35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Boosting Specificity of MEG Artifact Removal by Weighted Support Vector Machine*

Fang Duan, Montri Phothisonothai, Mitsuru Kikuchi, Yuko Yoshimura, Yoshio Minabe, Kastumi Watanabe, and Kazuyuki Aihara

Abstract—**An automatic artifact removal method of magnetoencephalogram (MEG) was presented in this paper. The method proposed is based on independent components analysis (ICA) and support vector machine (SVM). However, different from the previous studies, in this paper we consider two factors which would influence the performance. First, the imbalance factor of independent components (ICs) of MEG is handled by weighted SVM. Second, instead of simply setting a fixed weight to each class, a re-weighting scheme is used for the preservation of useful MEG ICs. Experimental results on manually marked MEG dataset showed that the method proposed could correctly distinguish the artifacts from the MEG ICs. Meanwhile, 99.72%±0.67 of MEG ICs were preserved. The classification accuracy was 97.91%±1.39. In addition, it was found that this method was not sensitive to individual differences. The cross validation (leave-one-subject-out) results showed an averaged accuracy of 97.41%±2.14.**

I. Introduction

Magnetoencephalography (MEG) provides non-invasive and real-time monitor of dynamic behavior and neural activity of the human brain on a millisecond time-scale. An advantage of MEG is that it does not face volume conduction effect as strong as electroencephalography (EEG) [1]. In addition, compared with other brain imaging methods, such as functional magnetic resonance imaging (fMRI) which is limited in temporal resolution to second timescales, MEG has better resolution in both temporal and spatial domains. Similar to EEG recording, MEG can be contaminated by physiological artifacts from various sources, such as eye movements, muscular contractions, cardiac signals, sudden high-amplitude changes, and environmental noise [2]. In recent years, independent component analysis (ICA) [3] is achieving very successful results in separating artifacts from brain signals [4, 5].

With ICA, the raw brain signals could be decomposed into different independent components (ICs) effectively. However, these ICs include artifactual ICs and meaningful MEG ICs. In general, ICA-based artifact removal methods suggest identifying and removing the artifactual ICs by manual inspection. But this process is very exhausting especially for multichannel signals like MEG which usually contain more than 100 channels. In order to remove artifacts automatically, support vector machine (SVM) was used to discriminate artifactual ICs from MEG ICs [6, 7]. By this way, the artifact could be removed effectively. However, there are still some problems which have not been taken into consideration. First, the balance of datasets needs to be taken into account. In fact, for MEG data, the ICs representing artifacts and brain activities are commonly imbalance. Further, when faced with imbalanced datasets, the performance of SVM may deteriorate significantly [8]. Second, there are still researches that use raw MEG even by exhausting visual inspection[9] to cut out relatively high quality MEG epochs, because there is an argument that ICA may eliminate not only artifacts but also useful MEG components. Therefore, we propose a method by which the useful MEG ICs can be well retained, while denoising. Considering artifacts and MEG ICs as positive class and negative class respectively, it is inevitably misclassifying the negative class into positive class when the specificity of classifier is not 100%, and which will result in an insufficient preservation of MEG ICs.

We propose a method to solve the two problems mentioned above by using weighted SVM to handle the imbalance ICs dataset and boost specificity of classifier by re-weighting the sample's weight of the negative class. The advantage of the proposed method is shown in a manually marked MEG ICs dataset.

II. Methods

A. Independent Component Analysis (ICA) Method

Independent component analysis (ICA) is now an important tool to separate empirical datasets. For the multichannel time series $X(t)$, ICA method is trying to estimate the mixing matrix W and the source time series $s(t)$ which is statistically independent each other in each time:

$$X(t) = \mathbf{W}s(t). \tag{1}$$

ICA mainly considers to separate time independent sources from linearly mixed data. In brain signal processing, we assume that meaningful brain activity is independent of artifact. Therefore, ICA can separate raw data to sources of brain activity and sources of artifact respectively. Therefore, ICA can be used to separate the components produced by muscle movement and cardiac artifact from the data. Generally, ICA performances are good in the field of artifact

* This work was supported by the Aihara Project, the FIRST program from JSPS, initiated by CSTP, the Japan Society for the Promotion of Science (JSPS).

F. Duan and K. Aihara are with the department of Electrical Engineering and Information Systems, the University of Tokyo, Tokyo 153-8904 Japan (phone:+81-3-5452-6697; e-mail: duan@sat.t.u-tokyo.ac.jp).

M. Phothisonothai and K. Watanabe are with the Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo 153-8904 Japan (e-mail: montri@fennel.rcast.u-tokyo.ac.jp).

Y. Yoshimura, M. Kikuchi, and Y. Minabe are with the Research Center for Child Mental Development, Graduate School of Medical Science, Kanazawa University, Kanazawa 920-8641 Japan (e-mail: minabe@med.m.kanazawa-u.ac.jp).

and noise detection because the non-Gaussian criterion in ICA estimation makes ICA more sensitive to non-Gaussian components. Meanwhile, interesting components of oscillations in brain are not far from Gaussian in statistics [10]. Artifacts usually show strong non-Gaussianity. In this work, we used FastICA [11] to calculate the sources and mixing matrix. For more details, see ref [6].

B. *Weighted Support Vector Machine (WSVM)*

WSVM was proposed for unbalance datasets [8]. There are three options for unbalance data: cost sensitive learning, oversampling the minority class, and undersampling the majority class. WSVM applies cost sensitive learning on SVM. In this work, WSVM was applied by modified C-Support Vector Classification (C-SVC) [12]. Given set of training samples with labels $\{(x_1, y_1),...,(x_l, y_l)\}$, $y_i \in \{1,-1\}$, WSVM solves the following optimization problem:

$$\begin{aligned} &\arg\min_{\omega,b,\xi} \quad \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{l} w_i\xi_i \\ &s.t.\ \ y_i(\omega^T\phi(x_i)+b) \geq 1-\xi_i, \\ &\xi_i \geq 0, i=1,...,l, \textstyle\sum_{i=1}^{l} w_i = l, \end{aligned} \tag{2}$$

where $\phi(x_i)$ maps x_i into a higher-dimensional space from the input space, $C > 0$ is the regularization constant and w_i is the *i*th weight parameter corresponding to the *i*th sample. Class 1 is the artifact class; class -1 is the MEG class. Usually, we solve the following dual form of (2):

$$\begin{aligned} &\arg\min_{\omega,b,\xi} \quad \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j k(x_i, x_j) - \sum_{i=1}^{l}\alpha_i \\ &s.t.\ \ \sum_{i=1}^{l} y_i\alpha_i = 0, 0 \leq \alpha_i \leq w_i C, \textstyle\sum_{i=1}^{l} w_i = l, \end{aligned} \tag{3}$$

where $k(x_i, x_j) = \phi(x_i)^T\phi(x_j)$ is the kernel function. In this research, $k(x_i, x_j) = \exp(-\| x_i - x_j \|^2 / 2\sigma^2)$, the RBF kernel, was used. Parameters C and σ were decided by 5 fold cross-validation at the beginning. The initial value of w_i are given as follows:

$$w_i = \begin{cases} \dfrac{l}{\sum_{j=1}^{l}(1+y_j)}, & y_j = 1, \\[2ex] \dfrac{l}{\sum_{j=1}^{l}(1-y_j)}, & y_j = -1. \end{cases} \tag{4}$$

In this work, WSVM classifier was trained by LIBSVM Tool: Weights for data instances [13].

C. *Re-weighting SVM*

In order to provide the classifier with better specificity, the re-weighting technique was used to update the weights of training samples as shown in Table I. The accuracy, specificity and sensitivity are defined as follow:

TABLE I. SUMMARY OF C. RE-WEIGHTING SVM

1. **Input:** a set of training samples with labels $\{(x_1, y_1),...$,
   $y_i \in \{1,-1\}$; the step of w_i , w_{step} ; the maximal number of cycles *T*.
2. **Initialize:** $t = 0$; the weights of training samples w_i^t as (4); learning C and σ on the weighted training set by cross-validation before iteration; calculate initial accuracy ε .
3. **While** $\varepsilon < 1$ and $t < T$
   a) $t = t + 1$.
   b) Train a RBFSVM classifier, $f_t(x)$, on the weighted training set.
   c) Calculate the accuracy, specificity and sensitivity of $f_t(x)$, ε , $\varepsilon_{\text{spec}}$ and ε_{sen} on training set.
   d) If $\varepsilon_{\text{spec}} < 1$
   $w_i^t = w_i^{t-1}\exp(-y_i f_t(x_i) w_{\text{step}}) / K_t, y_i f_t(x_i) = -1, y_i = -1$,
   K_t is a constant to normalize $\sum w_j^t$ equal to l .
   Else
   $w_i^t = w_i^{t-1}\exp(-y_i f_t(x_i) w_{\text{step}}) / K_t, y_i f_t(x_i) = -1, y_i = 1$,
   K_t is a constant to normalize $\sum w_j^t$ equal to l .
4. **Output:** $f(x) = f_t(x)$

$$\varepsilon = \frac{\text{FN} + \text{FP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}}, \tag{5}$$

$$\varepsilon_{\text{spec}} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \tag{6}$$

$$\varepsilon_{sen} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{7}$$

where TN, TP, FN, and FP are the numbers of true negative, true positive, false negative, and false positive, respectively. We focused on boosting specificity of classifier under acceptable sensitivity. Therefore, we re-weighted the weights of the MEG class (real negative) in order to get better specificity, and updated the weights of the artifact class (real positive) only when specificity of classifier was equal to 1.

D. *Feature Extraction*

For training classifier, it is needed to input some features to WSVM. In this research, we extracted five features, the same as ref [6, 7], from each component which obtained by ICA.

1) *Kurtosis*

Kurtosis, also called the fourth-order cumulant is a classical measure of non-Gaussianity. For a Gaussian x kurtosis is equal to 1. It has been successfully used to detect artifacts in ref [14]. For most of non-Gaussian random variables, kurtosis is not equal to 1. In this work, a log normalized kurtosis was defined for zero mean random variable x as follows:

$$K = \log(E[x^4] / 3\sigma^4), \tag{8}$$

where σ is the standard deviation of x .

2) *Probability Density*

Usually cardiac artifact has a peak at minimum or maximum point of its histogram [6]. So we estimated the probability density function and cumulated the mean ratio between both ends and highest peak in the IC's histogram as (9).

$$PD = (P(x = x_{\min}) + P(x = x_{\max})) / \max P(x) \quad (9)$$

where $P(\bullet)$ is the probability density function obtained by histogram.

3) Central Moment of Frequency

Central moment of frequency (CMoF) was employed as the third feature. This feature could show the dominated frequency of each IC. The response of muscle movement usually shows different CMoF with the MEG component. CMof is defined as

$$CMoF = \frac{1}{M}\sum_{f=1}^{40} fP_n(f), \quad (10)$$

where $P_n(f)$ is the power spectral density corresponding to f . M is the number of frequency bins.

4) Spectral Entropy

The spectral entropy (SE) can quantify the flatness of the frequency spectrum [5] as follows:

$$SpecEn = -\frac{1}{\log M}\sum_{f=1}^{40} P_n(f)\log[P_n(f)]. \quad (11)$$

5) Fractal Dimension

Fractal dimension has been successfully used to extract feature from brain signals [15]. This value can describe the smoothness of the brain signal waveform. In this work, the variance fractal dimension (VFD) [6] was used to estimate the FD of ICs. The normalized FD is given as follow:

$$FD = 1 - \frac{1}{4}\lim_{\Delta t \to 0}\frac{\log_2 D(\Delta x)}{\log_2 \Delta t}, \quad (12)$$

where Δx is equal to $x(t) - x(t - \Delta t)$. The limit in (12) can be estimate by the slope of the least-square fitted line of the plot $\log_2 n_k$ and $\log_2 D(x(t) - x(t - n_k))$. In this work, we set $n_k = 2^k$, $k = 1, \ldots, [\log_2 N] - 2$, and N is the length of IC.

III. Experiments and Results

A. MEG Data

MEG datasets included 10 healthy children's MEG data (5 boys, 5 girls). The average age of children was 65.4 months. In order to keep the subjects stay calm, the subjects were allowed to watch video during the experiment. MEG data were recorded by a whole-head coaxial gradiometer system (PQ 1151R; Yokogawa/KIT) for children with 151 channels superconducting quantum interface device in a magnetically shielded room under resting state. The measurement of data was approved by the Ethics Committee of Kanazawa University Hospital.

The data were separated to 1326 ICs by FastICA. In this work, we did numerical experiments on 956 ICs in the dataset which was marked as artifactual ICs and MEG ICs by manual inspection. The experimental dataset included 646 MEG ICs and 310 artifacts. The artifacts included cardiac, ocular, muscular and sudden high-amplitude change artifacts. We omitted other 370 ICs because those ICs were automatically not manually marked as ocular artifacts in the dataset by threshold method in previous research which may not real artifacts but meaningful ICs [6].

B. Experiment I

In the first experiment, we used all 956 ICs to test the proposed algorithm. The maximal number of iterations was set to 50. The results are shown in Figure 1. Accuracies were obtained using 10×10 cross-validation. The SVM parameter C and σ were all decided for each trial separately also by 10×10 cross-validation. At the beginning of iteration, the classification accuracy, specificity and sensitivity of classifier were 97.73%±1.55, 97.71%±1.94 and 97.77%±2.47, respectively. As shown in Figure 1, the specificity of classifier was boosted up by re-weighting samples in the MEG class. At the 39th iteration, the specificity of classifier reached its maximal value 99.72%±0.67. After the 39th iterations, the value of specificity varied between 99.69% and 99.72%, indicating that the specificity of the classifier became saturated. With the increase of iterations, sensitivity continued to decline. At the 39th iteration, sensitivity was 94.02%±4.14. After 50 iterations, sensitivity was 91.67%±4.99. The classification accuracy reached the maximal value 98.26%±1.38 at the 22nd iteration. At the 39th iteration, classification accuracy was 97.91%±1.39 which was still higher than the initial classification accuracy. After 50 iterations, the accuracy with 97.12%±1.58 was still acceptable.

C. Experiment II

In this experiment, we tested the generalization property of classifier trained by proposed method. By dividing the subjects into a training group and a testing group, the trained classifier obtained from the training group was used to test the data in the testing group. Figure 2 showed the average and standard deviation of classification accuracy, specificity and

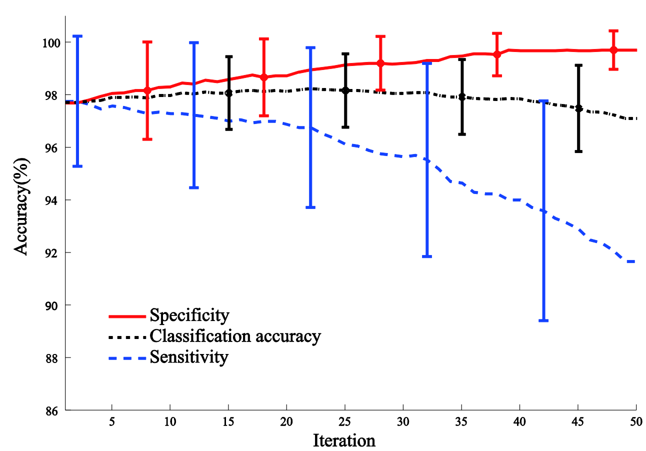

Figure 1. Specificity, sensitivity and classification accuracies with their standard deviation (%) across iterations on the dataset. All accuracies are obtained by 10×10 fold cross-validation.

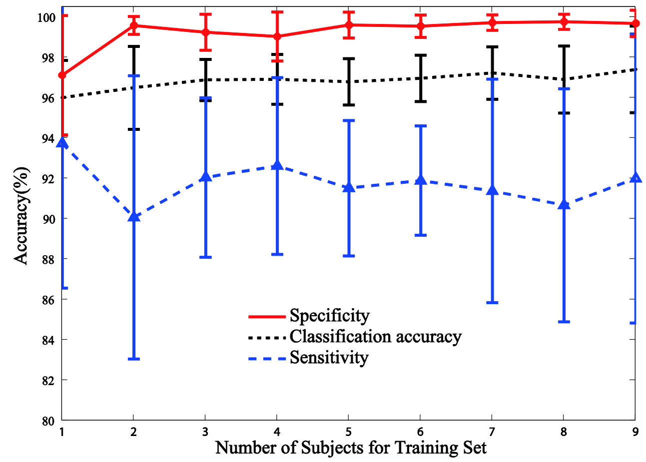

Figure 2. Specificity, sensitivity and classification accuracies with their standard deviation (%) across number of subjects.

sensitivity. Here, the result was obtained by 10 validations. In 10 validations, all subjects had equal chance to be in the training group and the testing group. It was found that the classifier trained from certain subjects can be used to classify other subjects' IC. Specificity and classification accuracy gradually increase with the number of subjects in the training set. The specificity, classification accuracy and sensitivity with leaving one subject out cross-validation were 99.69%±0.65, 97.41%±2.14 and 92.01%±7.16, respectively.

IV. DISCUSSION

The experiment I showed that after several iterations the specificity of classifier reached the maximal value. The specificity was nearly 100%. It means that the proposed method can preserve all MEG ICs after the artifacts removal. Meanwhile sensitivity was still acceptable. After the specificity is saturated, continuing to re-weight the samples' weights cannot increase the specificity. Also the sensitivity will keep decreasing, and the re-weighting just over-fits the negative class. The sensitivity becomes unacceptable if iterations keep going. Therefore, the iterative process can stop when the specificity is saturation. The classifier will have the best specificity and acceptable sensitivity. Also the classification accuracy will be higher than the one without iterations.

Besides, from Table I, we can find that the proposed method is similar to the re-weighting process of the adaboost algorithm [16]. Adaboost was proposed to enhance the classification accuracy of weak learner. In ref [17], authors indicated that strong learner like SVM also can be boosted by adaboost algorithm. In our method, the iterative process trained several classifiers which could be intergraded by adaboost. Then a better classification performance might be able to obtain.

From the results of experiment II, it was found that this method was not sensitive to individual differences. Meanwhile, we can say the features which was proposed in ref [6] could represent the character of noisy ICs across different subjects. As we know, neurophysiological signals usually have peculiarity for each subject. Some researches employed features which can represent peculiarity of each subject like the relevance between IC and certain manually selected ocular IC [18]. But features in ref [6] extract the common character of all subjects. Furthermore, classifier which was well trained by sufficient dataset can be applied on all MEG signals. The algorithm proposed can train a classifier to automatically identify the MEG epoch. These features will disuse exhausting visual inspection for massive multichannel MEG data.

REFERENCES

[1] S. P. van den Broek, F. Reinders, M. Donderwinkel et al., "Volume conduction effects in EEG and MEG," Electroencephalography and Clinical Neurophysiology, vol. 106, no. 6, pp. 522-534, 1998.

[2] R. Vigário, J. Sarela, V. Jousmiki et al., "Independent component approach to the analysis of EEG and MEG recordings," Biomedical Engineering, IEEE Transactions on, vol. 47, no. 5, pp. 589-593, 2000.

[3] L. Te-Won, Independent component analysis: theory and applications: Boston: Kluwer Academic Publishers, 1998.

[4] S. Y. Shao, K. Q. Shen, C. J. Ong et al., "Automatic EEG artifact removal: a weighted support vector machine approach with error correction," Biomedical Engineering, IEEE Transactions on, vol. 56, no. 2, pp. 336-344, 2009.

[5] D. Mantini, R. Franciotti, G. L. Romani et al., "Improving MEG source localizations: an automated method for complete artifact removal based on independent component analysis," NeuroImage, vol. 40, no. 1, pp. 160-73, 2008.

[6] M. Phothisonothai, H. Tsubomi, A. Kondo et al., "Linear and nonlinear features for automatic artifacts removal from MEG data based on ICA." in Proc. Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific,2012, pp. 1-9.

[7] M. Phothisonothai, F. Duan, H. Tsubomi et al., "Artifactual Component Classification from MEG Data using Support Vector Machine," in Proc. 5th Biomed. Eng. Int. Conf. (BMEiCON), 2012, pp. 1-5.

[8] R. Akbani, S. Kwek, and N. Japkowicz, "Applying support vector machines to imbalanced datasets," Machine Learning: ECML 2004, pp. 39-50, 2004.

[9] V. Tsiaras, P. G. Simos, R. Rezaie et al., "Extracting biomarkers of autism from MEG resting-state functional connectivity networks," Computers in Biology and Medicine, 2011.

[10] A. Hyvärinen, P. Ramkumar, L. Parkkonen et al., "Independent component analysis of short-time Fourier transforms for spontaneous EEG/MEG analysis," NeuroImage, vol. 49, no. 1, pp. 257-271, 2010.

[11] A. Hyvärinen, and E. Oja, "A fast fixed-point algorithm for independent component analysis," Neural computation, vol. 9, no. 7, pp. 1483-1492, 1997.

[12] C. Cortes, and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273-297, 1995.

[13] C. C. Chang, and C. J. Lin, "LIBSVM: a library for support vector machines," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 2, no. 3, pp. 27, 2011.

[14] J. Escudero, R. Hornero, D. Abásolo et al., "Artifact removal in magnetoencephalogram background activity with independent component analysis," Biomedical Engineering, IEEE Transactions on, vol. 54, no. 11, pp. 1965-1973, 2007.

[15] M. Phothisonothai, and M. Nakagawa, "EEG-based classification of motor imagery tasks using fractal dimension and neural network for brain-computer interface," IEICE TRANSACTIONS on Information and Systems, vol. 91, no. 1, pp. 44-53, 2008.

[16] R. E. Schapire, and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," Machine Learning, vol. 37, no. 3, pp. 297-336, Dec, 1999.

[17] X. Li, L. Wang, and E. Sung, "AdaBoost with SVM-based component classifiers," Engineering Applications of Artificial Intelligence, vol. 21, no. 5, pp. 785-795, 2008.

[18] L. Shoker, S. Sanei, and J. Chambers, "Artifact removal from electroencephalograms using a hybrid BSS-SVM algorithm," Signal Processing Letters, IEEE, vol. 12, no. 10, pp. 721-724, 2005.