

Training HMM Structure and Parameters with Genetic Algorithm and Harmony Search Algorithm

Kwang-Eun Ko*, Seung-Min Park*, Junheong Park* and Kwee-Bo Sim[†]

Abstract – In this paper, we utilize training strategy of hidden Markov model (HMM) to use in versatile issues such as classification of time-series sequential data such as electric transient disturbance problem in power system. For this, an automatic means of optimizing HMMs would be highly desirable, but it raises important issues: model interpretation and complexity control. With this in mind, we explore the possibility of using genetic algorithm (GA) and harmony search (HS) algorithm for optimizing the HMM. GA is flexible to allow incorporating other methods, such as Baum-Welch, within their cycle. Furthermore, operators that alter the structure of HMMs can be designed to simple structures. HS algorithm with parameter-setting free technique is proper for optimizing the parameters of HMM. HS algorithm is flexible so as to allow the elimination of requiring tedious parameter assigning efforts. In this paper, a sequential data analysis simulation is illustrated, and the optimized-HMMs are evaluated. The optimized HMM was capable of classifying a sequential data set for testing compared with the normal HMM.

Keywords: Hidden markov model, Genetic algorithm, Harmony search algorithm, Optimization

1. Introduction

During the past decades, studies on electrical power quality have become an important issue in the electric power system (EPS) industry. The main cause of the power quality issues is that a number of microelectronic devices are very sensitive to subtle changes such as a disturbance-induced variation of the voltage amplitude, frequency and so on [1, 2]. The increase of these devices induces the need for measuring and assessing the effect of power transient factors such as disturbances, stability, and so on. Especially, the electric power transient disturbances associated with degradation of power quality are non-stationary and transient factors that are expressed by nature and points of sharp variation such as singularities in transient signals usually carry the most important information about the disturbance. The transient disturbances on power system include voltage sag, voltage swell, capacitor switching transients, notching, noise, interruptions and harmonics [15]. There is a need for classification of the type of disturbances in order to maintenance of the quality of power system. The electric transient disturbance types are classified by the application of machine learning and pattern recognition algorithms such as neural network, fuzzy inference system and support vector machine [3, 16]. Recent literature reports several studies that electric transient disturbance classification techniques are based on

the combination of wavelet transform and machine learning algorithms [1, 2, 5, 15]. For example, the neural network based method is limited by the performance of the neural network itself. Thus, the correct classification rate may not be especially high. The fuzzy adaptive resonance theory and least square support vector machine methods are used to solve the electric transient stability problem as shown in [5] and [6]. Although they presented a good evolution, the experimental results are not practical.

Among a number of these machine learning algorithms, hidden Markov Models (HMM) are receiving scientific attention as one of the most successful techniques for classifying the wavelet-based electric transient disturbances [16]. Although the HMMs were originally applied to speech recognition, they have been proved highly successful for modeling the other types of data sequences such as electrical harmonics and types of transient disturbance [2, 16]. However, the success of HMMs owes much to their ability to encode target data sequence in their structure. This means that the performance of HMMs rely on the result of structure learning. Also, HMMs allow many unknown and missing quantities to be trained through the proper parameter learning process. Therefore, an automatic means of optimizing the structure and parameter learning process of HMMs would be highly desirable, but it raises important issues: model interpretation and complexity control. Won suggested a training method for HMM structure with genetic algorithm. Chau shows that using GA for HMM parameter learning has a better performance than un-optimized HMM case [14, 17]. Also, other heuristic algorithm, such as particle swarm optimization, is proposed

[†] Corresponding author: School of Electrical and Electronics Engineering, Chung-Ang University, Seoul, Korea (kbsim@cau.ac.kr)

* School of Electrical and Electronics Engineering, Chung-Ang University, Seoul, Korea (kkeun@cau.ac.kr, sminpark@cau.ac.kr, antiline@cau.ac.kr)

to optimize the training process of HMM by Yang [18]. Inspired by the previous approaches, we present an optimization strategy for the HMM structure and parameter training using GA and harmony search (HS) algorithm. The optimized HMM can be used for not only electric transient disturbance problem, but also versatile problems.

In this paper, we investigate the use of GA and HS algorithm as the optimization strategy for the structure and parameter learning of HMM. We begin by giving a brief overview of related works of GA, HS algorithm, HMM and introducing our notation. We then present the proposed method for solving the optimization problem associated with the HMM structure and parameter learning. And, we present some experimental results produced by the proposed method.

2. Overview of Related Works

2.1 Hidden Markov Model

HMMs are mainly used for modeling time varying data sequences. It has been widely used in speech recognition. Also, HMMs are a simple, dynamic Bayesian network forms [9, 10]. If we denote a sequence of observations \mathbf{v}_t and a sequence of hidden states s_t ($t = 1, 2, \dots, T$), then HMMs have to solve the joint probability distribution of states and observations that can be factorized as follows:

$$P(s_1, \dots, s_T, y_1, \dots, y_T) = P(s_1)P(y_1 | s_1) \prod_{t=2}^T P(s_t | s_{t-1})P(y_t | s_t) \quad (1)$$

For the real-valued case of observation vectors, $P(y_t | s_t)$ can be modeled in many different forms such as a Gaussian distribution. The training process for HMM structure and parameters is composed of the inference (expectation) step and learning (maximization) step. That famous forward-backward algorithm and Baum-Welch algorithm allow us to train the structure and parameter of HMMs. Details of the HMM are described in [10].

2.2 Genetic algorithm

The GA represents an advanced numerical search and heuristic optimization method inspired by the biological theory of evolution. In general, GA is initialized with little knowledge about the given matter to be solved and a searching process is performed in parallel for a complex and vast search space.

In general, GA consists of populations, chromosome, fitness function and genetic operations. The population represents a set of proper solutions. And each individual in the population represents a potential solution to a specified object problem. The search space for the problem solution

is defined in this population representation. Each of the variables that compose an individual is known as chromosome. The chromosomes are commonly coded into a string to form the individual. Each individual in the population is evaluated by a fitness function in order to determine how fit is the solution. The GA maintains a population of n possible solutions, i.e., individuals, with associated fitness values evaluated according to the fitness function. Details of the GA are described in [8].

2.3 Harmony Search Algorithm with Parameter-Setting Free Technique

The harmony search (HS) algorithm is an emerging meta-heuristic optimization method that has been successfully applied to diverse scientific and engineering problems. Other meta-heuristic algorithms, such as GA, simulated annealing and particle swarm optimization, have a major disadvantage in that they require a considerable technique for setting of the initial model parameters. In order to solve this problem, Geem proposed the HS algorithm with respect to the parameter-setting-free technique [13]. The basic HS algorithm tries to find an optimal solution vectors that satisfies an objective function related with given problems. The algorithm starts with randomly generated vectors that are stored at the harmony memory (**HM**) as follows:

$$\mathbf{HM} = \begin{bmatrix} \hat{y}_1^1 & \hat{y}_2^1 & \dots & \hat{y}_n^1 & | & \mu(\hat{\mathbf{y}}_1) \\ \hat{y}_1^2 & \hat{y}_2^2 & \dots & \hat{y}_n^2 & | & \mu(\hat{\mathbf{y}}_2) \\ \vdots & \vdots & \dots & \vdots & | & \vdots \\ \hat{y}_1^{hms} & \hat{y}_2^{hms} & \dots & \hat{y}_n^{hms} & | & \mu(\hat{\mathbf{y}}_{hms}) \end{bmatrix} \quad (2)$$

where, n =dimension of the decision vectors, $\mu(\hat{\mathbf{y}}_i)$ =objective function and hms =the size of harmony memory.

A $\hat{\mathbf{y}}^{New}$ is updated by the heuristic operations of HS algorithm. Details are described in the Table below [9, 13].

Table 1. HS operations of HS algorithm

Operation	Mechanism	Selection criteria
Random playing	Randomly chosen $\hat{\mathbf{y}}$ if $\hat{\mathbf{y}}^{\min} \leq \hat{\mathbf{y}} \leq \hat{\mathbf{y}}^{\max}$, then $\hat{\mathbf{y}}^{new} \leftarrow \hat{\mathbf{y}}$	1 - <i>HMCR</i>
Memory consideration	Chosen $\hat{\mathbf{y}}$ from HM $\hat{\mathbf{y}}^{new} \leftarrow \hat{\mathbf{y}} \in \{\hat{\mathbf{y}}^1, \hat{\mathbf{y}}^2, \dots, \hat{\mathbf{y}}^{hms}\}$	<i>HMCR</i>
Pitch adjustment	Chosen $\hat{\mathbf{y}}$ from HM $\hat{\mathbf{y}}^{NEW} \leftarrow \begin{cases} \hat{\mathbf{y}} + \Delta \\ \hat{\mathbf{y}} - \Delta \\ \hat{\mathbf{y}} \end{cases}$	$HMCR \times PAR \times 0.5$ $HMCR \times PAR \times 0.5$ $HMCR \times (1 - PAR)$

(*HMCR*: harmony memory considering rate, *PAR*: pitch adjustment rate, Δ : amount of increment)

Also, Fig. 2 shows the basic scheme of the HSA routine based on the above three operations.

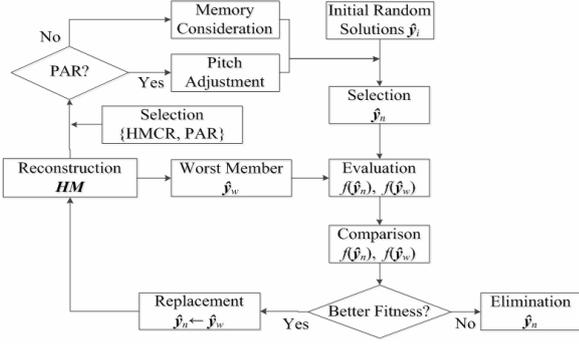


Fig. 1. The basic scheme of HS algorithm routine

3. HMM based Sequential Data Classification

3.1 GAs for optimizing HMM structure

To prove that GA is potentially useful for evolving HMMS, we implemented a standard GA process where a population of HMMS is evolved. The genetic operations are composed of mutation and crossover. For the mutation to be useful, it should make changes that cause minimal disruption so that the new HMM has a high probability of having fitness close to that of the un-mutated HMM. The crossover takes place between two HMMS and exchanges states based on probability. At each generation, some proportion of the HMMS is trained with typical Baum-Welch on a training set. The fitness is measured via an evaluation set and the fitter HMMS are selected. Finally, the mutation and crossover procedure is performed by the selected members in order to form the next generation. Fig. 2 shows the optimization process for HMM structure using GA-based strategy.

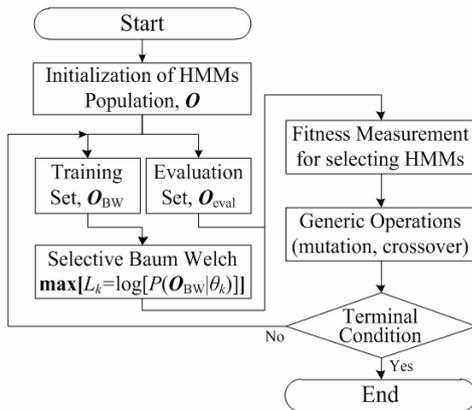


Fig. 2. The GA-based HMM structure optimization

3.2 HSA for optimizing HMM structure

The sequential data set can be classified by the decoding process of HMM. In order to more accurately classify the hidden state, an automatic means of optimizing the HMM

parameters would be highly desired. In this paper, HSA was used for optimizing the HMM parameters.

The model parameters of HMM mainly consisted of two matrices, \mathbf{A} and \mathbf{B} . \mathbf{A} is the $n \times n$ state-transition matrix including the probability distribution of state displacement. And \mathbf{B} is $n \times m$ observation symbol matrix including k observation probability distribution a given i -th state.

\mathbf{A} and \mathbf{B} can be formulized as follows:

$$\mathbf{A} = \{a_{ij} = P[s_j | s_i]\}, \sum_{j=1}^n a_{ij} = 1, 1 \leq i, j \leq n \quad (3)$$

$$\mathbf{B} = \{b_{jk} = P[o_t = v_k | s_j]\}, \sum_{k=1}^m b_{jk} = 1, 1 \leq k \leq m, 1 \leq j \leq n \quad (4)$$

The training of HMM parameter is firstly performed by the forward-backward algorithm. The following equations define the forward-backward variables $\{\alpha_t(i), \beta_t(i)\}$ and the probability expectation values $\{\gamma_t(i), \xi_t(i, j)\}$ [10].

$$\alpha_t(j) = \left[\sum_{j=1}^N \alpha_{t-1}(j) a_{ij} \right] b_j(o_t) \quad (5)$$

$$\beta_t(j) = \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \beta_{t+1}(j) \quad (6)$$

$$\gamma_t(i) = \frac{\alpha_t(i) \beta_t(i)}{\sum_{i=1}^N \alpha_t(i)} \quad (7)$$

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(i)}{\sum_{i,j=1}^N \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(i)} \quad (8)$$

In this case, the initial parameter setting of HMM can be defined as the following equations.

$$[a_{ij}] = \arg \max_{a_{ij}} \left\langle \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \right\rangle \quad (9)$$

$$[b_{jk}] = \arg \max_{b_j(o_t)} \left\langle \frac{\sum_{t=1}^{T-1} \gamma_t(j) b_j(o_{t+1})}{\sum_{t=1}^T \gamma_t(j)} \right\rangle \quad (10)$$

The HMM parameters have to be encoded into a harmony form that have real-valued probabilities. Fig. 3 shows the string \mathbf{h}_t that acts as a harmony and it consists of two main parts: \mathbf{A} and \mathbf{B} respectively.

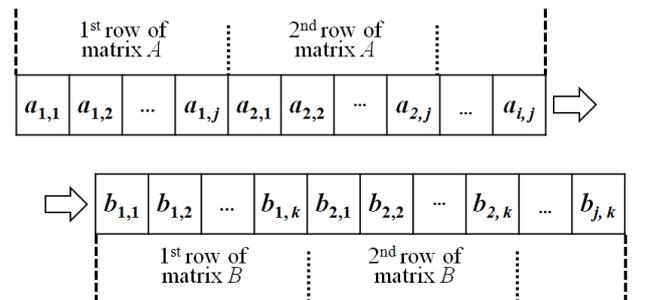


Fig. 3. Harmony representation of HMM parameters

The sum of Eq. (9) and Eq. (10) is used as an objective function for HS algorithm-based optimization process of HMM parameter.

$$f(\hat{\mathbf{y}}) = \sum_{i=1}^{T-1} \xi_i(i, j) / \sum_{i=1}^{T-1} \gamma_i(i) + \sum_{i=1}^{T-1} \gamma_i(j) / \sum_{i=1}^T \gamma_i(j) \quad (11)$$

As much as the harmony memory size (H), the initial solution vectors, i.e. initial harmonies $\hat{\mathbf{y}}_1 = [\mathbf{a}^1, \mathbf{b}^1]$, are generated by selective Baum-Welch algorithm according to the Eq. (9), (10) and (11).

$$\mathbf{HM}_{HMM} = \begin{bmatrix} a_{1,1}^1 \cdots a_{1,n}^1 \cdots a_{n,n}^1 & b_{1,1}^1 \cdots b_{1,k}^1 \cdots b_{n,k}^1 & f(\hat{\mathbf{y}}_1) \\ a_{1,1}^2 \cdots a_{1,n}^2 \cdots a_{n,n}^2 & b_{1,1}^2 \cdots b_{1,k}^2 \cdots b_{n,k}^2 & f(\hat{\mathbf{y}}_2) \\ \vdots & \vdots & \vdots \\ a_{1,1}^H \cdots a_{1,n}^H \cdots a_{n,n}^H & b_{1,1}^H \cdots b_{1,k}^H \cdots b_{n,k}^H & f(\hat{\mathbf{y}}_H) \end{bmatrix} \quad (12)$$

New harmony, $\hat{\mathbf{y}}^n = [\mathbf{a}^n, \mathbf{b}^n] = [a_{1,1}^n, \dots, a_{i,j}^n, b_1^n, \dots, b_{j,k}^n]$ is improved by the one of the HS operations: random playing, memory consideration, and pitch-adjustment. Once the new harmony is obtained, the existing harmonies of \mathbf{HM} are further checked whether it satisfied the worst condition. If the new harmonies are better than the worst one in \mathbf{HM} in terms of $f(\hat{\mathbf{y}}_i)$, then the worst one is replaced by a new one until the termination condition are satisfied.

By using HS algorithm, we can describe the optimization process for HMM parameters as shown in Fig. 4.

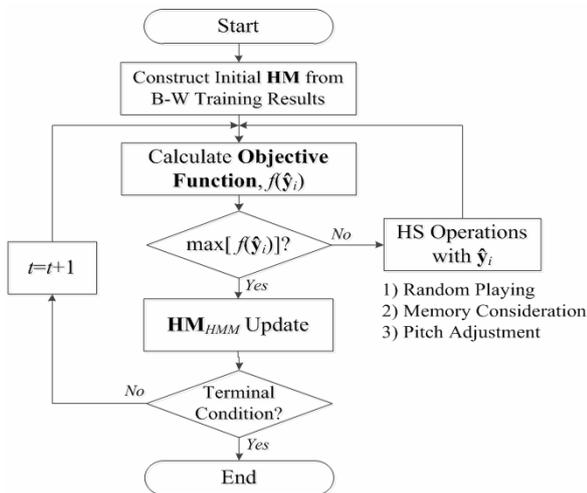


Fig. 4. The HSA based HMM parameter optimization

The initial parameter set consisting of HMMs was not considered by HS algorithm, because of the new parameter-setting-free technique that contains three operations (random selection, memory consideration, pitch adjustment) for every variable in harmony memory (\mathbf{HM}). Also, HSA has the advantage of being free from divergence of solutions and local optima [13].

4. Experimental Results

The experimental results of HMM optimization are presented in this section. To discover if the optimization strategies are useful for improving the performance of HMM, we have to implement a standard HMM, i.e., un-optimized HMM. In this experiment, we selected the speech signals classification problem, where the speech signals are composed of ten number words (0, 1, ..., 9). Thus, the configuration of HMM used ten states left-right model and a set of 60 samples (observation sequences). These configurations have the desirable property of being able to readily model sequence signal change over time. The initial setting of model parameters (\mathbf{A} , \mathbf{B}) is randomly generated by a uniform distribution on the interval [0, 1] and 10 experiments are conducted.

In this experiment, the HMM structure and parameters are trained by GA and HS algorithm with the same 10 observation training sequences. In GA-HMM structure training, the following conditions are used: population size=30, crossover rate = 0.01, and mutation rate = 0.0001. In HS-HMM parameter training, the following conditions are used: harmony memory size=10, harmony memory consideration rate = 0.8, and pitch adjustment rate = 0.4.

The performance measurement index of the proposed method HMM consists of two values: p_{corr} and p_{wrong} , which are average log accuracy of the HMM generated by 60 training observation sequences of correct and wrong classified [12]. These parameters can indicate the standard performance of HMM. The optimized HMM has mostly higher average log probabilities at p_{corr} and lower average log probabilities at p_{wrong} than the un-optimized HMM as shown in Table 2.

Table 2. A comparison between the optimized HMM with HS algorithm and un-optimized HMM in terms of average log accuracy

Experiments #	Sample Types	Optimized HMM with HS algorithm		Un-optimized HMM	
		p_{corr}	p_{wrong}	p_{corr}	p_{wrong}
1	①-②	-1.9268	-8.0732	-4.7359	-7.2714
2	①-③	-1.6605	-8.3395	-4.2125	-8.6137
3	①-②	-1.4182	-8.5818	-4.9843	-7.5914
4	②-①	-1.9484	-8.0516	-4.3908	-7.7634
5	①-③	-1.9326	-8.0674	-4.3876	-7.1007
6	②-①	-2.0621	-8.0917	-4.9811	-7.3825
7	①-②	-1.9083	-8.0917	-4.3481	-7.7351
8	①-③	-1.7257	-8.2743	-4.0567	-7.9328
9	③-①	-1.7410	-8.2590	-4.4860	-7.7514
10	②-①	-2.5436	-7.4564	-4.9251	-8.2254

The type of samples used in experiment is as follows. (① left hand, ② right hand, ③ foots)

5. Conclusion

In this paper, we propose a method for optimizing HMM in order to analyze sequential data. Generic HMMs play an important role in handling sequential data sets such as speech, biological signals, and electric transient. Therefore, this paper presented an optimization strategy for HMM using GA and HS algorithm. The optimization strategy is composed of two major procedures: structure optimization and parameter optimization. The structure optimization of HMM is performed by GA-based method. And the parameter optimization of HMM is performed by HS algorithm. The experimental results of the proposed method indicated that the optimized HMM has a higher probability in classifying the global maxima or at least local-maxima with better performance to generate training observation sequences than the un-optimized HMM. Thus, it can be optimized in a wide range of applications including the systems that cannot be solved by calculation-based algorithms.

Acknowledgements

This work was supported by the National Research Foundation of KOREA [NRF] grant funded by the KOREA government [MEST] [No.2011-0029861].

References

- [1] A. Santoso, E. J. Powers and W. M. Grady, "Power Quality Disturbance Identification using Wavelet Transform and Artificial Neural Network," *Proc. Int. Conf. Harmonics and Quality of Power*, pp.615-618, 1996
- [2] J. Chung, E. J. Powers, W. M. Grady, and S. C. Bhatt, "Electric Power Transient Disturbance Classification using Wavelet-based Hidden Markov Models," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 6, pp.3662-3665, 2000
- [3] W. Min, A. V. Mamishev, "Classification of Power Quality Events using Optimal Time-Frequency Representations-Part 1: Theory," *IEEE Transactions on Power Deliver*, vol. 19, Issue 3, pp.1488-1495, 2004
- [4] M. Negnevitsky, M. Ringrose, "A Neuro-Fuzzy System for Recognition of Power Quality Disturbances," *IEEE Power Engineering Society General Meeting*, vol. 3, pp.2295-2300, 2005
- [5] M. C. G. Silveira, A. D. P. Lotufo, C. R. Minussi, "Transient Stability Analysis of Electrical Power Systems using a Neural Network based on Fuzzy ARTMAP," *Proc. IEEE Bologna PowerTech Conference*, 2003
- [6] A.W. Noor Izzri, A. Mohamed, I. Yahya, "A New Method of Transient Stability Assessment in Power Systems using LS-SVM," *The 5th Student Conference on Research and Development*, 2007
- [7] M. Azab, "Harmonic Elimination in Three-Phase Voltage Source Inverters by Particle Swarm Optimization," *Journal of Electrical Engineering & Technology*, vol. 6, no. 3, pp. 334-341, 2011
- [8] D. Whitley, "A Genetic Algorithm Tutorial", *Statist. Comput.*, vol. 4, pp. 65-85, 1994
- [9] Z. W. Geem, "State-of-the-Art in the Structure of Harmony Search Algorithm," *Recent Advances in Harmony Search Algorithm*, Studies in Computational Intelligence, vol. 270, pp. 1-10, 2010
- [10] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of IEEE*, Vol. 77, pp. 257-286, 1989
- [11] H. K. Lee, S. J. Choi, "PCA+HMM+SVM for EEG pattern classification," *Proceedings of seventh international symposium on signal processing and its applications*, vol. 1, pp. 541-544, 2003
- [12] C. W. Chau, S. Kwong, C. K. Diu, W. R., Fahrner, "Optimization of HMM by a Genetic Algorithm," *IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1727-1730, 1997
- [13] Z. W. Geem and K. B. Sim, "Parameter-Setting-Free Harmony Search Algorithm," *Applied Mathematics and Computation*, vol. 217, no. 8, pp.3881-3889, 2010
- [14] K. -J. Won, A. Pruger-Bennett and A. Krogh, "Training HMM Structure with Genetic Algorithm for Biological Sequence Analysis," *Bioinformatic*, vol. 20, no. 18, pp. 3613-3619, 2004
- [15] D. T. Nguyen and T. A. Hoang, "Analysis of Power Transient Disturbances using Wavelet Transform Modulus Maxima Technique," *Proceedings of the Australasian Universities Power Engineering Conference*, Brisbane, pp. 190-195, 2000
- [16] N. Perera, A. D. Rajapakse, "Power System Transient Classification for Protection Relaying," *The 13th Int. Conf. Harmonics and Quality of Power*, pp. 1-6, 2008
- [17] C. W. Chau, S. Kwong, C. K. Diu and W. R. Fahrner, "Optimization of HMM by a Genetic Algorithm," *Proc. ICASSP*, pp.1727-1730, 1997
- [18] F. Yang and C. Zhang, "An Effective Hybrid Optimization Algorithm for HMM," *Proc. Fourth International Conference on Natural Computation*, pp.80-84, 2008



Kwang-Eun Ko He received the B.S. and M.S. degrees from the Department of Electrical and Electronics Engineering, Chung-Ang University, Seoul, Korea, in 2007 and 2009 respectively. He is currently a candidate for the Ph.D. degree in the School of Electrical and Electronics Engineering at Chung-Ang University. His research interests include Machine Learning, Emotion Recognition, and Multimodal Intention Recognition, etc.



Seung-Min Park He received the B.S. degree from the Department of Electrical and Electronics Engineering, Chung-Ang University, Seoul, Korea, in 2010. He is currently Master course in the School of Electrical and Electronics Engineering at Chung-Ang University. His research interests include Pattern Recognition, Brain Computer Interface, Evolutionary Computation, etc.



Junheong Park He received the B.S. degree from the Department of Electrical and Electronics Engineering, Chung-Ang University, Seoul, Korea, in 2011. He is currently Master course in the School of Electrical and Electronics Engineering at Chung-Ang University. His research interests include Machine Learning, Brain-Computer Interface, Gesture Recognition, etc.



Kwee-Bo Sim He received the B.S. and M.S. degrees in Department of Electronic Engineering from Chung-Ang University, Seoul, Korea, in 1984 and 1986 respectively, and Ph. D. degree in Department of Electronic Engineering from the University of Tokyo, Japan, in 1990. Since 1991, he has been a faculty member of the School of Electrical and Electronic Engineering at the Chung-Ang University, where he is currently a Professor. His research interests are Artificial Life, Neuro-Fuzzy and Soft Computing, Evolutionary Computation, Learning and Adaptation Algorithm, Autonomous Decentralized System, Intelligent Control and Robot System, Artificial Immune System, Evolvable Hardware, and Artificial Brain etc.