https://doi.org/10.5392/IJoC.2018.14.4.024

Improving the Subject Independent Classification of Implicit Intention By Generating Additional Training Data with PCA and ICA

Sang-Hoon Oh
Division of Information Communication Convergence Eng.
Mokwon University, Daejeon, 302-729, Korea

ABSTRACT

EEG-based brain-computer interfaces has focused on explicitly expressed intentions to assist physically impaired patients. For EEG-based-computer interfaces to function effectively, it should be able to understand users' implicit information. Since it is hard to gather EEG signals of human brains, we do not have enough training data which are essential for proper classification performance of implicit intention. In this paper, we improve the subject independent classification of implicit intention through the generation of additional training data. In the first stage, we perform the PCA (principal component analysis) of training data in a bid to remove redundant components in the components within the input data. After the dimension reduction by PCA, we train ICA (independent component analysis) network whose outputs are statistically independent. We can get additional training data by adding Gaussian noises to ICA outputs and projecting them to input data domain. Through simulations with EEG data provided by CNSL, KAIST, we improve the classification performance from 65.05% to 66.69% with Gamma components. The proposed sample generation method can be applied to any machine learning problem with fewer samples.

***Key words**: Implicit Intention, Subject Independent BCI, Support Vector Machine, Principal Component Analysis, Independent Component Analysis.*

1. INTRODUCTION

When developing an intelligent system, we usually implement speech recognition, image understanding and natural language processing within the system [1], [2]. Also, emotion is a fundamental characteristic of human and developing a human-like learning system requires emotion recognition by analyzing people's speech, gesture, and facial expression [3]. Additionally, there have been developments of BCI (brain-computer interface) because the brain is the fundamental source of information generated by humans.

BCIs have focused on explicitly-expressed intentions for the assistance of physically impaired patients [4]-[6]. When impaired patients want to do something, recognizing explicitly-expressed intentions and performing the recognized actions are essential for the impaired patients' well-being. In ordinary life, humans explicitly express their intentions in various ways. However, sometimes in a critical moment, humans do not express their intention but implicitly do. So, intelligent user-interface should have a capability of understanding implicit intentions.

Research on implicit intention has focused on hidden intention or lie detection, i.e., whether or not the users' explicitly-expressed intention is the same as the actual intention [7], [8]. Most recently, Dong et al. focused on another type of implicit intention, i.e., unexpressed intention, whether or not a user agrees with the others during conservation or sentence reading [9], [10].

Many BCIs are based on electroencephalography (EEG) signals since EEG has non-invasive nature and high temporal resolution. Accordingly, Dong et al. reported the EEG-based subject dependent classification of implicit intention during self-relevant sentence reading [9]. They used five time-frequency components extracted from EEG signals, and the best accuracy was 75.5% with Gamma components. In subject dependent classifications, we can only classify the implicit intention of users whose EEG signals are already recorded and analyzed. On the contrary, there is an effort of subject independent classification for classifying implicit intention of unanalyzed users [11]. However, the best accuracy of subject independent classification with Gamma components was about 10% lower than that of the subject dependent classification by Dong et al.

Because of the life-ethics problem, it is very strict to have the permission of gathering EEG signals and, consequently, there are not enough number of EEG signals for classification of implicit intention [9], [11]. When we train machine learning models for some applications, there must be enough number of training data because the machine learning models find their solutions from training data. Fewer training data lead to poor performance of machine learning, and more training data result in better performance.

* *Corresponding author, Email: ohsanghoon16@gmail.com*
Manuscript received Jul. 25, 2018; revised Oct. 23, 2018; accepted Oct. 24, 2018

In this paper, for improving the subject independent classification of implicit intention, we propose to generate additional training data through PCA (principal component analysis) and ICA (independent component analysis) of EEG data. Firstly, we perform PCA to reduce the dimension of EEG data by removing redundant components. Secondly, we train ICA network with the dimension-reduced data to obtain statistically independent components. Thirdly, we add i.i.d.(independent, identically distributed) Gaussian noises to the outputs of ICA network and back-project them to EEG data domain. The back-projected data is the additionally generated training data through PCA and ICA. Finally, we train the SVM(support vector machine) classifier with whole training data and additionally generated EEG data for subject independent classification of implicit intention.

In section 2, we briefly introduce PCA and ICA which are essential for data generation. Section 3 describes details of the data generation process through PCA, ICA, the addition of Gaussian noises, and back-projection from ICA outputs to EEG data domain. Also, we describe the subject independent classification by SVM in section 3. Finally, section 4 concludes this paper.

2. PRINCIPAL COMPONENT ANALYSIS AND INDEPENDENT COMPONENT ANALYSIS

2.1. PCA (PRINCIPAL COMPONENT ANALYSIS)

Data is often high-dimensional, and they will typically lie close to a much lower dimensional manifold. By reducing the data dimension, we can reduce time complexity with less computation and space complexity with fewer parameters. Also, a simple model is more robust on small datasets and more interpretable. Thus, dimension reduction is essential for processing high-dimensional data.

PCA is a linear projection method onto lower dimensional space while preserving as much information as possible [12]. Let's assume that we have d-dimensional zero-mean random vector $\mathbf{x}=[x_1,x_2,\cdots,x_d]^T$ and a projection unit vector $\mathbf{w}=[w_1,w_2,\cdots,w_d]^T$ where $\|\mathbf{w}\|=\left(\mathbf{w}^T\mathbf{w}\right)^{1/2}=1$. Then, the projected result is given by

$$z = \mathbf{w}^T\mathbf{x} \tag{1}$$

and we want to select a projection unit vector which corresponds to the direction of maximum variance in $\mathbf{x}$. Continuously, we find another direction along which variance is maximized but restrict the search to directions orthonormal to all previously selected directions.

To achieve the above description, we construct a correlation matrix of $\mathbf{x}$ as

$$\mathbf{S} = E\{\mathbf{x}\mathbf{x}^T\} \tag{2}$$

and solve the eigenvalue problem given by

$$\mathbf{S}\mathbf{w}_j = \lambda_j\mathbf{w}_j, \quad j=1,2,\ldots,d\ , \tag{3}$$

where $\mathbf{w}_j$ is the eigenvector and λ_j is the corresponding eigenvalue. Here, $E\{.\}$ is the expectation operator. We have d eigenvectors with corresponding eigenvalues in Eq. (3). By arranging the eigenvalues in decreasing order $\lambda_1>\lambda_2>\lambda_3.....>\lambda_d$, the associated vectors are given by

$$\mathbf{W}=[\mathbf{w}_1,\mathbf{w}_2,\mathbf{w}_3,....\mathbf{w}_d]\ . \tag{4}$$

Then, the projected data representation is

$$\mathbf{z}=[z_1,z_2,z_3,....z_d]^T=\mathbf{W}^T\mathbf{x}\ . \tag{5}$$

Moreover, we can have a dimensionality-reduced vector by selecting $\mathbf{z}'=[z_1,z_2,z_3,....z_k]^T$ where $k<d$.

PCA removes the second order dependencies among data components $z_1,z_2,z_3,....z_d$ because of the orthogonality of projection vectors. However, there are high order dependencies among them.

2.2. ICA (INDEPENDENT COMPONENT ANALYSIS)

ICA is also a linear projection method to find statistically independent directions in data. Let's assume that sources, $s_i,\ i=1,\cdots,n$, are unknown and mutually independent with zero means. We have sensor signals which are a spontaneous mixture of sources given by

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{6}$$

where $\mathbf{s}=[s_1,s_2,\cdots,s_n]^T$ is a source signal vector, $\mathbf{x}=[x_1,x_2,\cdots,x_n]^T$ is a sensor signal vector, and $\mathbf{A}$ is a mixing matrix. ICA is an algorithm to recover the sources from sensor signals except for scales and a permutation of indices, that is, to find an unmixing matrix $\mathbf{V}$ such that

$$\mathbf{u} = \mathbf{V}\mathbf{x} \tag{7}$$

corresponds to the source signal vector. There are many approaches to find the unmixing matrix[13]. Here, we introduce the InfoMax algorithm which maximizes the entropy of $\mathbf{y}=g(\mathbf{u})$ where $g(.)$ is a probability distribution function of sources [14].

The joint entropy of $\mathbf{y}$ is given by

$$H(\mathbf{y})=-E\{\log p(\mathbf{y})\} \tag{8}$$

where $p(.)$ denotes the probability density function(p.d.f.). Since

$$p(\mathbf{y}) = \frac{p(\mathbf{x})}{|J(\mathbf{x})|}, \tag{9}$$

we have

$$H(\mathbf{y}) = E\{\log|\mathrm{J}(\mathbf{x})|\} - E\{\log p(\mathbf{x})\}. \tag{10}$$

Here, $J(\mathbf{x})$ is the Jacobian matrix. Then, for maximizing $H(\mathbf{y})$, the unmixing matrix is updated by

$$\Delta\mathbf{V} \propto \frac{\partial H(\mathbf{y})}{\partial \mathbf{V}} = \frac{\partial}{\partial \mathbf{V}}\log|\det(\mathbf{V})| + \sum_{i=1}^{N}\frac{\partial}{\partial \mathbf{V}}\log\left|\frac{\partial y_i}{\partial u_i}\right| \\ = (\mathbf{V}^T)^{-1} - \varphi(\mathbf{u})\mathbf{x}^T \tag{11}$$

Where

$$\varphi(u_i) = -\frac{\frac{\partial p(u_i)}{\partial u_i}}{p(u_i)} \tag{12}$$

is the score function. Using the natural gradient for efficient learning [15], we have an updating equation of the unmixing matrix

$$\Delta\mathbf{V} \propto \frac{\partial H(\mathbf{y})}{\partial \mathbf{V}}\mathbf{V}^T\mathbf{V} = [\mathbf{I} - \varphi(\mathbf{u})\mathbf{u}^T]\mathbf{V}. \tag{13}$$

Compared to PCA, ICA networks remove the higher order dependencies among data components.

3. GENERATION OF TRAINING DATA

We use the EEG data [9] provided by CNSL(computational neuro-systems laboratory), KAIST, for subject independent classification of implicit intention. A total of nine subjects were participated voluntarily in the study, and none of them had a history of mental disorder, significant physical illness, head injuries, neurological disorder, or alcohol or drug dependencies. Throughout testing experiments for reliable and honest responses, one trial data from Subject 5 and eleven trial data from Subject 6 were removed from the dataset. So, we have 653 trial data [9].

Dong et al. [9] extracted the time-frequency representations of EEG signals at thirty electrodes by Morlet wavelet transform. After studying of fMRI, the five time-frequency components were selected as (1) the gamma component (35Hz - 45Hz) between 350 ms and 550 ms, (2) the beta2 component (20Hz - 26Hz) between 300 ms and 450 ms, (3) the beta1 component (14Hz - 17Hz) between 800 ms and 1,000 ms, (4) the alpha component (9Hz - 12Hz) between 300 ms and 700 ms, and (5) the theta component (5Hz - 7Hz) between 400 ms and 1,000 ms after the onset of the contents [9]. Thus, we have 150 dimensional EEG data consisted of the five components at thirty electrodes. Fig. 1 shows the average of the five time-frequency components on the scalp topographic map.

Since there are not separate sets of training and test data in the datasets, we use 1-out of-9 validation method for training and evaluating a classifier. That is, EEG data from randomly chosen eight subjects among nine ones are used for training and those from another subject are used for evaluating the test performance of a classifier. Finally, we average the nine cases for performance evaluation.

Among many machine learning models, we select the support vector machine (SVM) with radial basis function (RBF) kernels because of its' better performance with less training samples [16]. However, 653 trial data are not enough to train an SVM classifier. So, we propose to generate additional training data through PCA and ICA.

Firstly, we perform PCA to find primary principal components given by Eq. (5) and remove redundant components among EEG data of 150 dimensional. Since we have nine cases of training sets, we perform PCA and investigate the eigenvalues for each training set. Figure 2 is the plot of averaged eigenvalues of nine cases. The eigenvalues are very large in the primary principal component and decrease very sharply. We select the fifty principal components for data dimension reduction. Thus, we remove one hundred components and the dimensionality-reduced vector is

$$\mathbf{z}' = [z_1, z_2, z_3, \ldots z_{50}]^T. \tag{14}$$

Also, there are no second-order dependencies among the fifty principal components.

Secondly, to find out independent components, we train an ICA network with the dimensionality-reduced data by PCA. Here, we use the extended ICA algorithm to train the 50 x 50 sized ICA network [17]. After training of ICA networks

$$\mathbf{u} = \mathbf{V}\mathbf{z}', \tag{15}$$

we have fifty mutually independent components. That is, there are no higher-order dependencies among outputs of the ICA network.

Thirdly, we add i.i.d. Gaussian noises to the output of ICA network. Since ICA outputs are mutually independent, adding i.i.d. noises may not disturb the data distribution of ICA outputs.

Fourthly, we reversely project the ICA output with Gaussian noises on the input layer of ICA network, which correspond to $\mathbf{z}'$ given by Eq. (14) and (15).

Fifthly, we make 150-dimensional vector $\mathbf{z}$ by adding 100 zeros to $\mathbf{z}'$. Finally, we can get the generated sample from projection $\mathbf{x} = \mathbf{W}^{T^{-1}}\mathbf{z} = \mathbf{W}\mathbf{z}$ since the inverse matrix of orthogonal maxtrix is equal to its transpose.

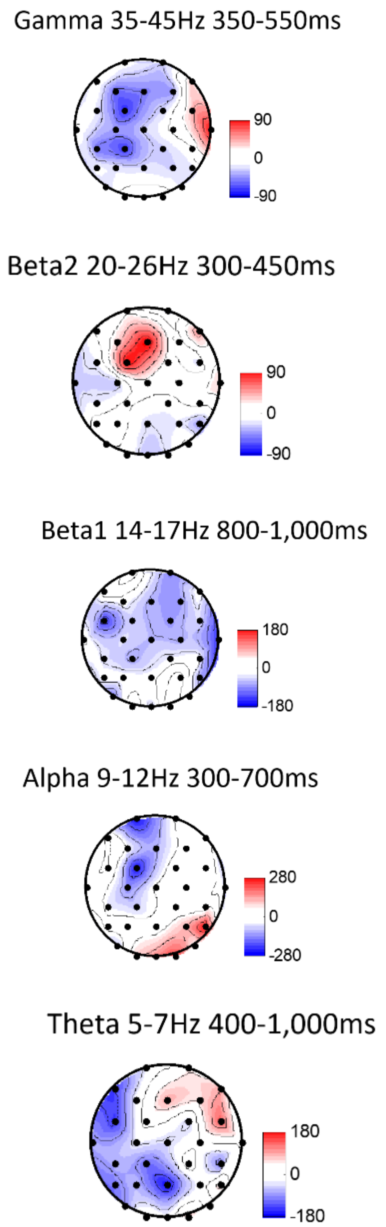

Fig. 1.The five components on the scalp topographic map.

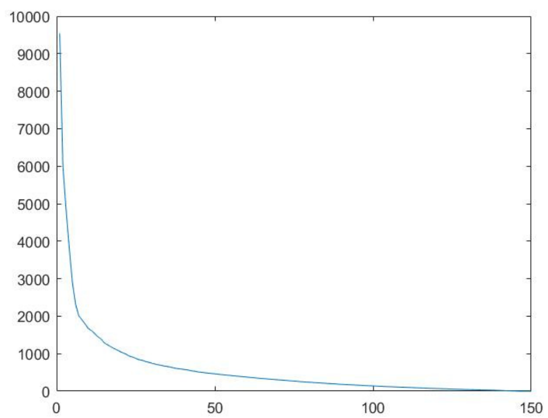

Fig. 2. Averaged eigenvalues of nine EEG data sets.

After finishing the sample generation, we train the SVM classifier with RBF kernels using LIBSVM toolbox [18]. We generate three times the number of training samples. The standard deviation of Gaussian noise is six-times of that for ICA outputs. We summarize the sample generation steps as follows.

[Sample Generation and Classifier Training Procedure]

Step 1: Store training samples.

Step 2: Make zero mean vectors of training samples and Calculate the correlation matrix of them.

Step 3: Conduct PCA on the correlation matrix.

Step 4: Sort the eigenvectors with decreasing order of eigenvalues.

Step 5: Calculate the 150 principal components of the whole training samples and remove 100 components with the least significant eigenvalues.

Step 6: Train 50 x 50 ICA network with the 50-dimensional principal components of the whole training samples.

Step 7: Present an input data to the ICA network and calculate the output of ICA network.

Step 8: Add Gaussian noises to the output of ICA network.

Step 9: Reversely project the 50-dimensional ICA output with noises to the input domain of ICA network.

Step 10: Do zero-padding of 100-dimensional data to the 50-dimensional data at the input domain of ICA network.

Step 11: Project the 150-dimensional data to the input domain of PCA, which corresponds to a generated sample.

Step 12: Repeat Step 7~Step 11 for the whole training samples.

Step 13: Train a subject independent classifier with whole training and generated samples.

In Fig. 3, we plot the classification accuracies for training and test samples with and without the sample generation method. The curves are the averages of 1-out of-9 validations. Also, "TG" denotes the simulation results with the proposed sample generation method. We can find that the sample generation method improves the average accuracy for training samples. For test samples, the best accuracy with the proposed sample generation method is 66.69% with four Gamma components that is higher than 65.05% without the sample generation. So, we can argue that the sample generation method can improve the classification accuracy for the implicit intention problem. Also, we can apply the sample generation method to any machine learning problem with fewer samples.

If we add i.i.d. noises to the dimensionality-reduced vector $\mathbf{z}'$ given by Eq. (14), the additive noises may disturb the data distribution since the components of $\mathbf{z}'$ are not independent but uncorrelated. So, training ICA network after PCA projection is important to add independent noises without disturbing the data distribution. Let's derive the p..d.f. of the generated samples. Since the outputs of ICA network are mutually independent, the p.d.f. of ICA outputs is given by

$$p(\mathbf{u}) = \prod_{i=1}^{50} p(u_i), \tag{16}$$

where $p(u_i)$ is the p.d.f. of the i-th ICA output. Since, $\mathbf{z}' = \mathbf{V}^{-1}\mathbf{u}$, we have

$$p_{\mathbf{z}'}(\mathbf{z}') = \frac{p(\mathbf{u})}{|\mathbf{V}^{-1}|} = \frac{p(\mathbf{V}\mathbf{z}')}{|\mathbf{V}^{-1}|}. \tag{17}$$

After constructing the vector $\mathbf{z} = [\mathbf{z}'^{T}; \mathbf{0}\ ]^{T}$ by zero padding, the p.d.f. of $\mathbf{z}$ is given by

$$p_{\mathbf{z}}(\mathbf{z}) = p_{\mathbf{z}'}(\mathbf{z}') \tag{18}$$

since the p.d.f. of zero vector is one. So, the p.d.f. of generated sample $\mathbf{x}$ is given by

$$p_{\mathbf{x}}(\mathrm{x}) = \frac{p_{\mathbf{z}}(\mathbf{W}^{T}\mathbf{x})}{|\mathbf{W}|} = \frac{p_{\mathbf{z}'}([\ \mathbf{W}^{T}\mathbf{x}\ ]_{z'})}{|\mathbf{W}|} = \frac{p(\mathbf{V}[\ \mathbf{W}^{T}\mathbf{x}\ ]_{z'})}{|\mathbf{W}||\mathbf{V}^{-1}|}, \tag{19}$$

where $[\ \mathbf{W}^{T}\mathbf{x}\ ]_{z'}$ is the dimensionality-reduced vector through PCA. If we know the p.d.f. of the i-th ICA output $p(u_i)$, we can generate samples with the same p.d.f. of training samples. However, there is no information of the p.d.f. of training samples. Therefore, we assume Gaussian p.d.f. of $p(u_i)$ and add the Gaussian noise to the output of ICA network.

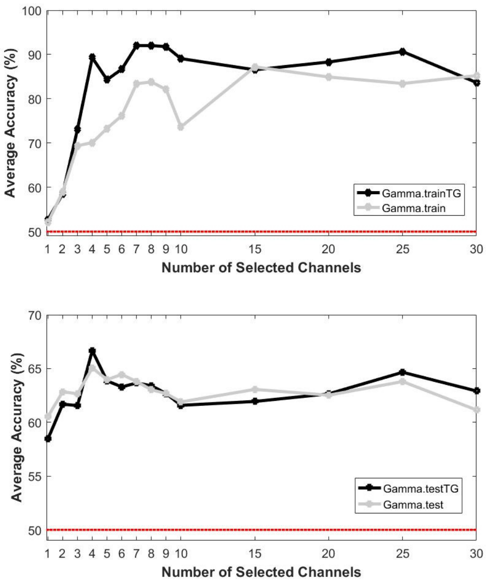

Fig. 3. Classification accuracy of implicit intention for Gamma component with increasing number of selected electrodes. “Gamma.train” and “Gamma.test” denote the accuracies for training and test samples, respectively. Also, “TG” denotes the simulation results with the proposed sample generation method.

4. CONCLUSION

In this paper, we proposed the sample generation method to improve the subject independent classification accuracy of implicit intention based on EEG signals. Because of the life-ethics problem, there are not enough number of EEG signals for classification of implicit intention. Fewer training data lead to poor performance of machine learning, and it is necessary to generate more training data for better performance.

In this point of view, we generated the training samples based on PCA followed by ICA and back-projection of ICA outputs with additive Gaussian noises to EEG data domain. We verified that the proposed sample generation improved the best test performance of implicit intention from 65.05% to 66.69% with Gamma components. The proposed sample generation method can be applied to any machine learning problem with fewer samples.

ACKNOWLEDGMENT

The author would like to thank Dr. Suh-Yeon Dong and Professor Soo-Young Lee for their helpful support and discussions.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, “Deep Learning,” Nature, vol. 521, May. 2015, pp. 436-444.

[2] A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” Advances in Neural Information Processing 25, 2012, MIT Press, Cambridge, MA.

[3] B. K. Kim, J. Roh, S. Y. Dong, and S. Y. Lee, “Hierarchical Committee of Deep Convolutional Neural Networks for Robust facial Expression Recognition,” J. Multimodal User Interfaces, vol. 10, issue. 2, Jun. 2016, pp. 173-189.

[4] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, “Brain-Computer Interfaces for Communication and Control,” Clin. Neurophysiol, vol. 113, 2002, pp. 767-791.

[5] B. Blankertz, G. Dornhege, M. Krauledat, K. R. Muller, V. Kunzmann, F. Losch, and G. Curio, “The Berlin Brain-Computer Interface: EEG-based Communication without Subject Training,” IEEE Trans. Neural Sys. and Rehabil., vol. 14, no. 2, 2006, pp. 147-152.

[6] J. H. Lim, H. J. Hwang, C. H. Han, K. Y. Jung, and C. H. Im, “Corrigendum: Classification of Binary Intentions for Individuals with Impaired Oculomotor Function: ‘eyes-closed’ SSVEP-based Brain-Computer Interface,” J. Neural Eng., vol. 10, 2013, pp. 1-9.

[7] L. A. Farwell and E. Donchin, “The truth Will Out: Interrogative Polygraphy (“Lie Detection”) with Event-related brain Potentials,” Psychophysiology, vol. 28, 1991, pp. 531-547.

[8] C. Davatzikos, K. Ruparel, Y. Fan, D. G. Shen, M. Acharyya, J. W. Loughead, R. C. Gur, and D. D. Langleben, “Classifying Spatial Patterns of Brain Activity with Machine Learning Methods: Application to Lie Detection,” Neuroimage, vol. 28, 2005, pp. 663-668.

[9] S. Y. Dong, B. K. Kim, and S. Y. Lee, “EEG-based Classification of Implicit Intention during Self-relevant Sentence Reading,” IEEE Trans. Cybernetics, vol. 46, Nov. 2016, pp. 2535-2542.

[10] S. Y. Dong, B. K. Kim, and S. Y. Lee, “Implicit Agreeing/Disagreeing Intention While Reading Self-relevant Sentences: A Human fMRI Study,” Social Neuroscience, vol. 11, issue. 3, 2016, pp. 221-232.

[11] S. H. Oh, “Subject Independent Classification of Implicit Intention Based on EEG Signals,” Int. Journal of Contents, vol. 12, no. 3, Sep. 2016, pp. 12-16.
[12] S. Haykin, *Neural Networks*, MacMillan, 1994.
[13] T. W. Lee, M. Girolami, and T. J. Sejnowski, “A Unifying Information Theoretic Framework for Independent Component Analysis,” Computers & Mathematics with Applications, vol. 31, no. 11, Mar. 2000, pp. 1-21.
[14] A. J. Bell and T. J. Sejnowski, “An Information-Maximisation Approach to Blind Separation and Blind Deconvolution,” Neural Computation, vol. 7, no. 6, Nov. 1995, pp. 1004-1034.
[15] S. Amari, “Natural Gradient Works Efficiently in Learning,” Neural Computation, vol. 10, Feb. 1998, pp. 251-276.
[16] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer Science+Business Media, LLC, New York, 2006.
[17] T. W. Lee, M. Girolami, and T. J, Sejnowski, “Independent Component Analysis Using an Extended Infomax Algorithm for Mixed Sub-Gaussian and Super-Gaussian Sources,” Neural Computation, vol. 11, 1999, pp. 417-441.
[18] C. C. Chang and C. J. Lin, “LIBSVM: A library for support vector machines,” ACM Trans. Intell. Syst. Technol., vol. 2, no. 3, Apr. 2011, pp. 1-27. http://www.csie.ntu.edu.tw/~cjlin/libsvm

Sang-Hoon Oh
He received his B.S. and M.S degrees in Electronics Engineering from Pusan National University in 1986 and 1988, respectively. He received his Ph.D. degree in Electrical Engineering from Korea Advanced Institute of Science and Technology in 1999. From 1988 to 1989, he worked for LG semiconductor, Ltd., Korea. From 1990 to 1998, he was a senior research staff in Electronics and Telecommunication Research Institute (ETRI), Korea. From 1999 to 2000, he was with Brain Science Research Center, KAIST. In 2000, he was with Brain Science Institute, RIKEN, Japan, as a research scientist. In 2001, he was an R&D manager of Extell Technology Corporation, Korea. Since 2002, he has been with the Department of Information Communication Engineering, Mokwon University, Daejeon, Korea, and is now an associate professor. Also, he was with the Division of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, USA, as a visiting scholar from August 2008 to August 2009. His research interests are machine learning, speech signal processing, pattern recognition, and bioinformatics.