REMOVAL OF EEG ARTIFACTS FOR BCI APPLICATIONS USING FULLY BAYESIAN TENSOR COMPLETION

Yu Zhang[1*], *Qibin Zhao*[2], *Guoxu Zhou*[2], *Jing Jin*[1], *Xingyu Wang*[1], *Andrzej Cichocki*[2,3]

[1] Key Laboratory for Advanced Control and Optimization for Chemical Processes,
East China University of Science and Technology, China
[2] Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Japan
[3] Skolkowo Institute of Science and Technology, Moscow, Russia
[*] Email: yuzhang@ecust.edu.cn

ABSTRACT

High accuracy of electroencephalogram (EEG) classification can hardly be achieved if the signals are contaminated by severe artefacts. One helpless way to avoid such artefacts is usually to directly discard the severely disturbed EEG segments. This study considers a more elegant way that tries to recover the disturbed segments from other undisturbed segments. The possible artefacts in EEG are treated as missing values. A Bayesian tensor factorization (BTF) based method is proposed to implement EEG completion for artefact removal. By specifying a sparsity-inducing hierarchical prior, the underlying low-rank tensor is discovered from incomplete EEG tensor with automatically inferred model parameters. The EEG missing values are effectively predicted with robustness to overfitting. Effectiveness of the BTF algorithm is demonstrated on EEG data recorded from seven subjects in a brain-computer interface paradigm based on event-related potentials.

***Index Terms*—** Electroencephalogram, Brain-computer interface, Bayesian inference, Tensor completion

1. INTRODUCTION

Electroencephalogram (EEG) technique provides a good way to investigate the potential neural response in our brain to a specific mental task [1]. Since EEG recording requires relatively simple and inexpensive equipment, it has been most widely adopted for the development of brain-computer interface (BCI) system. BCI is developed to build a direct connection between a human brain and an external device through decoding (classifying) EEG, especially for recovering the communication capability of disabled people.

Task-related EEG responses, such as event-related potentials (ERPs), are usually very weak and likely to be contaminated by various artefacts (noises or outliers) that may be caused by limb movement, eye blink, environment interference or device instability, and so on [2, 3]. In the past few years, numerous methods have been proposed to extract significant features from noised EEG for mental task classification [4, 5, 6]. However, good classification performance could still hardly be achieved if the artefacts in EEG are extreme large [7]. One helpless way to avoid such artefacts is usually to directly discard the severely disturbed EEG segments. Subsequently, the curse-of-dimensionality will most probably occur, especially in the context of BCI since the number of training samples is usually limited when taking into account the system practicability [8].

Tensor, as a multiway extension of matrix, is a natural representation of multidimensional structure of EEG data [9, 10, 11, 12, 13]. Tensor factorization of incomplete data provides a powerful approach to estimate the latent factors from partially observed entries by typically exploiting a CANDECOMP/PARAFAC (CP) multilinear model with a predefined rank [14]. Recently, tensor completion has attracted increasing research interest and been successfully applied to visual data analysis [15, 16]. In EEG recording, it is actually quite impossible that all of the segments are contaminated by severe artefacts. Therefore, to remove interference, we consider that a more elegant way is to recover the disturbed segments from other undisturbed segments.

To this end, we propose to treat the possible outliers as missing values and implement completion by CP tensor factorization on multiway EEG data. The specified CP rank determines the effective dimensionality of the latent space and is considerably important for CP factorization. However, the selection of rank is usually quite challenging and computational expensive. Bayesian inference provides an effective approach to automatically estimate the model parameters [17, 18, 19].

This study was supported in part by the National Natural Science Foundation of China under Grant 61305028, Grant 91420302, Grant 61573142, Fundamental Research Funds for the Central Universities under Grant WG1414005, Grant WG1516018, Grant WH1414022, the Shanghai Chenguang Program under Grant 14CG31, the Guangdong Natural Science Foundation under Grant 2014A030308009, JSPS KAKENHI under Grant 15K15955, Jinan Youth Star of Science and Technology Plan No.201406002.

In this study, we introduce a Bayesian tensor factorization (BTF) based method to implement EEG completion for artefact removal. By specifying a sparsity-inducing hierarchical prior, the underlying low-rank tensor is discovered from incomplete EEG tensor with automatically inferred model parameters including the CP rank. The missing values are effectively predicted with robustness to overfitting for improving the classification accuracy of EEG. The effectiveness of the BTF method is validated on EEG data recorded from seven subjects in an ERP-based BCI paradigm.

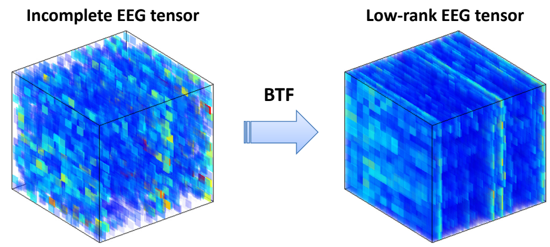

Fig. 1. An example of EEG completion via BTF.

2. BAYESIAN TENSOR FACTORIZATION FOR EEG COMPLETION

2.1. Basic notations and operations

A tensor is a multiway array of data, where the order of tensor is the number of dimensions. We denote by $\mathbf{a}$, $\mathbf{A}$, $\boldsymbol{\mathcal{A}}$ vector, matrix and tensor, respectively. An Nth-order tensor is denoted as $\boldsymbol{\mathcal{A}} = (\mathcal{A})_{i_1,\ldots,i_N} \in \mathbb{R}^{I_1\times\ldots\times I_N}$. The inner product of a set of vectors is defined as a sum of element-wise products

$$\langle \mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \ldots, \mathbf{a}^{(N)} \rangle = \sum_i \prod_n a_i^{(n)}. \tag{1}$$

The outer product of a set of vectors results in a rank-1 tensor

$$\boldsymbol{\mathcal{X}} = \mathbf{a}^{(1)} \circ \mathbf{a}^{(2)} \circ \cdots \circ \mathbf{a}^{(N)}, \tag{2}$$

where $x_{i_1\ldots i_N} = a_{i_1}^{(1)} a_{i_2}^{(2)} \cdots a_{i_N}^{(N)}$.

2.2. EEG completion based on BTF

Consider a 3th-order EEG tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I_1\times I_2\times I_3}$ (channel $\times$ point $\times$ trial). Assume the temporal points in some channels from some trials are disturbed by severe artefacts. Treating these artefact points as missing values, $\boldsymbol{\mathcal{Y}}$ becomes an incomplete tensor. We denote by Ω a set of 3-tuple indices so that $\mathcal{Y}_{i_1,i_2,i_3}$ is observed if $(i_1, i_2, i_3) \in \Omega$, and define a binary tensor $\boldsymbol{\mathcal{O}}$ of the same size as $\boldsymbol{\mathcal{Y}}$ as an indicator of observed entries. Consider the following CP model

$$\boldsymbol{\mathcal{Y}} = \sum_{r=1}^{R} \mathbf{a}_r^{(1)} \circ \mathbf{a}_r^{(2)} \circ \mathbf{a}_r^{(3)} + \boldsymbol{\varepsilon}, \tag{3}$$

where R denotes the CP rank and $\boldsymbol{\varepsilon} \sim \prod_{i_1,i_2,i_3} \mathcal{N}(0, \tau^{-1})$ is the noise term. The mode-n factor matrix $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n\times R}$ is obtained as

$$\mathbf{A}^{(n)} = [\mathbf{a}_1^{(n)}, \ldots, \mathbf{a}_{i_n}^{(n)}, \ldots, \mathbf{a}_{I_n}^{(n)}]^T, \;\; n = 1, 2, 3. \tag{4}$$

The likelihood is then written as

$$p(\boldsymbol{\mathcal{Y}}_\Omega | \{\mathbf{A}^{(n)}\}_{n=1}^3, \tau) = \prod_{i_1=1}^{I_1} \prod_{i_2=1}^{I_2} \prod_{i_3=1}^{I_3} \mathcal{N}\left(\mathcal{Y}_{i_1 i_2 i_3} | \langle \mathbf{a}_{i_1}^{(1)}, \mathbf{a}_{i_2}^{(2)}, \mathbf{a}_{i_3}^{(3)} \rangle, \tau^{-1} \right)^{\mathcal{O}_{i_1 i_2 i_3}}, \tag{5}$$

where τ denotes the noise precision, $\mathcal{Y}_{i_1 i_2 i_3}$ is generated from multiple R-dimensional latent vectors $\mathbf{a}_{i_1}^{(1)}$, $\mathbf{a}_{i_2}^{(2)}$, and $\mathbf{a}_{i_3}^{(3)}$. This likelihood allows us to model the multilinear interaction structure.

The CP rank R indicates the extent of low rank approximation and is a crucial parameter for controlling the effective dimensionality of the latent space. However, the accurate determination of R is usually quite challenging and computationally expensive. Bayesian inference provides an elegant way to automatically determine the model complexity. Accordingly, we specify a sparsity-inducing prior over hyperparameters that control the variance related to each dimensionality of the latent space, respectively. As a result, a Bayesian tensor factorization (BTF) based method is proposed to automatically infer the optimal rank R and effectively avoid overfitting for EEG tensor completion.

A prior governed by hyperparamters $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_R]$ is defined over the latent factors

$$p(\mathbf{A}^{(n)} | \boldsymbol{\lambda}) = \prod_{i_n=1}^{I_n} \mathcal{N}(\mathbf{a}_{i_n}^{(n)} | \mathbf{0}, \boldsymbol{\Lambda}^{-1}), \;\; n = 1, 2, 3, \tag{6}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{\lambda})$ is the precision matrix shared by latent factor matrices in all modes, and each λ_r controls rth component in $\mathbf{A}^{(n)}$. The prior (6) is closely related to sparse Bayesian learning and will enforce column sparsity of the factor matrices. We further define a hyperprior over $\boldsymbol{\lambda}$

$$p(\boldsymbol{\lambda}) = \prod_{r=1}^{R} \boldsymbol{\Gamma}(\lambda_r | c_0^r, d_0^r), \tag{7}$$

where $\boldsymbol{\Gamma}(x|a, b) = \frac{b^a x^{a-1} e^{-bx}}{\Gamma(a)}$ denotes a Gamma distribution. The same sparsity can be obtained at the different three modes to infer the rank of tensor during factorization since the priors are shared across different latent matrices.

For Bayesian inference, a hyperprior over the noise precision τ is also needed

$$p(\tau) = \boldsymbol{\Gamma}(\tau | a_0, b_0). \tag{8}$$

We collect all unknown latent variables and hyperparameters into $\Psi = \{\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}, \boldsymbol{\lambda}, \tau\}$. Through combining the

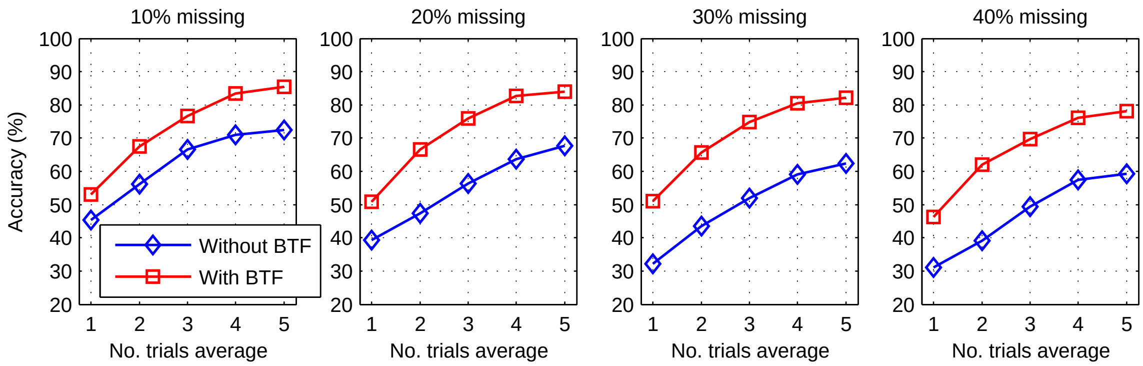

Fig. 2. Averaged ERP classification accuracy obtained by STDA only and by BTF for EEG completion followed by STDA at different missing ratios, in using one to five trials average, respectively.

likelihood (5), priors (6) and hyperpriors (7) and (8), the joint distribution can be written as

$$p(\boldsymbol{\mathcal{Y}}_\Omega, \Psi) = p\left(\boldsymbol{\mathcal{Y}}_\Omega \middle| \{\mathbf{A}^{(n)}\}_{n=1}^{3}, \tau\right) \prod_{n=1}^{3} p\left(\mathbf{A}^{(n)} \middle| \boldsymbol{\lambda}\right) p(\boldsymbol{\lambda}) p(\tau). \tag{9}$$

When given the observed EEG data, the full posterior distribution of all variables in Ψ is given as

$$p(\Psi|\boldsymbol{\mathcal{Y}}_\Omega) = \frac{p(\Psi, \boldsymbol{\mathcal{Y}}_\Omega)}{\int p(\Psi, \boldsymbol{\mathcal{Y}}_\Omega) d\Psi}. \tag{10}$$

The posterior distribution of Ψ can be estimated by a deterministic approximate inference under variational Bayesian framework [16, 20]. The predictive distribution over missing entries $\boldsymbol{\mathcal{Y}}_{\backslash\Omega}$ is then inferred as

$$p(\boldsymbol{\mathcal{Y}}_{\backslash\Omega}|\boldsymbol{\mathcal{Y}}_\Omega) = \int p(\boldsymbol{\mathcal{Y}}_{\backslash\Omega}|\Psi) p(\Psi|\boldsymbol{\mathcal{Y}}_\Omega) d\Psi. \tag{11}$$

An example of EEG completion via the BTF-based algorithm is depicted in Figure 1.

2.3. Experimental Evaluation

EEG data were recorded from seven subjects using the g.USBamp amplifier, at 256 Hz sampling rate, with band-pass filtering between 0.1 and 30 Hz. In the experiment, eight arrow commands were placed on a screen to simulate a navigation control. Specified objects were presented in a block-randomized sequence on the arrows as visual stimuli to elicit ERPs. Each subject completed 16 experimental runs. In each run, the presentation block was repeated five times, in each of which an object was presented on one arrow position once with a duration of 100 ms and ISI of 80 ms. The subjects were indicated to focus attention on the cued arrow positions and silently count the number of times the objects appeared. From each object presentation, a data segment of 700 ms was extracted and downsampled by a 12-point moving average. That is, an EEG data matrix of size 16×15 (channel $\times$ point) was derived from each trial. A total of 640 trials consisting of 80 targets and 560 non-targets were derived from each subject, where each 40 trials (5 targets and 35 non-targets) corresponded to one command selection.

The effectiveness of BTF-based method on EEG completion was investigated at different missing ratios (10%, 20%, 30% and 40%). Spatial-temporal discriminant analysis (STDA) [4] was adopted for comparison, which performed collaboratively multiway filtering to extract discriminative features from noised EEG for ERP classification, and has shown better performance over many other competing algorithms. The classification accuracy was compared between in using STDA only and in using BTF for EEG completion followed by STDA. For each subject, randomly chosen half data (i.e., 320 trials) were used for feature extraction and classifier calibration, while the remaining half for test. The classification accuracy was evaluated on repeated running of the program for 50 times, in using one to five trials average, respectively.

3. RESULTS

Figure 2 shows the ERP classification accuracy averaged over subjects, derived by STDA only and by BTF for EEG completion followed by STDA at different missing ratios, in using one to five trials average, respectively. It can be seen that accuracy was obviously enhanced by using BTF for EEG completion before feature extraction and classification. Table 1 further presents the statistical results of accuracy difference by using paired t-test between STDA only and BTF followed by STDA. The proposed BTF algorithm significantly improved the classification accuracy in using different numbers of trials average at all 10% to 40% missing ratios.

Note that the effects of BTF might become relatively limited when either very few (e.g., 1% to 2%) or too many (e.g., 90% to 95%) data are missing. Hence, we recommend that the proposed BTF-based algorithm is a good choice for EEG completion to improve classification accuracy when the miss-

Table 1. Statistical analysis for accuracy difference between STDA only and BTF followed by STDA, obtained by the paired t-test at different missing ratios.

Missing ratio	No. trials average				
	1	2	3	4	5
10%	$p < .05$	$p < .05$	$p < .005$	$p < .01$	$p < .01$
20%	$p < .01$	$p < .001$	$p < .005$	$p < .01$	$p < .05$
30%	$p < .001$	$p < .001$	$p < .001$	$p < .005$	$p < .05$
40%	$p < .05$	$p < .005$	$p < .005$	$p < .001$	$p < .005$

ing ratio is neither extremely low nor extremely high.

4. CONCLUSIONS

In this study, we proposed a Bayesian tensor factorization (BTF) based method to implement EEG completion for artefact removal. The possible outliers was treated as missing values. With a sparsity-inducing hierarchical prior, the underlying low-rank tensor is discovered from incomplete EEG tensor with automatically inferred model parameters. Experimental studies were carried out on EEG data recorded from seven subjects in an ERP-based BCI paradigm. The results demonstrated that ERP classification accuracy was significant improved by using BTF for EEG completion before classification in comparison with directly performing classification without BTF.

5. REFERENCES

[1] Y. Zhang, Q. Zhao, J. Jin, X. Wang, and A. Cichocki, "A novel BCI based on ERP components sensitive to configural processing of human faces," *J. Neural Eng.*, vol. 9, no. 2, pp. 026018, Apr. 2012.

[2] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 4, no. 2, pp. R1–R13, 2007.

[3] D.J. Krusienski, E.W. Sellers, F. Cabestaing, S. Bayoudh, D.J. McFarland, T.M. Vaughan, and J.R. Wolpaw, "A comparison of classification techniques for the P300 speller," *J. Neural Eng.*, vol. 3, no. 4, pp. 299–305, 2006.

[4] Y. Zhang, G. Zhou, Q. Zhao, J. Jin, X. Wang, and A. Cichocki, "Spatial-temporal discriminant analysis for ERP-based brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 2, pp. 233–243, Mar. 2013.

[5] F. Qi, Y. Li, and W. Wu, "RSTFC: A novel algorithm for spatio-temporal filtering and classification of single-trial EEG," *IEEE Trans. Neural Netw. Learn. Syst.*, Feb. 2015, In press, DOI: 10.1109/TNNLS.2015.2402694.

[6] F. Cong, A.-H. Phan, Astikainen P., Zhao Q., Wu Q., Hietanen J., Ristaniemi T., and Cichocki A., "Multi-domain feature extraction for small event-related ptentials throughnonnegative multi-way array decomposition from low dense array EEG," *Int. J. Neural Syst.*, vol. 23, pp. 1350006, 2013.

[7] J. Li, Z. Struzik, L. Zhang, and A. Cichocki, "Feature learning from incomplete EEG with denoising autoencoder," *Neurocomputing*, vol. 165, pp. 23–31, Oct. 2015.

[8] Y. Zhang, G. Zhou, J. Jin, Q. Zhao, X. Wang, and A. Cichocki, "Aggregation of sparse linear discriminant analysis for event-related potential classification in brain-computer interface," *Int. J. Neural Syst.*, vol. 24, no. 1, pp. 1450003, Feb. 2014.

[9] A. Cichocki, D. Mandic, A.-H. Phan, C. Caiafa, G. Zhou, Q. Zhao, and L. De Lathauwer, "Tensor decompositions for signal processing application," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, Mar. 2015.

[10] G. Zhou, Q. Zhao, Y. Zhang, and A. Cichocki, "Linked component analysis from matrices to high order tensors: applications to biomedical data," *Proc. IEEE*, 2015, In press, DOI: 10.1109/JPROC.2015.2474704.

[11] T. Kolda and B. Bader, "Tensor decomposition and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[12] F. Cong, Q. Lin, L. Kuang, X. Gong, P. Astikainen, and T. Ristaniemi, "Tensor decomposition of EEG signals: a brief review," *J. Neurosci. Meth.*, vol. 248, pp. 59–69, 2015.

[13] G. Zhou, A. Cichocki, Q. Zhao, and S. Xie, "Nonnegative matrix and tensor factorizations: an algorithmic perspective," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 54–65, May 2014.

[14] Q. Zhao, G. Zhou, L. Zhang, A. Cichocki, and S. I. Amari, "Bayesian robust tensor factorization for incomplete multiway data," *IEEE Trans. Neural Netw. Learn. Syst.*, 2015, In press, DOI: 10.1109/TNNLS.2015.2423694.

[15] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, Jan. 2013.

[16] Q. Zhao, A. Cichocki, and L. Zhang, "Bayesian CP factorization of incomplete tensors with automatic rank determination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1751–1763, Jan. 2015.

[17] W. Wu, C. Wu, S. Gao, B. Liu, Y. Li, and X. Gao, "Bayesian estimation of ERP components from multicondition and multichannel EEG," *NeuroImage*, vol. 88, pp. 319–339, Mar. 2014.

[18] W. Wu, Z. Chen, X. Gao, Y. Li, E. Brown, and S. Gao, "Probabilistic common spatial patterns for multichannel EEG analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 639–653, Mar. 2015.

[19] Y. Zhang, G. Zhou, J. Jin, Q. Zhao, X. Wang, and A. Cichocki, "Sparse Bayesian classification of EEG for brain-computer interface," *IEEE Trans. Neural Netw. Learn. Syst.*, 2015, In press, DOI: 10.1109/TNNLS.2015.2476656.

[20] J. Minn and C. Bishop, "Variational message passing," *J. Mach. Learn. Res.*, vol. 6, pp. 661–694, 2005.