International Journal of Fuzzy Logic and Intelligent Systems, vol. 10, no. 1, March 2010, pp. 59−64

Membership Function-based Classification Algorithms for Stability improvements of BCI Systems

Hong-Gi Yeom and Kwee-Bo Sim*

School of Electrical and Electronics Engineering, ChungAng University
221, HeukSeok-Dong, DongJak-Gu, Seoul, 156-756, Korea
Tel : +82-2-820-5319, Fax : +82-2-817-0553, E-mail : kbsim@cau.ac.kr

Abstract

To improve system performance, we apply the concept of membership function to Variance Considered Machines (VCMs) which is a modified algorithm of Support Vector Machines (SVMs) proposed in our previous studies. Many classification algorithms separate non-linear data well. However, existing algorithms have ignored the fact that probabilities of error are very high in the data-mixed area. Therefore, we make our algorithm ignore data which has high error probabilities and consider data importantly which has low error probabilities to generate system output according to the probabilities of error. To get membership function, we calculate sigmoid function from the dataset by considering means and variances. After computation, this membership function is applied to the VCMs.

Key Words : Classification Algorithms, Optimal Hyperplanes, Membership Functions, Support Vector Machines, Variance Considered Machines.

1. Introduction

Intelligent systems perform various complex job as a result of advanced technology. For example, the system recognizes speech to help disabled people and they can detect human's face based on LDA or PCA or Bayesian [1-3]. Furthermore, intelligent robot can interact with you because the robot knows your body gestures [4].

To perform this task, a lot of techniques are required such as preprocessing, feature extraction and data classification. Above all, one of the most important skills is pattern classification. Pattern classification is the act of taking in raw data and taking an action based on the "category" of the pattern [5]. The accuracy of pattern classification affects diverse problems, like security systems or selling expensive salmon at the price of sea bass.

One of the most used classification algorithm is a neural network. The neural network repeatedly adjusts the weights of the connections in the network so as to minimize a measure of the desired output vector [6]. This method is widely used because it is available by learning without system modeling. However, the three main problems encountered when minimizing the empirical risk using the backpropagation method are as follows [7].

1) The empirical risk functional has many local minima.
2) Convergence to a local minimum can be rather slow.
3) The sigmoid function has a scaling factor which affects the quality of the approximation.

For this reason, Vapnik suggested new classification method, Support Vector Machines (SVMs), in 1995 [8]. It has maximal margin using support vectors and this method can separate nonlinear data linearly with kernel function [9].

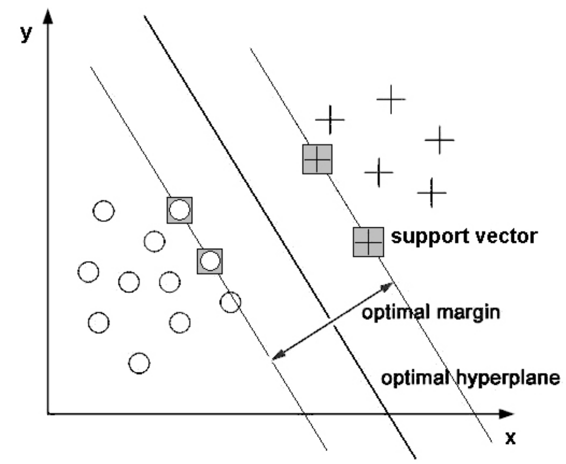

Fig 1. Discriminant of Support Vector Machine (SVM). SVM finds W which has maximal margin. However, it determines optimal hyperplanes on the center of support vectors of the two classes. It can cause more error.

Therefore, the SVM has been used in various fields such as financial time series forecasting, image processing and circuit design [10-12].

The SVM implements the following idea: it maps the input vectors into some high-dimensional feature space Z through some non-linear mapping [8]. In this space, the SVM constructs

Manuscript received Dec. 7, 2009; revised Jan. 30, 2010.
* Corresponding author
This work was supported by a National Research Foundation of Korea Grant funded by the Korean Government (2009-0083866)

optimal hyperplanes which have maximal margin using the support vector as shown in Fig. 1.

SVM is a good way to find a vector W which is a slope of optimal hyperplanes. Nevertheless, SVM neglects to consider variances and prior probabilities. $P(w_k)$ Therefore, the algorithm is incomplete in reducing error probabilities. For this reason, we proposed a modified algorithm variance considered machine (VCM). This method uses the same vector W of SVM; however, it changes the values of b, y-intercept, by considering variances and prior probabilities. Hence, VCMs have maximal margin and the algorithms decrease reducible error probabilities.

However, in case of Brain-Computer Interface (BCI), it is required to modify the algorithm when control the system according to the results of the classification [13]. BCI is a system for communication and control with thoughts [14]. It can help severely motor-disabled persons to communicate and control their environments through many applications [15-18]. BCI systems can substitute for the loss of normal neuromuscular outputs by enabling people to interact with their environment through brain signals rather than thought muscles [19]. Thus, for examples, a person can use electrophysiological signals such as electroencephalographic (EEG) activity or cortical neuronal activity to indicate “yes” or “no” to control a cursor on a computer screen or to control a neuroprosthetic arm [20].

Although, data, which is close to hyperplanes, has high error probabilities, many classification algorithm consider the data as others. It makes sense if the system have to make a decision whether it belongs to class A or class B. However, as we mentioned before, it is essential to control the system according to error probabilities in BCI case. Therefore, data which have high error probabilities should affect less influence to output than data of low error probabilities.

Fig. 2 and Fig. 3 represent Time/Frequency Analysis of Electroencephalogram (EEG) signals on C3 and C4, electrode locations, when subjects imagine movements of their right or left arm. In the figure, axis of X expresses time flow and axis of Y means frequency. And the colors represent magnitude of the signals and the vertical dotted line is a point that subjects imagine the movements. Therefore, the changes of EEG signals are shown according to the flow of time and frequency.

We can see differences of the signals at alpha and beta waves from the figures. However, classification of subject’s intention is not easy because there are changes according to subjects. Although same person makes experiments, the data of the signals is different.

In case of not only BCI but also many other system, it is necessary to ignore the data which can occur error.

In this paper, we apply the concept of fuzzy to classification methods to solve these problems by expressing the results of discrimination as grade of membership. Although it is a simple idea it will be useful on systems which are permitted to ignore high error probability data. In the second session, we introduced new classification algorithms, Variance Considered

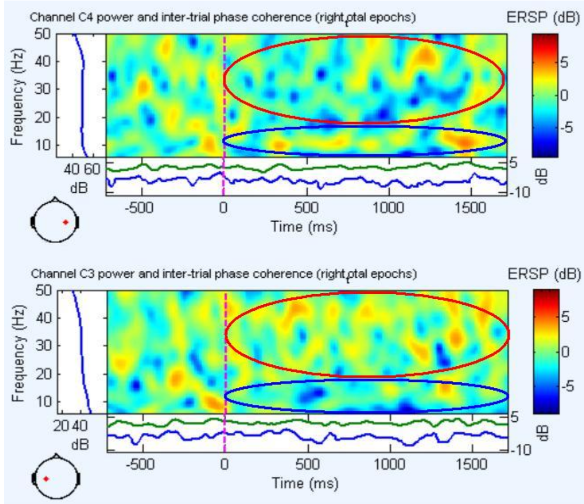

Fig. 2 ERSP comparison during the imaginary movement of right arm at C3 and C4 [10]

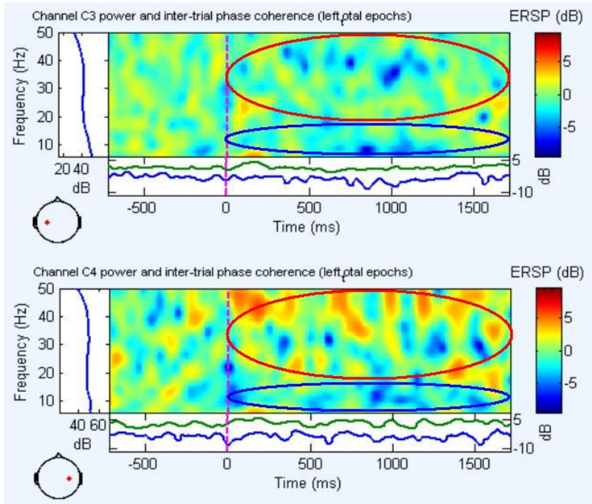

Fig. 3 ERSP comparison during the imaginary movement of left arm at C3 and C4

Machines (VCMs), and calculated sigmoid function as a membership function for m-Variance Considered Machines (m-VCMs). In the third session, we showed Matlab simulation results of Support Vector Machines (SVMs) and classification algorithm based on membership function. In section 4, we summarize our study and give our conclusions.

2. Proposed Algorithms

We will introduce proposed algorithms, after illustrations of optimal hyperplanes of SVMs, VCMs and concept of membership functions.

2.1 Optimal Hyperplanes of SVMs

For the explanation of optimal hyperplanes of SVMs, suppose we are given dataset of labeled training patterns (1).

The data are linearly separable if there are a vector w and a scalar b such that the inequalities (Eq. (4)) are valid for all elements of the training set (Eq. (3)):

$$(y_1, x_1), \dots, (y_n, x_n), \qquad y_i \in \{-1, 1\}, \tag{3}$$

$$\begin{aligned} w \cdot x_i + b \geq 1 \qquad & if\ y_i = 1, \\ w \cdot x_i + b \leq -1 \qquad & if\ y_i = -1, \end{aligned} \tag{4}$$

There is only one direction $w/|w|$ which satisfies the condition that the distance between projections of the training vectors of the two different classes is maximal, as shown in Fig. 1. This distance $\rho(w,b)$ is calculated as follows:

$$\rho(w,b) = \min_{\{x:y=1\}} \frac{x \cdot w}{|w|} - \max_{\{x:y=-1\}} \frac{x \cdot w}{|w|}. \tag{5}$$

The optimal hyperplane (w_0, b_0) should maximize the distance in Eq. (5) subject to inequality constraints (Eq. (4)) to have maximal margin. Therefore, Eq. (6) is computed from Eqs. (5) and (4). This means that the optimal hyperplane is the unique hyperplane that minimizes $w \cdot w$:

$$\rho(w_0, b_0) = \frac{2}{|w_0|} = \frac{2}{\sqrt{w_0 \cdot w_0}}. \tag{6}$$

2.2 Variance-Considered Machines

VCMs use the same vector w as SVMs to decide the direction of the optimal hyperplanes in order to obtain maximal margin. However, SVMs construct optimal hyperplanes on the center between support vectors of each class without consideration of variances and prior probabilities. This can cause more errors when each class has different variances and prior probabilities. Therefore, we propose modified algorithms which shift the optimal hyperplanes according to the variances and prior probabilities. The key idea of VCMs is that the algorithm shifts the optimal hyperplanes from the middle of the support vectors to the point that has the same probability of belonging to each class.

The proposed VCM algorithms only move the position of the optimal hyperplanes to reduce errors. Thus, kernel functions of SVMs can be applied to VCMs, and VCMs can also classify nonlinear data.

Because variances on the optimal hyperplane direction do not affect pattern classification results, we only considered variances on the vector u which is orthogonal to the optimal hyperplane, as illustrated in Fig. 4. To formalize this idea, consider the set of data $\{x_1^{(i)}; i = 1, \dots, m\}$ marked by a circle and the set of data $\{x_2^{(j)}; j = 1, \dots, n\}$ marked by a cross in Fig. 4. The projected vector of x onto u is expressed as $x^T u$. Therefore, the mean and variance of projected data of $x_1^{(i)}$ are represented by Eq. (7), and the mean and variance of $x_2^{(j)} u$ are expressed as Eq. (8).

$$\mu_1 = E\left\{ {x_1^{(i)}}^T u \right\},\ \sigma_1^2 = E\left\{ \left(\mu_1 - {x_1^{(i)}}^T u \right)^2 \right\}, \tag{7}$$

$$\mu_2 = E\left\{ {x_2^{(j)}}^T u \right\},\ \sigma_2^2 = E\left\{ \left(\mu_2 - {x_2^{(j)}}^T u \right)^2 \right\}. \tag{8}$$

The probability density function (PDF) belonging to the class c_k can be calculated from these means and variances by Eq. (9) if we assume that the data follow a normal distribution:

$$P(u \mid c_k) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left[-\frac{1}{2} \left(\frac{u - \mu_k}{\sigma_k} \right)^2 \right]. \tag{9}$$

This is the class-conditional probability density function. Bayes' formula was used to find the point where the posterior probabilities of each class are equal. For convenience of calculation, the following discriminant function can be substituted for Bayes' formula [5]:

$$g_k(u) = \ln P(u \mid c_k) + \ln P(c_k), \tag{10}$$

where $P(c_k)$ is the prior probability. From Eqs. (9) and (10), discriminant functions can be computed as follows:

$$g_k(u) = -\frac{1}{2}\left(\frac{u - \mu_k}{\sigma_k} \right)^2 - \frac{1}{2}\ln 2\pi - \frac{1}{2}\ln \sigma_k + \ln P(c_k), \tag{11}$$

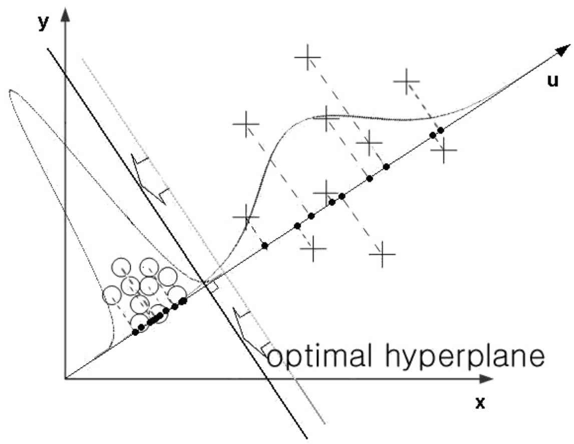

Fig. 4 Discriminant of Variance Considered Machine (VCM). VCM uses same W of SVM. However, it changes y intercept by considering variances and prior probabilities [11].

where $-\frac{1}{2}\ln 2\pi$ is independent of k. Therefore, $-\frac{1}{2}\ln 2\pi$ can be omitted and $P(c_k)$ can be calculated as follows:

$$P(c_k) = \frac{n_k}{n_{all}}, \tag{12}$$

where n_k is the number of data c_k, and n_{all} is the number of all data. The point which has the same posterior probabilities for two classes can be found by

$$g_1(u) = g_2(u). \tag{13}$$

We denote this point $\hat{u}$. Finally, new optimal hyperplanes moved to the point $\hat{u}$ can be expressed as

$$w_0 \cdot x - \hat{u} = 0. \tag{14}$$

2.3 Membership Function based VCMs (m-VCMs)

m-VCMs use membership function instead of hyperplanes. Membership functions represent membership grades that data belongs to certain class. For examples, the membership function of class “Hot” can be expressed as 0~1 when temperature is 25~45ºC and it will be determined by user.

The proposed methods use sigmoid functions as membership functions. The integral of equation (3) is sigmoid function. It can be simplified as equation (4).

$$P(u \mid w_k) = \frac{1}{1 + \exp\left(-\dfrac{u - \mu_k + \sigma_k}{\sigma_k} \right)} \tag{4}$$

It is a sigmoid function shifted to the mean as much as twice standard deviation for relevance of the membership function. We use it as a membership function of m-VCM. The reason why we use sigmoid function instead of Gaussian function is that if we use Gaussian function as a membership function many data which are far away from the hyperplane will be considered unimportantly. And we move the center of sigmoid function to the optimal hyperplane. Therefore, data existing in the center of Class have high membership grades.

System output is determined according to the results of classification. At this time, output magnitude of class j is calculated as follows.

$$Output_j = p(w_j \mid x) - \sum_{i \neq j} P(w_i \mid x) \tag{5}$$

If $Output_j$ is lager then threshold, determined by user, then the output is generated as equation (5). For example, existing system control the wheel chair to only right or left direction. However, proposed methods turn the wheel chair 10º if $Output_j$ is 10% or 30º if $Output_j$ is 30%. Therefore, the system can be more robust.

3. Simulation

We generate random 50 data for each class on 2D space using Matlab program. Following Fig. 5 and Fig. 6 show how they separate classes.

Results classified by SVM with RBF kernel is represented in Fig. 5. We can know that SVM separates class well although it is a non-linear case. However, SVM neglect to consider error probabilities as we mentioned before.

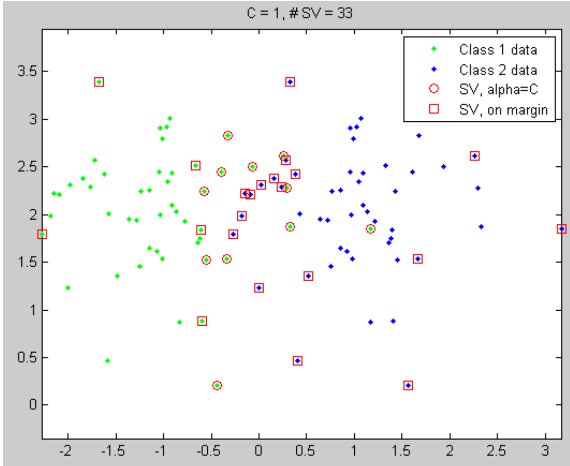

Fig. 5 Classification of Support Vector Machine (SVM) with RBF kernel. It separate non-linear well. However, data in the mixed area can have high error probabilities.

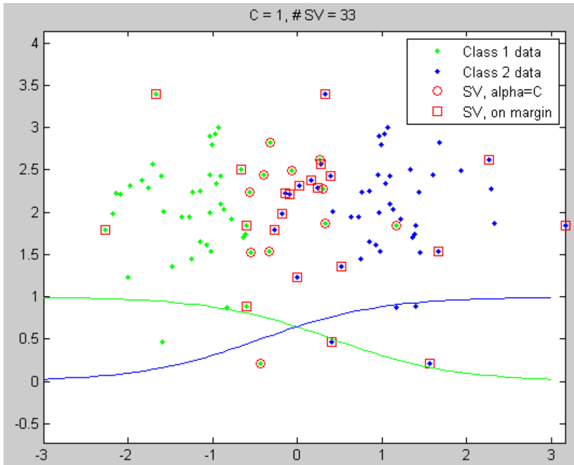

Fig. 6 Membership Function calculated by VCM. Green line shows the membership function of Class 1 and blue line represents the membership function of Class 2

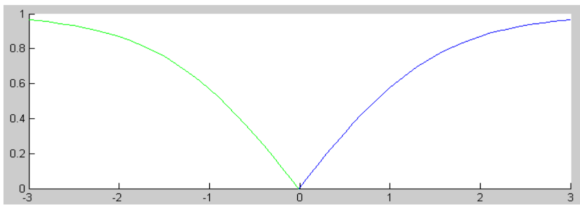

Fig. 7 System outputs corresponding to each data. Therefore, almost data near the hyper plane do not affect output.

For same dataset, membership functions of each class are calculated from the equation (4) and expressed in Fig. 6. From the equation (5), system outputs corresponding to each data are obtained and shown in Fig. 7.

We can get a new concept classification algorithm by applying $Output_j$, equation (5), or sigmoid function, equation (4), to the VCM. It is shown in Fig. 8.

This m-VCM takes advantages of VCM, maximum margin and low error probabilities, and it is more robust for input which has high error probabilities because of its membership function.

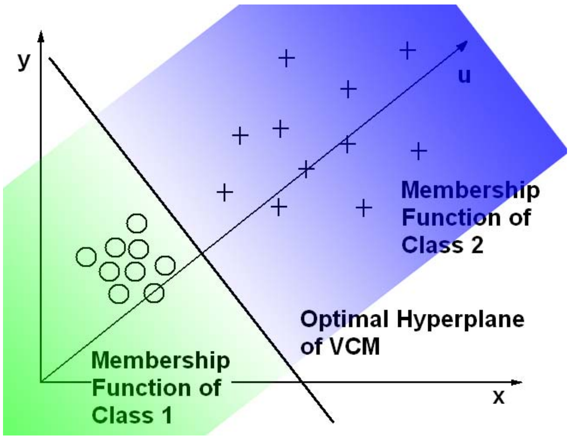

Fig. 8 Membership function based VCM (m-VCM). Changes of color represent the membership grade of each Class and strength of system output

4. Conclusion

Although many classification algorithms like SVM separate non-linear data well, the data, near hyperplanes, are likely to be error. Therefore, we applied concept of membership function to classification algorithm VCM. Sigmoid function is calculated from the dataset parameters as a membership function. In the third section, we compared simulation results of SVM and proposed methods. We believe that although it is a simple idea it will be very useful in various intelligent system.

References

[1] H. G. Yeom, I. H. Jang and K. B. Sim, "Variance Considered Machines: Modification of Optimal Hyperplanes in Support Vector Machines", *IEEE Int. Symp. Industrial Electronics*, Seoul, Korea, July, pp. 1144-1147, 2009.

[2] B. M. Belgacem, "Speech Recognition for Disabilities People", *Information and Communication Technologies, 2006. ICTTA "06. 2nd*, Vol. 1, pp.864 – 869, 2006.

[3] H.K Ekenel, J. Stallkamp, H. Gao, M. Fischer and R. Stiefelhagen, "Face Recognition for Smart Interactions", *IEEE International Conference on Multimedia and Expo*, Vol. 2-5 July pp.1007 – 1010, 2007.

[4] S. W. Lee, "Automatic Gesture Recognition for Intelligent Human-Robot Interaction", *7th International Conference on Automatic Face and Gesture Recognition*, Vol. 2-6 April 2006. pp. 645 – 650, 2006.

[5] O. D. Richard, E. H. Peter and G. S. David, *Pattern classification*, Wiley, New York, Second edition, 2001.

[6] D. E. Rumelhart, G. E. Hintont and R. J. Williams, "Learning representations by back-propagating errors", *Nature*, Vol.323 9 Oct. 1986.

[7] V. N. Vapnik, " An Overview of Statistical Learning Theory", *IEEE Transactions on neural networks*, Vol. 10, No. 5, Sep. 1999.

[8] C. CORTES and V. VAPNIK, "Support-vector networks", *Machine Learning*, 20, pp. 273–297, 1995.

[9] B. Schoölkopf, K. K. Sung, C. J. C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, "Comparing Support Vector Machines with Gaussian Kernels to Radial Basis Function Classifiers", *IEEE Transactions on signal processing,* Vol. 45, No. 11, Nov, 1997.

[10] L. J. Cao and F. E. H. Tay, "Support Vector Machine With Adaptive Parameters in Financial Time Series Forecasting", *IEEE Transactions on neural networks*, Vol. 14, No. 6, Nov, 2003.

[11] FU, Y., SHEN, R., and LU, H. "Watermarking scheme based on support vector machine for colour images", *Electronics Letters*, 40, (16), pp. 986–987, 2004.

[12] D. ANQUITA, S. RIDELLA and S. ROVETTA, "Circuital implementation of support vector machines", *Electronics Letters*, 34, (16), pp. 1596–1597, 1998

[13] H. G. Yeom and K. B. Sim, "m-Variance Considered Machines: Membership Function based Classification Algorithms", *The 10th International Symposium on Advanced Intelligent Systems (ISIS2009),* pp. 165~168, August 17-19, 2009.

[14] J. Lehtonen, P. Jylänki, L. Kauhanen, and M. Sama, "Online Classification of Single EEG Trials During Finger Movements", *IEEE Transaction on Biomedical Engineering*, Vol. 55, No. 2, Frb. 2008.

[15] L. Kauhanen, P. Jylanki, J. Lehtonen, P. Rantanen, H. Alaranta and M. Sams, "EEG-Based Brain-Computer Interface for Tetraplegics," *Hindawi Publishing Corporation Computational Intelligence and Neuroscience* Vol. 2007, Article ID 23864, 11pages doi:10.1155/2007/23864.

[16] L. Citi, R. Poli, C. Cinel, and F. Sepulveda, "P300-Based BCI Mouse With Genetically-Optimized Analogue Control," *IEEE Trans. On Neural System and Rehabilitation Engineering*, Vol. 16, No. 1, pp. 51-61, Feb. 2008.

[17] Gernot R. Müller-Putz, Reinhold Scherer, Gert Pfurtscheller, Rüdiger Rupp, "EEG-based neuroprosthesis control: A step towards clinical practice", *Neuroscience Letters*, 382, pp. 169-174, 2005.

[18] F. Cincotti, F. Aloise, S. Bufalari, G. Schalk, G. Oriolo, A, et al., "Non-Invasive Brain-Computer Interface System to Operate Assistive Device", *Proceedings of the 29th Annual International Conference of the IEEE EMBS* Cité Internationale, Lyne, France, Aug. 23-26, 2007.

[19] J. R.Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, T. M. Vaughan, "Brain–computer interfaces for

communication and control", *Clinical Neurophysiology*,Vol.113, Issue 6, pp.767 June 2002.

[20] J. J. Daly, J. R. Wolpaw, "Brain-computer interfaces in neurological rehabilitation", *Lancet Neurol*. Vol. 7, pp. 1032-1043, 2008.

Hong-Gi Yeom

He received his B.S and M.S. degree in the Department of Electrical and Electronics Engineering from Chung-Ang University, Seoul, Korea, in 2008 and 2010 respectively. His research interests include Brain-Computer Interface, Functional Electrical Stimulation, Machine Learning, Wearable Robot etc

Kwee-Bo Sim

He received his B.S. and M.S. degrees in the Department of Electronic Engineering from Chung-Ang University, Korea, in 1984 and 1986 respectively, and Ph.D. degree in the Department of Electronics Engineering from The University of Tokyo, Japan, in 1990. Since 1991, he is currently a Professor. His research interests include artificial life, emotion recognition, ubiquitous intelligent robot, intelligent system, computational intelligence, intelligent home and home network, ubiquitous computing and Sense Network, adaptation and machine learning algorithms, neural network, fuzzy system, evolutionary computation, multi-agent and distributed autonomous robotic system, artificial immune system, evolvable hardware and embedded system etc. He is a member of IEEE, SICE, RSJ, KITE, KIEE, KIIS, and ICROS Fellow.