_k^{(i)}$ is the k^{th} trial trial for class i) where C and M_i refer to the number of channels and the number of sampling points respectively. Let $R^{(i)}$ denote the covariance matrix of the dataset belongs to each class i, i.e.,

$$R^{(i)} = \frac{1}{M_i}\sum_{k=1}^{M_i} Y_k^{(i)} Y_k^{(i)T} \tag{1}$$

CSP is a classic method which aims to solve the BCI related problem especially for motor imaginary task. The key function of CSP is maximizing the covariance of one class, and at the same time, minimizing that of the other class [1]. From mathematical aspect, it is equal to a maximization ratio function given by

$$max_w \frac{w^T R^{(1)} w}{w^T R^{(2)} w} \ s.t.||w|| = 1 \tag{2}$$

Using Lagrangian multiplier method, equation (2) could be changed to another form given by (3), where is the eigenvector

Chao Huang is the master candidate of Biomedical Engineering Department in SEU of Biomedical Engineering, Southeast University, Nanjing, 210096 P. R. C. e-mail: oliverhuangchao@163.com

matrix ($W = [\omega_1, \omega_2, ...\omega_C]$) and λ is a diagonal matrix of the corresponding eigenvalues. Normally, the spatial filters refer to the eigenvectors which correspond to the few largest and smallest eigenvalues. Equation (3) could be solved by the generalized eigenvalue decomposition

$$R^{(1)}W = R^{(2)}W\Lambda \tag{3}$$

Given W, the feature for $Y_k^{(i)}$ would be written as

$$f_k^{(i)} = diag(W^T Y_k^{(i)} Y_k^{(i)T} W) \tag{4}$$

B. Probabilistic framework for learning CSP

The basic probabilistic framework for CSP is given in [8]:

$$X_k^{(i)} = AZ_k^{(i)} + \Xi_k^{(i)} \tag{5}$$

$$Z_k^{(i)} \sim (0, \Lambda^{(i)}), \Xi_k^{(i)} \sim N(0, \varphi^{(i)}), i = 1, 2$$

Unlike classical CSP, in this framework, $X_k^{(i)} \in R^{C\times 1}$ is the k^{th} observed sample for class i which is linearly constructed by $Z_k^{(i)}$, $\Xi_k^{(i)}$ and A. $Z_k^{(i)}$ refers to a latent factor that is assumed to be Gaussian distributed with zero mean and $\Lambda^{(i)}$ covariance for each class i. $\Xi_k^{(i)}$ can also be regarded as the a specific factor which is assumed to be a zero mean and $\varphi^{(i)}$ covariance Gaussian distribution. A represents the factor loading matrix. Wu demonstrated that, based on the following conditions, CSP could be expressed in a probabilistic way [8]:

1) W should equal to A^{-T}
2) $\Lambda^{(i)}$,$\varphi^{(i)}$are all diagnosed matrixes
3) The specific factor perishes to zero

C. Mixture Factor Analysis based Probabilistic model for learning CSP (MFAPCSP)

Based on the above probabilistic framework, we extend it further and propose a new mixture probabilistic model for CSP. Due to the fact that all the samples are centralized and filtered through a band pass filter (5∼33Hz), we can define the basic framework as:

$$X_k^{(i)} = \sum_{j=1}^{M} \pi_j A_j Z_{jk}^{(i)} + \Xi_K^{(i)} \tag{6}$$

where $Z_{jk}^{(i)} \sim N(0, \Lambda_j^{(i)}), \Xi_k^{(i)} \sim N(0, \varphi^{(i)}), i = 1, 2$. M is the number of the mixture components, and π_j refers to weighs of component j where $\pi_1 + \pi_2 + ... + \pi_M = 1$. The model parameters $\Theta = \{A_j, \Lambda_j^{(i)}, \varphi^{(i)}, \pi_j^{(i)}\}, (j = 1, 2, ...m; i = 1, 2)$ are calculated by the maximum likelihood estimation. The maximum likelihood function is defined as follows

$$Q = E(\log \prod_k \prod_j \prod_i p(X_k^{(i)}, Z_{jk}^{(i)}, \omega_j^{(i)})^{\omega_j^{(i)}}) \tag{7}$$

Following the classic EM-steps for mixture factor analysis [9], a specific EM algorithm for our mixture probabilistic framework for CSP could be obtained in the following steps:

Methods		
Record	Record the EEG data and the data preprocessing	
Initialization	Parameters	Initialization
	$\varphi^{(i)}$	unit matrix
	A_j	inverse transpose of matrix W calculated by classical CSP
	$\Lambda_j^{(i)new}$	WRW^T, where R refers to the covariance matrix of the EEG data
	$\pi_j^{(i)}$	empirical parameters which will be discussed in the later section
E-Step	Get supervise data $X_k^{(i)}$ and Compute $h_jk(i)$, $E(z_{jk}^{(i)})$ and $E(z_{jk}^{(i)} z_{jk}^{(i)T})$ for each mixture component	
M-Step	Solve the equation listed below(10)	
End	Calculate Q shown in (8) for each EM iteration (9) to (12) until $\left\|\frac{Q^n}{Q^{n-1}}\right\| - 1 < 10^{-6}$	

$$\begin{aligned}\varphi^{(i)new} =& \frac{1}{n} \sum_{j,k} h_{jk}^{(i)} [X_k^{(i)} X_k^{(i)T} - 2X_k^{(i)} E(z_{jk}^{(i)})^T A_j^T \\ &+ A_j E(z_{jk}^{(i)} z_{jk}^{(i)T})^T A_j^T]\end{aligned} \tag{8}$$

$$\Lambda_j^{(i)new} = \frac{1}{\sum_k h_{jk}^{(i)}} \sum_k h_{jk}^{(i)} E(z_{jk}^{(i)} z_{jk}^{(i)T})^T \tag{9}$$

Factor loading matrix A could be calculated through a Lyapunov equation

$$\begin{aligned}&\varphi_{(1)}^{-1} A_j \sum_k h_{jk}^{(1)} E(z_{jk}^{(1)} z_{jk}^{(1)T})^T + \varphi_{(2)}^{-1} A_j \sum_k h_{jk}^{(2)} E(z_{jk}^{(2)} z_{jk}^{(2)T})^T = \\ &\varphi_{(1)}^{-1} A_j \sum_k h_{jk}^{(1)} X_k^{(1)} E(z_{jk}^{(1)})^T + \varphi_{(2)}^{-1} A_j \sum_k h_{jk}^{(2)} X_k^{(2)} E(z_{jk}^{(2)})^T\end{aligned} \tag{10}$$

$$\pi_j^{(i)new} = \frac{1}{n} \sum_k h_{jk}^{(i)} \tag{11}$$

$$h_{jk}^{(i)} = \frac{\pi_j^{(i)} N(0, A_j \Lambda A_j^T + \varphi^{(i)})}{\sum_j \pi_j^{(i)} N(0, A_j Lambda A_j^T + \varphi^{(i)})} \tag{12}$$

In the end, we can get M components loading matrix A_j through MFAPCSP and each of them represents a spatial filter

(The whole procedure could be seen as 1). At last, we can integrate a $M \times C$ dimension feature f_k for X_k as (12):

$$f_k = \begin{bmatrix} f_k^1 \\ f_k^2 \\ \dots \\ f_k^M \end{bmatrix} f_k^j = diag(A^{-1} X_k X_k^T A^{-T}) \qquad (13)$$

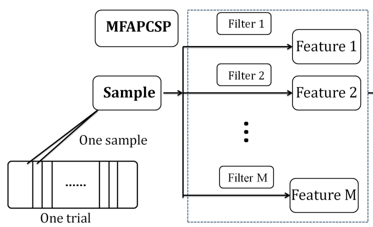

Fig. 1. M components construct the whole probabilistic framework in which each component works as a spatial filter independently. The weigh for each component j is π_j and $\pi_1 + \pi_2 + ... + \pi_M = 1$. Every filter projects the sample to its corresponding feature. In this frame work, one sample refers a column of its trail.

III. Experiment

The performance of MFAPCSP proposed in this paper is evaluated by calculating the classification accuracy on a motor imagery BCI datasets supplied by BCI Competition III dataset IVa [10] collected from five healthy subjects. The visual cues are indicated for 3.5 s which includes three motor imagery blocks: left hand, right hand and right foot. In our experiment, only left and right hand motor imagery data are used. All the datasets contain 280 trials, but different for their number of the training trials. All the signals are recorded through a 118-channel Ag/AgCl electrodes cap which is measured at positions of the extended international 10-20 system.

Compared to the original data recorded from 118 channels, it is more efficient to select ten channels (Fig. 2, AF3, AF4, F1, F2, F3, F4, F5, F6, FC3, and FC4) to finish the further analysis. After that, all the samples will be filtered by a band-pass filter (5∼33Hz) and centralized before classification. The detail of this experiment is shown in Fig. 2. It is really a difficult task for MFAPCSP to choose its best number of mixture components and their initial weighs. Lots of researches tried to solve this problem and nearly all of them focused on the GMM [11][12][13] or Bayes method [14]. However, to my short knowledge, little of them focus on the mixture factor analysis model. In this paper, we try to construct a probabilistic framework based on mixture factor analysis and choose $\{0.3, 0.7\}$ as an empirical Bayes initial component weigh. The results for other possible combination weighs could be seen in Fig. 3.

As we all know, the smaller size training dataset always, the worse classification result will be. Therefore, many Regularized Common Spatial Patterns (RCSP) [7] methods which aim to improve the accuracy of CSP classification were created and added to (3). In our experiment, we calculate and compare the classification accuracy between two methods (classical CSP vs. MFAPCSP) under the same condition (the same training samples). The result and its discussion will be shown in the following section.

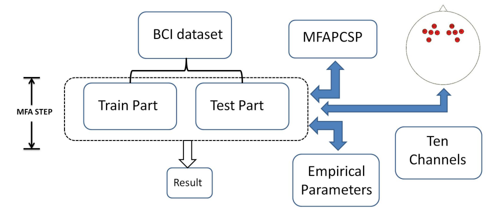

Fig. 2. Probability CSP experiment procedure. Experimental schema for BCI competition data is shown above. For every training group, we will finish MFA step five times and average their accuracy result as the final result. The size of training group ranges from tow to twenty

IV. Result and Discussion

As shown in IV, different components' combinations will lead to different results. Obviously, in this experiment, the results of two components MFAPCSP will be superior if the weigh combination for each class is chosen between 0.2 and 0.3 (e.g. 0.2 means the combination weigh is $\{0.2, 0.8\}$).

We could observe all the results from five subjects in

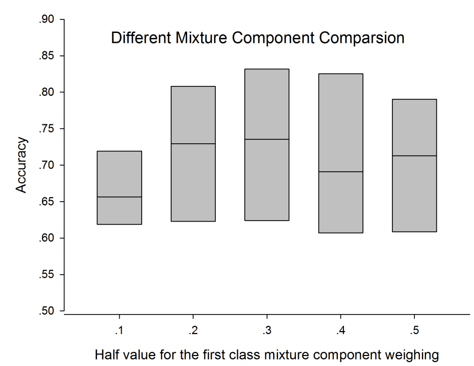

Fig. 3. A Comparison among different weighing for each mixture components for all the five subjects. The X-axis represents the half of the value of component combinations, there is no difference between $\{0.1, 0.9\}$ and $\{0.9, 0.1\}$ and same to the similar combination). It is obvious phenomenon that combinations between $\{0.2\ 0.8\}$ and $\{0.3, 0.7\}$ will lead to a better result compare to other pairs.

Fig. 4 to Fig. 8. As the framework structure described above shows 1, for MFAPCSP, $X_k^{(i)}$ represents a sample for every single trial rather than the whole trial used in classic CSP. So that, it will provide more samples to the training procedure and thus contributes to a higher accuracy than classical CSP. That is a big advantage for MFAPCSP when it tackles with classification work whose training group size is small. For this reason, in our experiment, it is clear that when the size of training sample is small (less than six trials), classical CSP lacks the ability to finish its job while MFAPCSP (the weigh of each class is $\{0.3, 0.7\}$) is much more competitive in the same circumstance. As the number of the training samples goes higher, the result of MFAPCSP

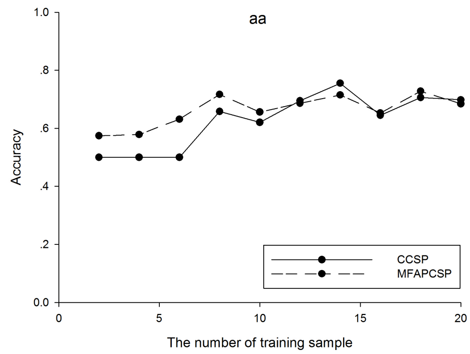

Fig. 4. Small size training datasets result for BCI competition III data IVa using {0.3,0.7} as the component initial weigh. The results for aa subject

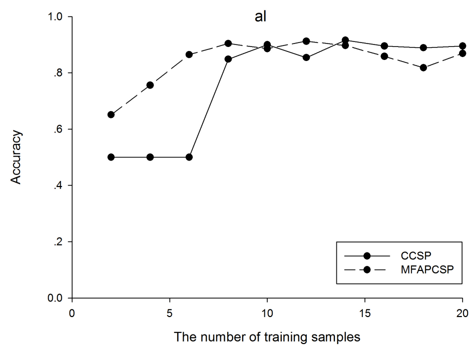

Fig. 5. Small size training datasets result for BCI competition III data IVa using {0.3,0.7} as the component initial weigh. The results for al subject

will be similar to classical CSP.

However, there are still some problems should be addressed in future studies. First, we still lack a suitable and efficient method to determine the number and weigh of each mixture component. In addition, the convergence rate for the MFAPCSP is still low and it should be improved when it comes to the daily life application. More importantly, MFAPCSP is not robust comparing to some methods proposed in RCSP [7]. A combination of these two algorithms might be much better.

V. Conclusion

A new method has been proposed and tested in this paper. Based on the previous studies, we had improved classical CSP and proposed MFAPCSP. Using the data supported by the 2005 BCI competition, we find that though there are also some problems which need to be addressed in the future such as the algorithm robust ability, efficiency, MFAPCSP is more suitable and accurate than classical CSP in a small size training dataset classification work. Therefore, further experiments and improvements are necessary to be done.

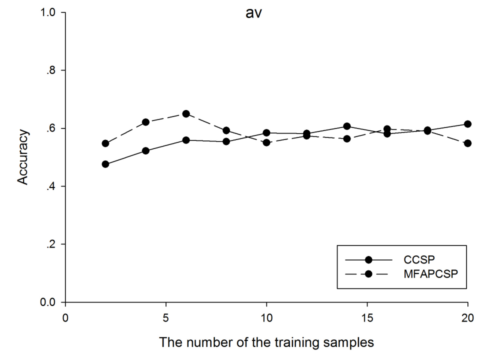

Fig. 6. Small size training datasets result for BCI competition III data IVa using {0.3,0.7} as the component initial weigh. The results for av subject

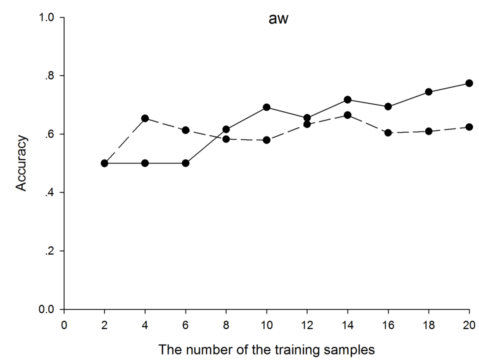

Fig. 7. Small size training datasets result for BCI competition III data IVa using {0.3,0.7} as the component initial weigh. The results for aw subject

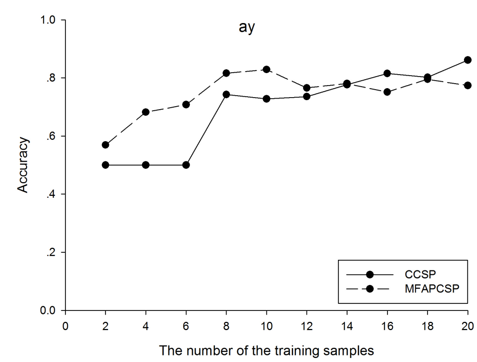

Fig. 8. Small size training datasets result for BCI competition III data IVa using {0.3,0.7} as the component initial weigh. The results for ay subject

References

[1] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller, "Optimizing spatial filters for robust eeg single-trial analysis," *Signal Processing Magazine, IEEE*, vol. 25, no. 1, pp. 41 –56, 2008.

[2] E. W. Sellers, D. J. Krusienski, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "A p300 event-related potential braincomputer interface (bci): The effects of matrix size and inter stimulus interval on performance," *Biological Psychology*, vol. 73, no. 3, pp. 242 – 252, 2006.

[3] G. Liu, G. Huang, J. Meng, and X. Zhu, "A frequency-weighted method combined with common spatial patterns for electroencephalogram classification in brainomputer interface," *Biomedical Signal Processing and Control*, vol. 5, no. 2, pp. 174 – 180, 2010.

[4] S. Lemm, B. Blankertz, G. Curio, and K.-R. Muller, "Spatio-spectral filters for improving the classification of single trial eeg," *Biomedical Engineering, IEEE Transactions on*, vol. 52, pp. 1541 –1548, sept. 2005.

[5] H. Lee and S. Choi, "Pca-based linear dynamical systems for multichannel eeg classification," in *Neural Information Processing, 2002. ICONIP '02. Proceedings of the 9th International Conference on*, vol. 2, pp. 745 – 749 vol.2, nov. 2002.

[6] J. Kittler and P. C. Young, "A new approach to feature selection based on the karhunen-loeve expansion," *Pattern Recognition*, vol. 5, no. 4, pp. 335 – 352, 1973.

[7] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve bci designs: Unified theory and new algorithms," *Biomedical Engineering, IEEE Transactions on*, vol. 58, pp. 355 –362, feb. 2011.

[8] W. Wu, Z. Chen, S. Gao, and E. Brown, "A probabilistic framework for learning robust common spatial patterns," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pp. 4658 –4661, sept. 2009.

[9] Z. Ghahramani and G. rey E. Hinton, "The em algorithm for mixtures of factor analyzers," *University of Toronto Technical Report*, feb. 1996.

[10] B. Blankertz, K.-R. Muller, D. Krusienski, G. Schalk, J. Wolpaw, A. Schlogl, G. Pfurtscheller, J. Millan, M. Schroder, and N. Birbaumer, "The bci competition iii: validating alternative approaches to actual bci problems," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 14, pp. 153 –159, june 2006.

[11] N. Vlassis and A. Likas, "A greedy em algorithm for gaussian mixture learning," *Neural Processing Letters*, vol. 15, pp. 77–87, 2002.

[12] W. Jin-Jia, J. Ke-Mei, and M. Chong-Xiao, "Gmm-based detection methods in eeg-based brain-computer interfaces," in *Pervasive Computing Signal Processing and Applications (PCSPA), 2010 First International Conference on*, pp. 779 –782, sept. 2010.

[13] H.-C. Kim, D. Kim, and S.-Y. Bang, "A pca mixture model with an efficient model selection method," in *Neural Networks, 2001. Proceedings. IJCNN '01. International Joint Conference on*, vol. 1, pp. 430 –435 vol.1, 2001.

[14] P. Sykacek, S. Roberts, M. Stokes, E. Curran, M. Gibbs, and L. Pickup, "Probabilistic methods in bci research," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 11, pp. 192 – 194, june 2003.