International Journal of Innovative Computing, Information and Control
Volume **5**, Number **12(B)**, December **2009**

pp. **4935–4943**

EMG PATTERN CLASSIFICATION USING HIERARCHICAL NETWORK BASED ON BOOSTING APPROACH

Masaru Okamoto, Yukihiro Matsubara

Department of Intelligent Systems
Graduate School of Information Sciences
Hiroshima City University
3-4-1, Ozuka-Higashi, Asa-Minami-Ku, Hiroshima, 731-3194, Japan
{ okamoto; matsubar }@hiroshima-cu.ac.jp

Keisuke Shima and Toshio Tsuji

Department of Artificial Complex Systems Engineering
Graduate School of Engineering
Hiroshima University
1-4-1, Kagamiyama, Higashi-Hiroshima, 739-8527, Japan
{ shima; tsuji }@bsys.hiroshima-u.ac.jp

Received July 2008; revised December 2008

Abstract. *This paper proposes a new electromyogram (EMG) pattern classification method using probabilistic neuralnetworks based on boosting approach* [1]*. Since the proposed method automatically constructs a suitable classification network from measured EMG signals, there is no need to set the structure of network in advance. To verify the feasibility of the proposed method, phoneme classification experiments are conducted using EMG signals measured from mimetic and cervical muscles. In these experiments, the proposed method achieved high classification rates.*
Keywords: Pattern classification, Probabilistic neural network, Electromyogram, EMG signals

1. **Introduction.** Electromyograms (EMG) pattern classification has been used to devise elaborate human-machine interfaces for people with physical disabilities [2, 3]. Generally, such pattern classification is performed by estimating the relationship between the EMG signals as feature vectors and the corresponding intentions as class labels. For general pattern classification problems, various soft computing approaches, such as fuzzy inference and self organizing maps (SOM) and so on, have been proposed [3-7], and in recent years, EMG pattern classification methods using various types of classifiers have also been proposed [8-11]. In particular, neural networks (NN) have been demonstrated as a promising classification tool, since their learning ability allows them to find optimum non-linear relationships between classes and feature patterns from data sets [8, 11-13]. However, to effectively use NNs as the classifiers for applications, several fundamental problems, such as the choice of network structure, learning convergence and local minima, must be solved.

A probabilistic neural network (PNNs), which estimates the probability density function of patterns, has been proven to be an efficient and important method for pattern classification. In particular, Tsuji et al. proposed a feedforward PNN, a log-linearized Gaussian mixture network (LLGMN) based on the Gaussian mixture model (GMM) and a log-linear model [14]. The LLGMN has been successfully applied to pattern classification of bioelectric signals, such as electromyograms [15] and electrocardiograms [14, 16], and has been used to develop human interface applications, such as prosthetic devices and EMG-based pointing devices [2, 17, 18].

However, to estimate the LLGMN parameters, the GMM number of each class must be fixed beforehand. When the GMM number is fixed at an unsuitable value, the LLGMN training cannot avoid convergence to a local minimum for some initial weights and training data. Therefore, better classification performance requires estimation of an optimum LLGMN structure. Also, in other classification methods, the estimation approach of suitable parameters is important. In order to overcome these problems, several methods, such as information criterion and the variational Bayes approach, have been widely used as the criterion for the structure of a model [19, 20]. In these methods, a suitable learning model can be selected based on discrimination accuracy, likelihood and model complexity. Although these methods select a suitable model structure, all possible models must be evaluated based on the criterion. Thus, it takes a long time to estimate a suitable model structure.

There has also been growing interest in a boosting approach for the construction of classification systems with simple classifiers [21-23]. A general boosting procedure can combine inaccurate and simple classifiers to improve the discrimination accuracy of a classification system. Therefore, this approach eliminates the need for evaluation of unnecessary models. Up to the present, we have proposed a hierarchical classification method that can automatically construct the classification models through a learning network [1]. In this method, the LLGMN is utilized in order to create a simple and weak classifier. The proposed method can estimate the number of LLGMNs corresponding to the pattern complexity, according to statistical information obtained from the training data.

In this paper, an EMG pattern classification method is proposed based on hierarchical pattern classification based on boosting approach. By using boosting approach, it is expected that the structure of the classification network can be estimated automatically corresponding to the complexity of the EMG patterns.

The next section provides the details of the LLGMN structure and shows the pattern classification method for constructing a suitable model using the boosting approach. In Section 3, the EMG pattern classification method is proposed. Then, the phoneme classification and experimental results are presented in Section 4. We discuss our findings and draw conclusions in the last section.

2. **Classification Method** [1].

In this method, the LLGMNs are used in order to create simple classifiers for the classification of input vectors to produce binary splits. The structure of each classifier is a hierarchical tree using the LLGMN as each non-terminal node. By combining classifiers based on a boosting approach, the network can discriminate complex data, and calculate a posteriori probability for the training data. The structure of LLGMN and the constructing algorithm are explained below.

2.1. **LLGMN** [14].

LLGMN is based on a log-linear model and a Gaussian mixture model (GMM). It calculates posteriori probability for the training data. In this method, LLGMN is utilized for partition at the non-terminal node of the hierarchical tree.

The structure of LLGMN is shown in Figure 1. In order to represent a normalized distribution corresponding to each component of GMM as weight coefficients of NN, the input vector $\boldsymbol{x}(\in \Re^D)$ is converted into the modified input vector $\boldsymbol{X}$ as follows:

$$\boldsymbol{X} = \{1, \boldsymbol{x}^T, x_1^2, x_1x_2, \cdots, x_2^2, \cdots, x_2x_D, \cdots, x_D^2\}^T \tag{1}$$

The first layer of LLGMN consists $H = 1 + D(D+3)/2$ units, which correspond to the dimension of the input vector $\boldsymbol{X}$, and the identity function is used for the activation function of each unit. The outputs of the first layer multiplied by weight $w_h^{(k,m)}$ are

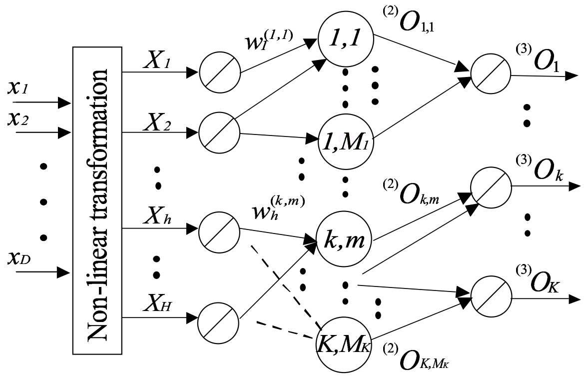

FIGURE 1. The structure of LLGMN

transmitted to the second layer. Where $w_h^{(K,M_K)} = 0$, K and M_K denote the number of classes (patterns) and components belonging to class M, respectively. In this layer, LLGMN calculates the posteriori probability of each Gaussian component $\{k, m\}$. The unit k in the third layer integrates the outputs of M_k units in the second layer.

$$^{(2)}I_{k,m} = \sum_{k=1}^{H} {}^{(1)}O_h w_h^{(k,m)} \tag{2}$$

$$^{(2)}O_{k,m} = \frac{\exp(^{(2)}I_{k,m})}{\sum_{k'=1}^{K}\sum_{m'=1}^{M_{k'}} \exp(^{(2)}I_{k',m'})} \tag{3}$$

The relationship between the input $^{(3)}I_k$ and the output O_k in the third layer is

$$^{(3)}I_k = \sum_{m=1}^{M_k} {}^{(2)}O_{k,m} \tag{4}$$

$$^{(3)}O_k = {}^{(3)}I_k \tag{5}$$

The output of the third layer $^{(3)}O_k$ corresponds to the posterior probability $P(k|\boldsymbol{x})$ of class k given the input vector $\boldsymbol{x}$, and the former can be used to evaluate the ambiguity of a classification result.

This network has the ability of adaptive learning for statistical properties of data. It can discriminate data with complex distributed structure, and in comparison to the conventional method [24] using normal distribution restricted the parameter.

2.2. **Structure of the network.** Initially, the network consists of C classifiers, corresponding to the number of classified classes. C is the number of classes of training data. Each classifier achieves a binary classification to calculate the posteriori probability of the cth class ($c = 1, 2, \cdots, C$). For binary classification, the parameter of LLGMN K is set as 2. $L_c^{(q)}(\boldsymbol{x})(c = 1, \cdots, C, q = 1, \cdots, Q_c)$ is the posteriori probability calculated by classifier, where Q_c is the number of classifiers used for the classification of the cth class added based on boosting approach. Then, the posteriori probability $O_c(\boldsymbol{x})$ is given as in Equation 6.

$$O_c(\boldsymbol{x}) = \max_{q=1,\cdots,Q_c} (L_c^{(q)}(\boldsymbol{x})) \tag{6}$$

The structure of this method is shown in Figure 2.

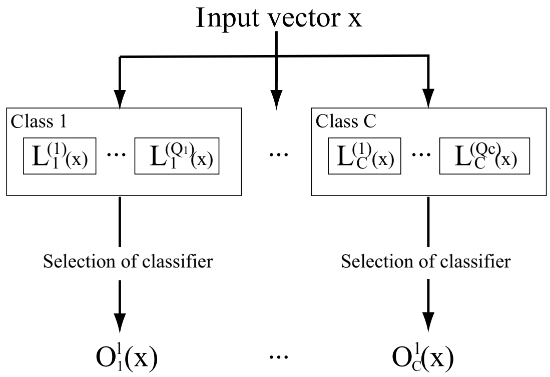

Figure 2. The structure of the proposed method

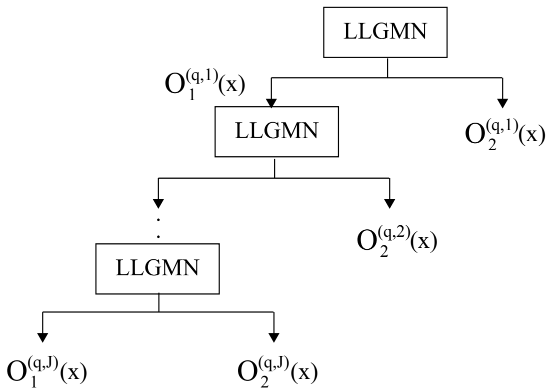

Figure 3. The structure of classifier.

2.3. **Learning of hierarchical classifier.** Structure of classifier is a hierarchical tree using LLGMN. When the learning of the cth class is performed, the training data is divided into two groups, G_c and $G_{\overline{c}}$, where G_c is a set obtained from the training data belonging to class c, and $G_{\overline{c}}$ is the complementary set of G_c. An example of constructed classifier is shown in Figure 3.

Consider a training set $\{\boldsymbol{x}^{(n)}, \boldsymbol{T}^{(n)}\}$ $(n = 1, \cdots, N)$, where $\boldsymbol{T}^{(n)} = (T_1^{(n)}, T_2^{(n)})$. If the input vector $\boldsymbol{x}^{(n)}$ belongs to class c, $T_1^{(n)} = 1$, and $T_2^{(n)} = 0$. An energy function according to the minimum log-likelihood training criterion can be derived as:

$$E = \sum_{n=1}^{N} J^{(n)} = -\sum_{n=1}^{N} \sum_{k=1}^{2} T_k^{(n)} \log^{(3)} O_k \tag{7}$$

In the training process, modification of the LLGMN's weight $\Delta w_h^{(k,m)}$ is defined as:

$$\Delta w_h^{(k,m)} = -\eta \sum_{n=1}^{N} \frac{\partial J^{(n)}}{\partial w_h^{(k,m)}} \tag{8}$$

$$\begin{aligned}\frac{\partial J^{(n)}}{\partial w_h^{(k,m)}} &= \frac{\partial}{\partial w_h^{(k,m)}}(-\sum_{k=1}^{2} T_k^{(n)} \log{}^{(3)} O_k) \\ &= ({}^{(2)}O_{k,m} - \frac{{}^{(2)}O_{k,m}}{{}^{(3)}O_k} T_k^{(n)}) X_h^{(n)} \end{aligned} \tag{9}$$

where $\eta > 0$ is the learning rate.

LLGMNs are added to avoid the misclassification of training data belonging to $G_{\overline{c}}$. To evaluate the misclassification accuracy of training data belonging to $G_{\overline{c}}$, an evaluation function is defined as Equation 12.

$$F' = \frac{|D(\overline{c}, c)|}{|G_{\overline{c}}|} \tag{10}$$

If F' is greater than the threshold Th', more LLGMNs are added hierarchically, and are trained using a two class set $D(c, c)$ and $D(\overline{c}, c)$. Then, the posteriori probability $L_c^{(q)}(\boldsymbol{x})$, which is calculated by the qth classifier, is defined as,

$$L_c^{(q)}(\boldsymbol{x}) = 1 - \sum_{j=1}^{J_q} \left(\left(\prod_{j'=0}^{j-1} {}^{(3)}O_1^{(q,j')}(\boldsymbol{x}) \right) {}^{(3)}O_2^{(q,j)}(\boldsymbol{x}) \right) \tag{11}$$

where J_q is the number of LLGMNs added to the qth classifier, $O_1^{(q,j)}(\boldsymbol{x})$ is the posteriori probability calculated by the jth LLGMN in the qth classifier and $O_1^{(q,0)}(\boldsymbol{x})$ is set to 1. By combining the LLGMN hierarchically to construct a network, the misclassification of data belonging to class c' can be avoided.

2.4. **Construction network.** In the proposed method, the addition and learning of the classifier is repeated for each class. A classifier is initially trained to classify the training data into G_c and $G_{\overline{c}}$. If $O_1(\boldsymbol{x}) > O_2(\boldsymbol{x})$, it is considered that x is classified into class c. Then, $D(c, \overline{c})$ is the data set belonging to G_c, and is classified into $G_{\overline{c}}$. An evaluation function that considers the training accuracy is defined as follows:.

$$F = \frac{|G_c| -| D(c, c)|}{|G_c|} \tag{12}$$

If F is greater than the threshold Th, a classifier is added for accurate discrimination. To train newly added classifier, training data $D(c, \overline{c})$ and $G_{\overline{c}}$ are used. Repeating the addition of classifiers until the evaluation function is less than the threshold Th allows model construction and classifier learning to take place simultaneously.

Through the above training, the model construction and training of the classifier are performed based on a boosting approach.

3. **EMG Pattern Classification.** The proposed EMG pattern classification consists of two parts: (1) EMG feature extraction and (2) classification network.

L channels of EMG signals are recorded using surface electrodes attached to muscles. The EMG signals are measured with a sampling frequency $f = 1000$ Hz, then rectified and filtered by a Butterworth filter (cutoff frequency: 1Hz). Each sampled EMG pattern, defined as $EMG_l(t)(l = 1, 2, \cdots, L)$ was normalised to make the sum of five channels equal to 1 using the following equation,

$$x_l^{(t)} = \frac{EMG_l(t) - EMG_l^{st}}{\sum_{l'=1}^{L}(EMG_{l'}(t) - EMG_{l'}^{st})}, \tag{13}$$

where EMG_l^{st} is the mean value of $EMG_l(t)$ measured while relaxing the muscles. The feature vectors $\boldsymbol{x}(t) = [x_1(t), x_2(t), \cdots, x_L(t)]$ are inputted into classification network. A power level is estimated from the EMG signals as

$$\alpha(t) = \frac{1}{L}\sum_{l=1}^{L}\frac{EMG_l(t) - EMG_l^{st}}{EMG_l^{max} - EMG_l^{st}}, \tag{14}$$

where EMG_l^{max} is the mean value of $EMG_l(t)$ measured under the maximum voluntary contraction. The power level is compared with a prefixed threshold M_d to determine whether the motion actually happened.

For pattern classification, a pattern classification network as described in Section 3 is employed. Using samples labelled with their corresponding motions, the structure of the network is constructed and network learns non-linear mapping between the EMG patterns and motions simultaneously.

The entropy of outputs is also calculated to prevent risk of misclassification. The entropy is defined as Equation 15.

$$H(\boldsymbol{x}(t)) = -\sum_{c=1}^{C} O_c(\boldsymbol{x}(t)) \log O_c(\boldsymbol{x}(t)). \tag{15}$$

If the entropy $H(\boldsymbol{x}(t))$ is less than the discrimination threshold Te, the specific motion with the largest probability is determined according to the Bayes' decision rule (shown in Equation 16). Otherwise, the determination is suspended.

$$Y(\boldsymbol{x}) = \arg\max_{c} O_c(\boldsymbol{x}) \tag{16}$$

4. **Experiments.** Phoneme classification based on EMG signals was conducted to examine performance of the proposed method. In the experiments, EMG signals measured from mimetic and cervical muscles were used to classify six Japanese phonemes ($C = 6$: /a/, /i/, /u/, /e/, /o/, and /n/). Experiments were performed for four subjects (A, B, C: healthy; D: a patient with cervical spine injury). After training, the subjects were asked to utter six phonemes in the order.

Five pairs of Ag/AgCl electrodes (NT-511G: NIHON KOHDEN Corp.) were attached to the subject's face (Depressor Anguli Oris, Zygomaticus Major, Masseter, Digastric, Depressor Labii Inferioris; a pair of electrodes on each muscle) with conductive paste. The EMG signals from five muscles were recorded (sampling frequency: 1KHz).

The parameters of proposed method were set as: $M_d = 0.25$, $Te = 0.7$, $Th = 0.8$ and $Th' = 0.8$.

4.1. **Classification results.** Five sets of randomly chosen initial weights were used to train each sample data. For the verification of the classification performance of the proposed method, single LLGMN, support vector machine (SVM) [25] and back-propagation neural network (BPNN) [26] classifiers were used for the comparison. BPNN had four layers (two hidden layers), the units of which were set as 2, 10, 10 and 4. Also, an SVM having a second-order polynomial kernel was used to perform a two-class classification. By cobining two-class classifiers, multiclass classification using SVMs was achieved.

The mean values and standard deviations of the classification rates using the proposed method and other medhotds are shown in Figure 4. As shown in the figure, the classification results of the proposed method are similar to those of a single LLGMN and SVM in the case of Subjects A and B. Comparing the classification rates of Subject C, it can be seen that the method using the proposed EMG pattern classification method outperformed the use of a single LLGMN and BPNN. In addition, in the case of Subject

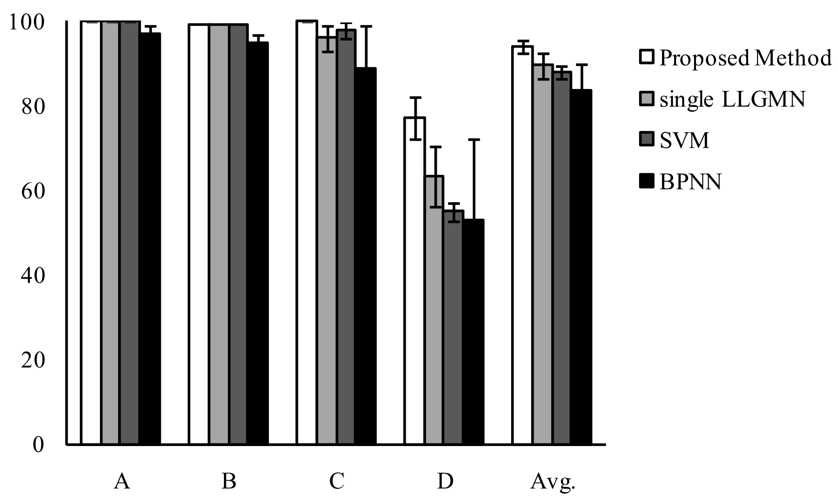

FIGURE 4. Discrimination results for four subjects

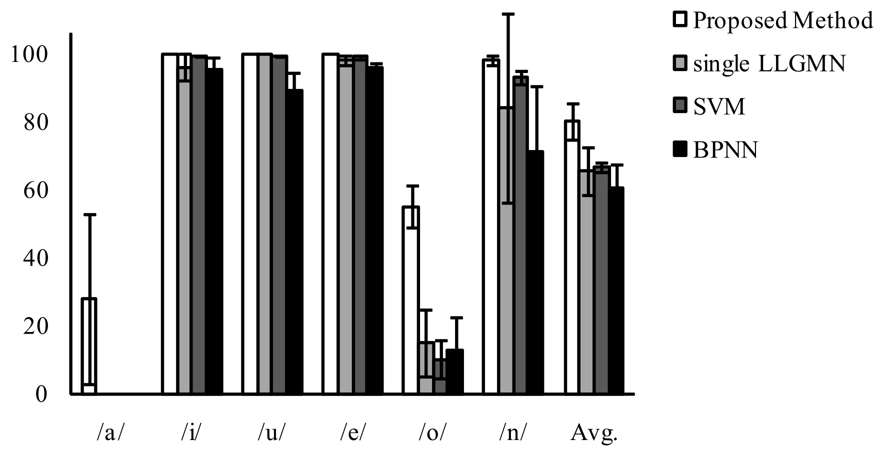

FIGURE 5. Discrimination results for subject D

D, the classification results of other methods degraded significantly more than that using the proposed method.

Figure 5 shows the mean values and standard deviations of the classification rates of each type of utterance for Subject D. From this figure, it can be seen that the classification accuracies of utterance /o/ and /n/ have improved by the proposed method. On the other hand, in the classification of utterance /a/, similar misclassifications occurred by using the proposed method and other methods. It considered that these results were caused by the ambiguous EMG pattern found with the utterance /a/. Table 1 shows an example of the number of added classifiers and LLGMNs in the network of Subject D. In contrast, the number of classifiers and LLGMNs for Subjects A and B were set as 1. Here, we infer that a better classification is achieved by adding the classifiers and LLGMNs for the estimation of the distribution of EMG patterns. It is clear that by adding LLGMNs to a

TABLE 1. The number of added classifiers and LLGMNs for subject D

phoneme	classifiers	LLGMNs
/a/	13	33
/i/	3	7
/u/	2	3
/e/	4	9
/o/	16	24
/n/	5	7

network for the estimation of the distribution of EMG signals, the proposed method can achieve a most accurate classification of all methods.

5. **Conclusions.** This paper proposes a novel classification method for EMG signals. This method uses a hierarchical pattern classification algorithm based on boosting approach for the estimation of a suitable network structure. In this algorithm, the structure of the classification network is automatically constructed by adding LLGMNs as classifiers to estimate the distribution of EMG signals for training.

To examine the classification capability and the accuracy of the proposed method, phoneme classification experiments were carried out with four subjects. In these experiments, the proposed method achieved the highest classification performance of all methods.

In future, we would like to improve the pre-processing method for EMG signals. It would also be interesting to theoretically study the effectiveness of boosting approach for EMG pattern classification.

Acknowledgment. The authors would like to express their gratitude to Mr. Motonobu Shigeto and Mr. Hiromi Koseki for collecting the experimental data.

REFERENCES

[1] M. Okamoto, K. Shima, Y. Matsubara and T. Tsuji, Pattern discrimination method with boosting approach using hierarchical neural trees, *Proc. of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol.222, no.7, pp.701-710, 2008.

[2] O. Fukuda, T. Tsuji, M. Kaneko and M. Ohtsuka, A human-assisting manipulator teleoperated by EMG signals and arm motion, *IEEE Trans. on Robotics and Automation*, vol.19, pp.210-222, 2003.

[3] K. Englehart, B. Hudgins and A. D. C. Chen, Continuous multifunction myoelectric control using pattern recognition, *Technology and Disability*, vol.15, pp.95-103, 2003.

[4] Z. Bien and H.-E. Lee, Effective learning system techniques for human-robot interaction in service environment, *Knowledge-Based Systems*, vol.20, issue.5, pp.439-456, 2007.

[5] T. Kohonen, *Self-Organizing Maps*, Springer, 1995.

[6] H. Benitez-Perez, F. Garcia-Nocetti and H. Thompson, Fault classification based upon self organizing feature maps and dynamic principal component analysis for inertial sensor drift, *International Journal of Innovative Computing, Information and Control*, vol.3, no.2, pp.257-276, 2007.

[7] F. V. Nelwamondo, T. Marwala and U. Mahola, Early classifications of bearing faults using hidden Markov models, Gaussian mixture models, mel-frequency cepstral coefficients and fractals, *International Journal of Innovative Computing, Information and Control*, vol.2, no.6, pp.1281-1299, 2006.

[8] R. Kato, H. Yokoi and A. Tamio, Real-time learning method for adaptable motion-discrimination using surface EMG signal, *Proc. of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.2127-2132, 2006.

[9] F. H. Y. Chan, Y. Yang, F. K. Lam, Y. Zhang and P. A. Paker, Fuzzy EMG classification for prosthetic control, *IEEE Transactions on Rehabilitation Engineering*, vol.8, no.3, pp.305-311, 2000.

[10] I. Kajitani, T. Hoshino, N. Kajihara, M. Iwata and T. Higuchi, An evolvable hardware chip and its application as a multi-function prosthetic hand controller, *Proc. of the 16th National Conference on Artificial Intelligence*, pp.182-187, 1999.
[11] A. Hiraiwa, K. Shimohara and Y. Tokunaga, EMG pattern analysis and classification by neural network, *Proc. of the IEEE Informational Conference on System, Man and Cybernetics*, pp.1113-1115, 1989.
[12] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford Univ. Press, 1995.
[13] D. E. Rumelhart, J. L. McClelland and R. J. Williams, Learning internal representations by error propagation, *Parallel Distributed Processing*, vol.1, pp.318-362, 1986.
[14] T. Tsuji, O. Fukuda, H. Ichinobe and M. Kaneko, A log-linearized Gaussian mixture network and its application to EEG pattern classification, *IEEE Trans. on Systems, Man and Cybernetics-Part C: Applications and Reviews*, vol.29, no.1, pp.60-72, 1999.
[15] T. Tsuji, H. Shigeyoshi, O. Fukuda, H. Ichinobe and M. Kaneko, Bio-mimetic control of an externally powered prothetic forearm based on EMG signals, *Trans. of the Japan Society of Mechanical Engineers*, vol.66, pp.2764-2771, 2000 (in Japanese).
[16] O. Fukuda, T. Tsuji and M. Kaneko, Pattern classification of EEG signals Using a log-linearized Gaussian mixture neural network, *Proc. of IEEE Int. Conf. on Neural Network*, pp.2479-2484, 1995.
[17] O. Fukuda, T. Tsuji and M. Kaneko, An EMG controlled robotic manipulator using neural networks, *Proc. of the IEEE Int. Workshop on Robot and Human Communication*, pp.442-447, 1997.
[18] O. Fukuda, T. Tsuji and M. Kaneko, An EMG-controlled pointing device using a neural network, *Proc. of the 1999 IEEE Int. Conf. on Systems, Man, and Cybernetics*, vol.4, pp.63-68, Tokyo, 1999.
[19] H. Akaike, A new look at the statistical model identification, *IEEE Trans. Automat. Contr.*, vol.19, pp.716-723, 1974.
[20] K. Watanabe and S. Watanabe, Lower bounds of stochastic complexities in variational bayes learning of Gaussian mixture models, *Proc. of the IEEE Conf. on Cybernetics and Intelligent Systems*, pp.99-104, 2004.
[21] R. E. Schapire, The strength of weak learnability, *Machine Learning*, vol.5, no.2, pp.197-227, 1990.
[22] Y. Freund, Boosting a weak learning algorithm by majority, *Information and Computation*, vol.121, no.2, pp.256-285, 1995.
[23] C. Domingo, T. Tsukiji and O. Watanabe, Partial Occam's razor its applications, *Information Processing Letters*, vol.64, pp.179-185, 1997.
[24] L. I. Perlovsky and M. M. McManus, Maximum liklihood neural networks for sensor fusion and adaptive classification, *Neural Networks*, no.4, pp.89-102, 1991.
[25] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*, Cambridge University Press, 2000.
[26] D. E. Rumelhart, J. L. McClelland and R. J. Williams, Learning internal representations by error propagation, *Parallel Distributed Processing*, vol.1, pp.318-362, 1986.