TELKOMNIKA, Vol.10, No.1, March 2012, pp. 129~136
ISSN: 1693-6930
accredited by DGHE (DIKTI), Decree No: 51/Dikti/Kep/2010

Isolated Word Recognition Using Ergodic Hidden Markov Models and Genetic Algorithm

Nyoman Rizkha Emillia, Suyanto, Warih Maharani
Faculty of Informatics, Telkom Institute of Technology
Jl. Telekomunikasi No.1, Ters. Buah Batu, Bandung 40257 Indonesia
Phone: +62-22-7564108, Fax: +62-22-7562721
e-mail: mieltheaatuh@gmail.com, suy@ittelkom.ac.id, wrh@ittelkom.ac.id

Abstrak

Penerjemah ucapan ke teks merupakan salah satu aplikasi pengenalan ucapan dimana sinyal ucapan diolah, dikenali, dan diubah ke dalam representasi tekstual. Hidden Markov model (HMM) merupakan metode yang banyak digunakan dalam proses pengenalan ucapan, namun tingkat akurasinya sangat dipengaruhi oleh optimalisasi proses ekstraksi ciri dan pemodelan yang digunakan. Pada penelitian ini digunakan metode algoritma genetik (GA) untuk mengoptimalkan kinerja Ergodic HMM pada sistem pengenalan ucapan. Pada sistem hibrid HMM-GA, GA digunakan untuk mengoptimalkan metoda Baum-Welch pada proses pelatihan sistem untuk meningkatkan performansi dari parameter-parameter Ergodic HMM yang menghasilkan akurasi yang relatif rendah pada saat pengujian sistem HMM. Berdasarkan penelitian ini didapatkan prosentase kenaikan tingkat akurasi sebesar 20% sampai 41%. Terbukti bahwa kombinasi GA dengan HMM dapat memberikan hasil yang lebih optimal, jika dibandingkan dengan sistem HMM yang hanya menggunakan metode Baum-Welch.

Kata kunci: *algoritma genetik, Baum-Welch, Ergodic HMM, hibrid HMM-GA, Hidden Markov model*

Abstract

Speech to text was one of speech recognition applications which speech signal was processed, recognized and converted into a textual representation. Hidden Markov model (HMM) was the widely used method in speech recognition. However, the level of accuracy using HMM was strongly influenced by the optimalization of extraction process and modellling methods. Hence in this research, the use of genetic algorithm (GA) method to optimize the Ergodic HMM was tested. In Hybrid HMM-GA, GA was used to optimize the Baum-Welch method in the training process. It was useful to improve the accuracy of the recognition result which is produced by the HMM parameters that generate the low accuracy when the HMM are tested. Based on the research, the percentage increases the level of accuracy of 20% to 41%. Proved that the combination of GA in HMM method can gives more optimal results when compared with the HMM system that not combine with any method.

Keywords*: Baum-Welc h, Ergodic HMM, genetic algorithm, Hidden Markov model, hybrid HMM-GA*

1. Introduction

There are several problems which arise in the Automatic Speech Recognition (ASR) specially speech to text system, such as the speaker adaption and how to make an effective language model [1]-[7]. Several methods can be used to modeling the speech signal, such as Hidden Markov Model (HMM) [1],[4],[8],[9]. HMM is the best method for modeling the speech signal because this method is represented in the form of states which the characteristics of the speech signal is also represented in the form of states [10]. Left- right model of HMM models commonly used for modeling the isolated words [11],[12], but in this research, we used the ergodic model to modeling speech signal. Ergodic HMM can directly model the sequence of verbal units (phonemes, words) articulated more than left-to-rigth model that have been used in [11]. During articulation, only certain verbal units may follow each other, but one may progress from one verbal unit if enough intermediate states are allowed [9].

Genetic Algorithms (GA) is one of the optimization methods that commonly used to optimize any methods; in this case, GA can improve the optimization of the Ergodic HMM model, which can affect accuracy of the recognition result [9],[11]. In recent years, study on

Received May 9th, 2010; Revised August 3rd, 2010; Accepted August 16th, 2010

Genetic Algorithm can be found in many research papers [6], [13]. They demonstrated different characteristics in Genetic Algorithm than others.

The main problem in this case is speaker adaptation and modeling isolated words using ergodic model in structure of HMM models [14]. The number of data used in the training process greatly affects the results of the training process. For example, if we have much variation data for the training process, the higher level of accuracy is given. Meanwhile, if the training data is less variation of the data, the very low level of accuracy will be given. Based on that statement the use of GA as a hybrid of HMM is expected to improve the accuration which is given if the problem is lack of training data provided.

The isolated words that we used are in Bahasa, Indonesian language, those are “satu”, “dua”,”tiga”, “empat”, “lima”, “enam”, “tujuh”, “delapan”, “Sembilan”, “nol” and “kosong”. In this ergodic HMM model, the words “nol” and “kosong” assumed have the same meaning that is zero.

2. Research Method

In this paper, genetic algorithm was applied to optimize the Baum-Welch algorithm in Ergodic HMM. The result between HMM system and hybrid HMM-GA was compared to analyze how GA can improve the accuracy in hybrid HMM-GA.

2.1. Hidden Markov Model (HMM)

Hidden Markov Model (HMM) is a natural and highly robust statistical method for automatic speech recognition [10],[14],[15]. Hidden Markov Model is a doubly embedded stochastic process. Stochastic process is a process in which observation is made indirectly, which can only be observed through another stochastic process that produces a row of observation [10],[14]. Modeling speech signals have a HMM parameter estimation process which is repeated and eventually obtained a set of parameters. Recognition process is done by looking for words that provide the greatest chance of occurrence of these observations in order HMM model.

HMM consist of basic elements ie, N (number of states in HMM) and M (Number of distinct observation symbols in each state) [10]. Symbols of individual observations from each state that would establish a series of observation can be expressed as $V = (v_1\ v_2 \ldots\ v_m)$, symbol of the probability of trasition from state i to state j is A, B is a symbol of probability for observation in a state, π is symbol of probability for initial state. HMM can be represented with a notation λ ($\lambda = (A, B, \pi)$) [10].

2.2. Genetic Algorithm

Genetic algorithm is a method that used to solve optimization problems based on natural selection, which is a process that refers to mechanism of biological evolution [16],[17]. Individuals who are fitter in a generation will have the ability to survive and reproduce better than individuals who are less fit. The process of biological evolution is expected at a certain generation, the population will contain more fit organisms. This method has the same basic concept with the process of biological evolution above [16]. GA modifies the population of individual solutions repeatedly. At each step the GA choose the best individuals who would be children on the parents to produce the next generation. GA uses three basic rules for forming the next generation of the current generation, ie Selection, crossover, mutation [16],[11],[12].

2.3. Hybrid HMM-GA

In every generation there is a section of train the HMM training set. HMM fitness values measured a set of evaluation and choosing the fitest chromosome. Then the chromosomes will be in to modify process using the genetic operators of GA before going to the next generation evolution process.

HMM topology that we used is ergodic HMM. Broadly speaking, the process of forming a model for each reference signal comprising a random population initialization, Baum-Welch reestimation, calculation of fitness value, genetic operation, and checking the condition of termination. Here is an example of ergodic topology is used (states=4):

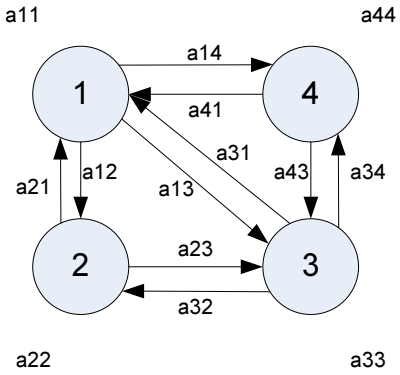

Figure 1. Ergodic HMM with States=4 [10]

The process of forming a model $\lambda = (A, B, \pi)$ for reference signals using Hybrid HMM-GA method can be drawn from the block diagram below:

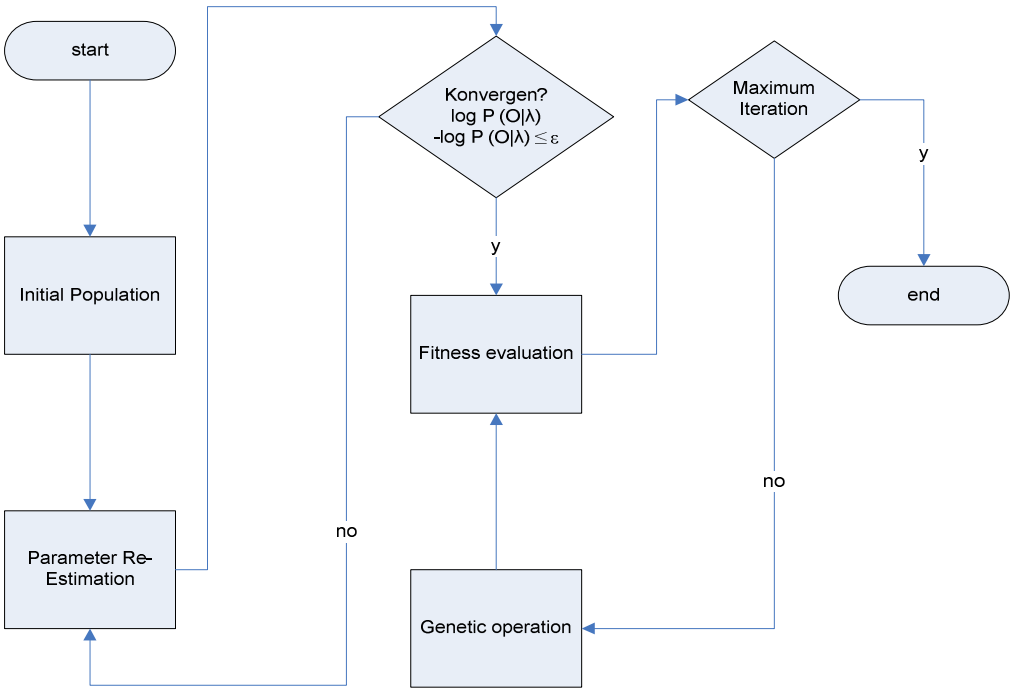

Figure 2. Block diagram of Hybrid HMM-GA

First of all randomly generated initial population, these populations include individuals who are HMM parameters namely, A, B, π. HMM model that we used is ergodic HMM model [9]. This is because the pronunciation of the word “nol” and “kosong” is not modeled separately, but used as one model [9],[11],[12]. Then do the process of Baum-Welch reestimate that are refered to as the training process due to the reestimates, each initial model will be adapted again and again until the difference in the logarithmic [11],[12]. The model results with prior reestimation smaller than a certain threshold value. In other words, the change of the log before the next logarithmic value, no significant change (often called convergent). In this research we used in the calculation of scaling parameters in the HMM model in order to obtain a new algorithm to calculate the HMM parameters.

The next step is to calculate the fitness value of each model is formed. Mathematically fitness function of the n models is expressed as:

$$f_n = \frac{P_n}{\sum_{i=1}^{N} P_i} \quad (1)$$

$$P_n = \frac{\sum_{i=1}^{M} \log(P(o_i \mid \lambda_n))}{M} \quad (2)$$

where:

P_n Is the average of probability the generated models

M is stating the number of rows of observations

N is stating the number of models in a population

The third step is each model in a population that has been computed its fitness value will then be carried out genetic surgery. Genetic operations may include selection, crossover, and mutation by firstly encode each model in a population into chromosomes [14]. Genetic operation is performed to improve the re-estimation technique Baum-Welch so that the population genetics of this operation is an optimal model [11],[12].

After all of the training process is completed, then is formed reference model which is used as a reference model in th recognition process. Recognition process carried out by finding a model that has the greatest probability of all existing models. This is done by performing forward procedure to the input speech observation sequence based on the parameters of HMM of each model of words. The model of words that generates the greatest probability of the input speech is a candidate for the result of recognition.

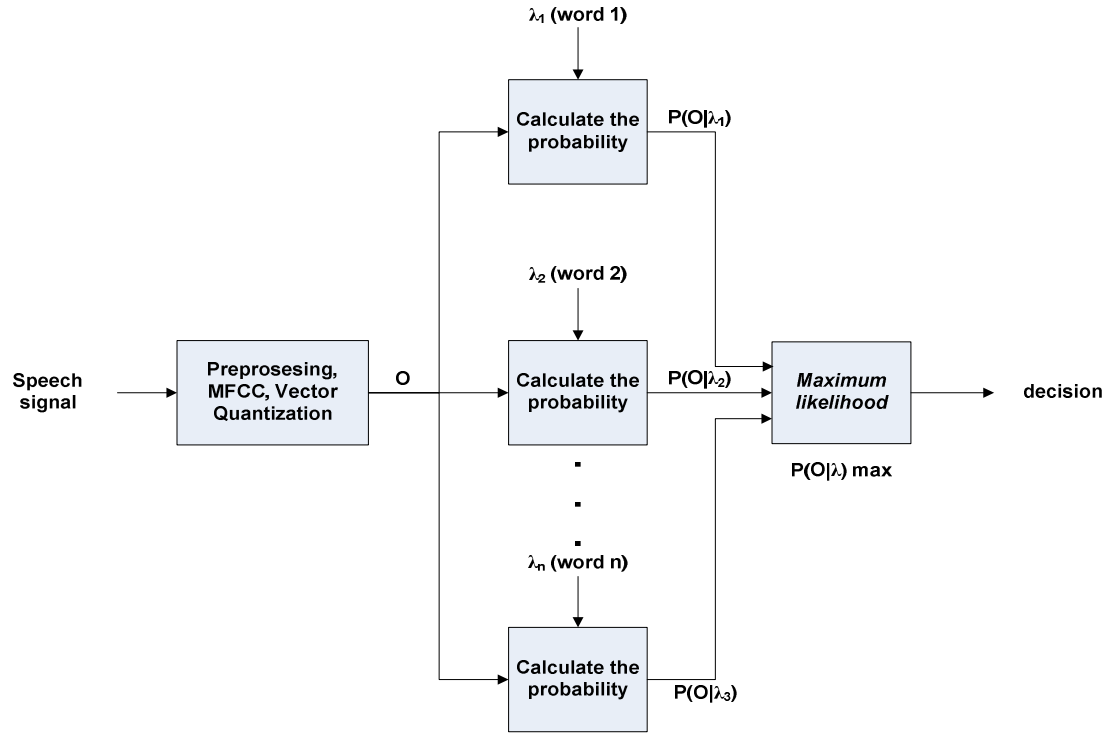

Figure 3. Block diagram of the recognition process [9]

The parameters that was used in testing process, as follows:

- Number of states that are used: 5,6,7,8,9,10, and 20.
- The size of codebook used: 16, 32, 64, and 128.
- HMM topology used: ergodic HMM.

These parameters were chosen, because these parameters can give varying result [5],[9],[11],[12]. The purpose of testing is to find the worst HMM parameters and test it using HMM-GA system. The accuracy was measured in each group of words is to show the performance of the system.

For the recognition process using the hybrid HMM-GA, the training process using the following parameters:

a. States: states-states that produce the worst accuracy in the HMM
b. The size of codebook: the number of clusters that produce the worst accuracy in the HMM
c. This is done to see how big the influence of GA on HMM system.

The GA parameters that used are:

a. Maximum number of individual : 10000 and 1000
b. Crossover probability: 0.9
c. Probability of Mutation: 1/total gene

GA will be analyzed how the influence of changes in number of individuals in a population on the accuracy of the system.

3. Results and Discussion

To find the HMM parameters which are the worst result will be tested against HMM system based on the combined size of the codebook with the amount of state. In this case, the amount of state for each test permanently we used the state 5,6,7,8,9,10,20. Tests conducted on a sound file that consists of 4 speakers in which 2 women and 2 men. The properties of the sound file that is, channel mono, 16000 Hz sampling frequency, bit resolution of 16 bits. Tests were done separately between male and female speakers. The following graph test results for female and male speakers:

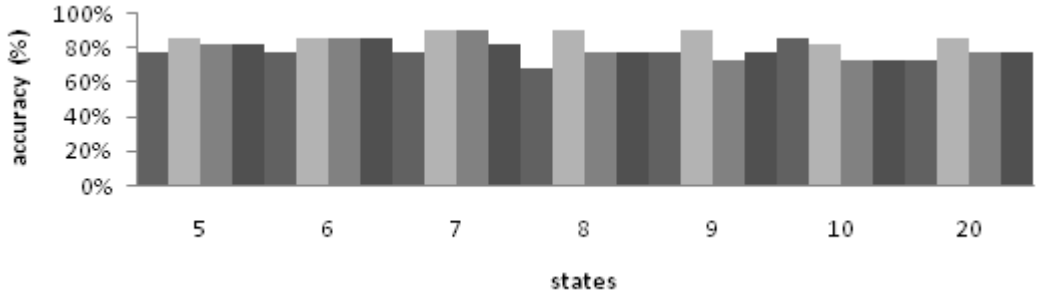

Figure 4. Test result for Female in a variation of the size of the codebook

Based on Figures 4 and 5, we can conclude that the lack of variation in training data can reduce the accuracy of the HMM method, where the lowest level of accuracy provided by the HMM system is in testing for male speaker, with codebook size 16 and the number of states 6. Therefore, testing the hybrid HMM-GA system further, using a codebook size of 16 and states 6 for male speakers.

The next test is used the HMM-GA system for the worst result in HMM systems. In this test, the GA parameters measured were the number of maximum number of individuals this is because the number of Max generations depends on max number of individuals and population size. Here can be seen from the data of HMM parameter optimization that gives the worst accuracy rate obtained HMM-GA Hybrid system for each parameter.

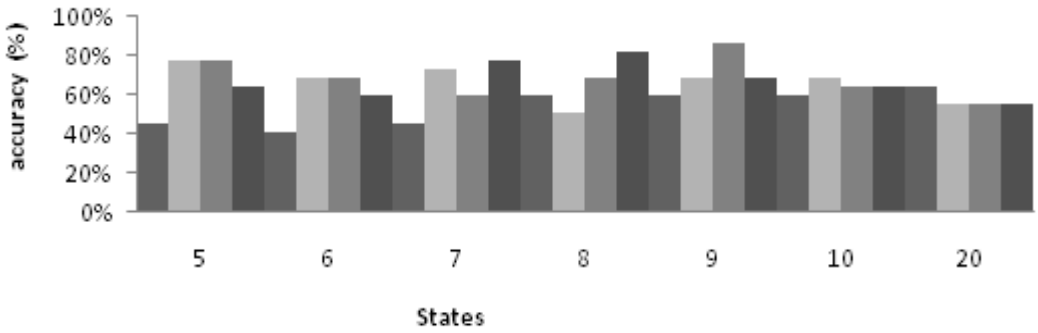

Figure 5. Test result for male in a variation of the size of the codebook

This data was obtained by conducting tests in which the max number of individuals is only 1000, which means not so much variation of individual chromosome with max number 10000. To test maxJumInd 1000 is used in combination with population size 50, 100, 500 and the max generation is calculated by maxJumIndividu divided by population size.

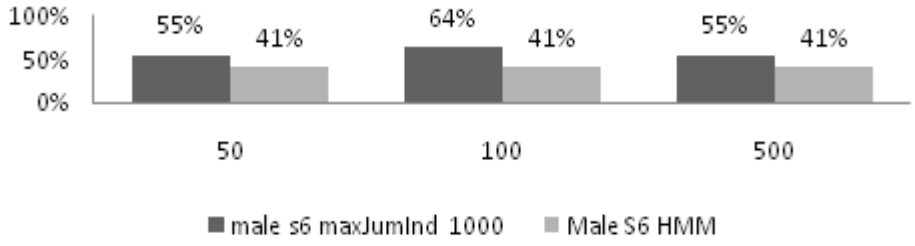

Figure 6. The accuray level for hybrid HMM-GA with MaxJumInd 1000

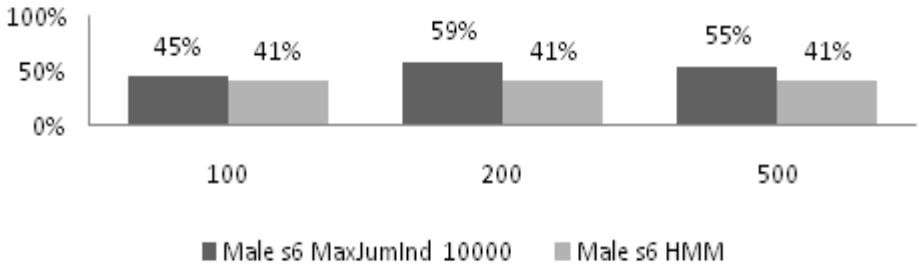

Figure 7. The accuray level for hybrid HMM-GA with MaxJumInd 10000

Based on Figures 6 and 7, can be seen that the increasing the level of accuracy provided by the system HMM-GA against HMM system, but a rise in the level of accuracy given is not fixed, along with the increase in the size of the population. It could be that given the level of accuracy worse or the same HMM system before optimizations. This is because the characteristic of the GA method is random and this system using the standard general replacement process. This method did generate a random population, which is resulted in the lack of certainty in obtaining optimal results. Therefore, it should be tested one by one all the combinations of parameters to obtain optimum results, but it should be, conducted observations of all combinations of parameters to get the combination of parameters that can deliver optimal results and also the condition of the speech signal has a noise that is not equal to zero can also be very influential in the process of recognition where it will lead to the original characteristics of the extraction of the speech signal will be more difficult.

3.1. Comparison of HMM system and HMM-GA system

Based on the results of tests performed on the HMM system, we can conclude that the increase in system performance is the worst outcome HMM with parameter codebook 16 and states 6, amounting to 20-40% of the test results on HMM-GA system. The value of this increase is the average of the HMM-GA parameters are tested to maxJumInd 1000, the percentage increase of 41%, while in maxJumInd 10000, the percentage increase of 29% only for this case.

The main cause of the recognition of the word that its performance is still poor due to the influence of conditions of each case in which for each case the performance can vary, which means the use of states and parameters used can not be as certained for each case. An increase in performance is due to the ability of GA in generating and optimizing the model for each class. Optimizing Performance of HMM systems using GA can be done by observing the parameters of GA to the system, then test one by one these combinations. With the addition using GA has the ability to optimize, but sometimes also can produce solutions that are not optimal, because basically the ability of this method in finding an optimal solution is sometimes influenced by the use of genetic operators of a chromosome that produces a rich poorer.

4. Conclusion

From the test results, it can be concluded that genetic algorithm is a significant effect on speech recognition systems because the algorithm is to optimize the HMM model for each training signal by finding a solution that no other individual parameters of HMM substantial set of solutions. The greater the population size is used, the greater the chance of finding a solution that can truly represent the training signal. But that does not mean that the greater of the degree of accuracy is given.

By using hybrid HMM-GA method, proven to increase the percentage of accuracy of speech recognition result that is, amounting to 20-40% in the parameter - parameter HMM system which produces the lowest or worst accuracy because of the lack of variation of training data.

References

[1] S Kwong, Chau W. Analysis of Parallel Genetic Algorithms on HMM Based Speech Recognition System. City University of Hong Kong. *IEEE Transactions on Consumer Electronics.* 1997; 43(4)
[2] Glass J. *Introduction to Automatic Speech Recognition*. 2003.
[3] Cho SJ. *Introduction to Hidden Markov Model and Its Application*. Samsung Advanced Institute of Technology (SAIT). 2005.
[4] Shing-Tai P, Ching-Fa C, Jian-Hong Z. *Speech Recognition via Hidden Markov Model and Neural Network Trained by Genetic Algorithm*. Ninth International Conference on Machine Learning and Cybernetics. Qingdao, 11-14 July 2010.
[5] SA Selouani, D O'Shaughnessy. Robustness of Speech Recognition Using Genetic Algorithms and A Mel-Cepstral Subspace Approach. *IEEE Transactions on Neural Network*. 2004.
[6] Shing-Tai P, hih-Hung W, Shih-Chin L. The Application of Improved Genetic Algorithm on The training of Neural Network for Speech Recognition. *IEEE Transactions on Neural Network*. 2007.
[7] Bhaskar MM, Maheswarapu S. A Hybrid Genetic Algorithm Approach for Optimal Power Flow. *TELKOMNIKA.* 2011; 9(2): 211-216.
[8] Cernock J. *Speech Recognition Based on Hidden Markov Models*. Brno University of Technology. 2007.
[9] Terry L. *Ergodic Hidden Markov Models for Visual-Only Isolated Digit Recognition*. Dissertation. Electical Engineering Northwestern University: Evanston, Illinois. 2007.
[10] Rabiner LJ, Biing-Hwang. *Fundamentals of Speech Recognition.* Prentice Hall International Edition.1993.
[11] Xueying Z, Yiping W, Zhefeng Z. A Hybrid Speech Recognition Training Method for HMM Based on Genentic Algorithm and Baum Welch Algorithm. *IEEE Transactions on Neural Network*. 2007.
[12] Mourad O, Raja NA. HMM Parameters Estimation Using Hybrid. IEEE. 2010.
[13] Chu SH. *Combination of GA and SDM to Improve ANN Training Effiiency.* Shu-Te University. MS Thesis. 2003.
[14] Preetham S, Kedar N, Lakshmi. *Speech Recognition using HMMs*. IIIT-Hyderabad. 2003.
[15] Jain AK, Duin RPW, Mao J. Statistical Pattern Recognition: A Review*. IEEE Transaction on Pattern Analysis and Machine Intelligence*. 2000.
[16] Kim FM, Tang KS, Kwong S. Genetic Algortihm: Concept and Design. Springer. 2001.

[17] Sutikno T, Facta M, Markadeh GRA. Progress in Artificial Intelligence Techniques: from Brain to Emotion. *TELKOMNIKA.* 2011; 9(2): 201-202.
[18] Lan ML, Pan ST, Lai CC. *Using Genetic Algorithm to Improve the Performance of Speech Recognition Based on Artificial Neural Network*. Dept.of Computer Science and Information Engineering, Shu-Te University Taiwan. Proceesings of the First International on Innovative Computing, Information and Control (ICICIC'06).