제 11 회 음성통신 및 신호처리 워크샵 논문집(제 SCAS-11권 1호)

A STUDY ON THE SIMULATED ANNEALING OF SELF ORGANIZED MAP ALGORITHM FOR KOREAN PHONEME RECOGNITION

°Myung-Kwang Kang, *Tae-Ock Ann, °Lee-Hyung Kim, °Soon-Hyob Kim

° Dept. of Computer Engineering
Kwang-Woon Univ. Seoul, Korea

* Dept. of Computer Engineering
Chŏn Buk Sanup Univ.

Abstract

In this paper, we describe the new unsupervised learning algorithm, SASOM(Simulated Annealing Self Organized Map). It can solve the defects of the conventional SOM(Self-Organized Map) that the state of network can't converge to the minimum point.

The proposed algorithm uses the object function which can evaluate the state of network in learning and adjusts the learning rate flexibly according to the evaluation of the object function. We implement the simulated annealing which is applied to the conventional network using the object function and the learning rate. Finally, the proposed algorithm can make the state of network converged to the global minimum. Using the two-dimensional input vectors with uniform distribution, we graphically compared the ordering ability of SOM with that of SASOM. We carried out the recognition on the new algorithm for all Korean phonemes and some continuous speech.

I. Introduction

In speech recognition, many researches have been progressed and various discriminators have been proposed. Recently, the speech recognition using HMM(Hidden Markov Model), NN(Neural Network) and hybrid method of both has been studied widely.

NN model can roughly be divided into two models according to the learning methods, supervised learning and unsupervised learning, respectively. In supervised learning, the information of desired output are provided by teacher and accomplished learning. In unsupervised learning, learning must somehow be accomplished based on observations of responses to inputs that we have no knowledge about[7]. One of the representative unsupervised learning models is SOM(Self Organizing Map) model proposed by T.Kohonen. In this paper, we propose the SASOM(Simulated Annealing Self Organizing Map) learning algorithm which can solve the defects of the conventional SOM algorithm.

The organization of this paper is as follows. In Section II, we review the theory of conventional SOM and show the defects of it. Section III describes the Simulated Annealing for improving SOM. Section IV explains the new algorithm,SASOM. Section V shows the result of experiments and we discuss the conclusion in Section VI.

II. Kohonen's SOM Algorithm

A. Conventional SOM Algorithm[1][2][4][5]

SOM algorithm is a feature classifier representing the probability distribution of input pattern as topological maps to the two dimensional output layer through competitive learning and lateral inhibition.

SOM learning consists of two processes, the selecting a similar cell to input pattern and the modifying weight vectors of the selected cell. Inner product or Euclidean distance measures their similarity.

Since all cells compete against other cells for being learned, the learning in SOM is called "competitive learning". The best matching vector, a winner, has the chance to be learned. In this point, the Kohonen network is a "winner-take-all fashion"[3]. The learning algorithm adjusts the weight vectors in the vicinity of the winning neuron according to learning rules. It is as follows;

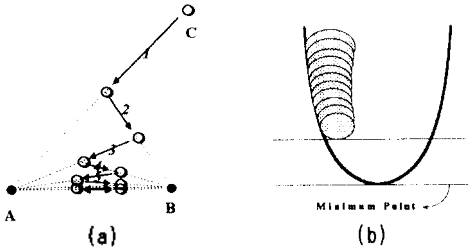
$$M_i(t+1) = \begin{cases} M_i(t) + \alpha(t)[\ X(t) - M_i(t)] & \text{if } i \in N_c(t) \\ M_i(t) & \text{if } i \notin N_c(t) \end{cases}$$

where, $X(t)$ is input pattern, M_i is weight vector, $N_c(t)$ is a neighbor cell which can be learned with winner cell. $\alpha(t)$ is adaptation gain, $0< \alpha(t)<1$, learning rate. $\alpha(t)$ is decreased monotonically with time.

We use the lateral inhibition which modifies the weights of neighborhood cells as well as that of the winner cell. Kohonen`s algorithm creates topological organized map of various features of input patterns. It is analogous to the basic functions of the biological neuron. The effect of lateral inhibition is similar to that of Mexican hat function. The neighborhood starts with large number, and is gradually decreased with time. Eventually, the neighborhood includes only the nearest cells around the winner cell. The advantages of Kohonen's SOM algorithm are unsupervised learning, representing unknown input pattern by producing the topological map which mapped input signal space into output layer, and continous learning through competitive learning.

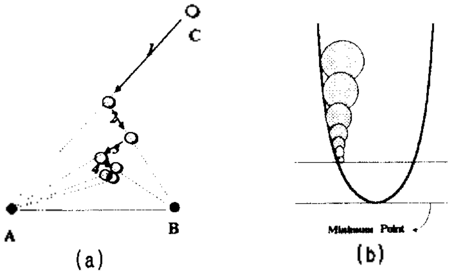
B. The Defects of SOM Algorithm[10]

It is not guaranteed that the state of network converges to the minimum point in learning. In case that the decreasing ratio of learning rate is near to 1, Fig. 1, because the learning includes unnecessary oscillation of the connected vector progresses, the state of the network moves slowly toward the minimum point. Because the termination of the learning is determined by the iteration number, it can not be assured for the state of network to converge the optimal state in learning. In case that the decreasing ratio of learning rate is much less than 1, because the learning rate rapidly decreases to zero, the subsequent learning is meaningless and the state of network may not converge to the optimal state, Fig. 2. We have to determine the changing pattern of learning rate carefully. If we know the probability distribution of learning pattern, we can make the network learned with determining easily the changing pattern of a proper learning rate to that distribution. Actually, not knowing the distribution, we have to determine the distribution by lots of experiments. Nevertheless, the changing pattern of learning rate which is selected by lots of experiments does not assure that network converges to global minimum point. Fig. 3 show that the state of network dependents on the initial weights in learning.



(a) Pattern A and B, and weight vector C
(b) The state of network

Fig. 1 In case that the decreasing ratio of learning rate is near to 1.



(a) Pattern A and B, and weight vector C
(b) The state of network

Fig. 2 In case that the decreasing ratio of learning rate is much less than 1.

III. Simulated Annealing[5][6]

For SOM algorithm adjusts the network weights, following the local slope of the error surface toward a minimum, the network may converge the local minimum point. It is the Simulated Annealing method that can solve the above defect. Simulated Annealing is modeled by the physical processing that heats until the solid melts completely in metal and cools until it becomes structure of crystal. In this process, free energy of solid state is minimized[7]. In order to minimize the free energy, sufficient time must be allowed at each temperature so that the solid reaches an equilibrium. In equilibrium, the solid follows the canonical probability distribution ;

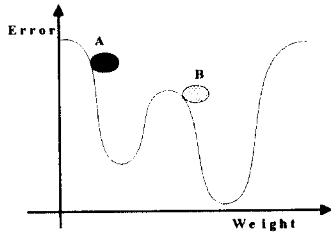


Fig. 3 The state of network dependents on the initial weights in learning.

$$P(E_r) \propto e^{-E_r/k_B T}$$

Consider the simple energy landscape shown in Fig. 4. Assume that we shake the whole system with fixed power($\Delta E_1 < p < \Delta E_2$) and the ball moves through the outer space of energy in dependent power P. If we shake the whole system with power P, the ball rolls down from E_s to the minimum point E_a with energy P. The ball can reach the minimum of E_b because the ball energy is higher than ΔE_1. But because the ball energy is lower than ΔE_2, it can't get up the hill, E_a again. By decreasing the power P which shakes the whole system gradually, the ball reaches E_b consequently.

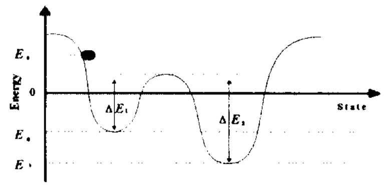


Fig. 4 The surface of energy that have two minimum point

IV. SASOM(Simulated Annealing of SOM)

Applying this method to network learning, we have to define the network energy which corresponds the energy level of annealing and the artificial temperature. And we have to establish the function with canonical probability distribution in equilibrium. The representative simulated annealing methods are Boltzman machine and Cauchy machine.

In order to evaluate the state of network, the proposed algorithm defines the object function as follows

$$\Phi = \frac{1}{N}\sum_{i,j} d^2(X_i - M_j)$$

i is index of learning pattern
j is the index of winner cell of ith learning pattern
N is the iteration number

If the N is small, we can evaluate the state of network for entire learning pattern. If the N is large, because the learning time increases excessively, we have to draw out the sample data which can represent the distribution of learning pattern [8]. And the network state is evaluated for that sample set. To implement the simulated annealing, we define the energy level through learning rate. When the learning continues with fixed learning rate, the state of network changes in Fig 5.

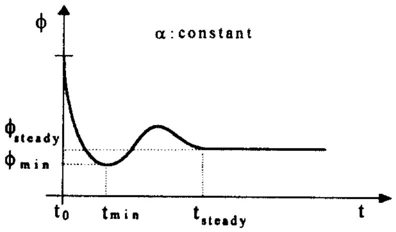


Fig. 5 The state of network with constant learning rate

Also, if α_1 is less than α_2, learning rate can be defined as the energy level of network because Φ_{α_1} is less than Φ_{α_2}. In the proposed algorithm, simulated annealing method uses a Cauchy machine. Artificial temperature of Boltzman machine decreases in inverse log function and Cauchy machine decreases in inverse linear function[9]. Because of the large changing probability into the high energy level, Cauchy machine can converge to the global minimum point more fast than Boltzman machine. Artificial temperature of Cauchy machine is as follows;

$$T(t) = T_0/(1+t), \qquad T_0 : \text{initial artificial temperature}$$

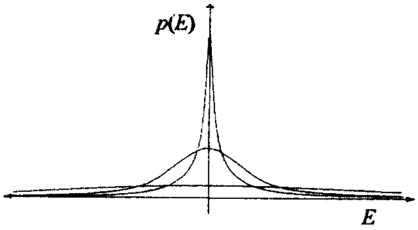


Fig. 6 The probability distribution according to the change of artificial temperature

and probability distribution is given by;

$$P(E) = T(t)/(T(t)^2 + E^2)$$

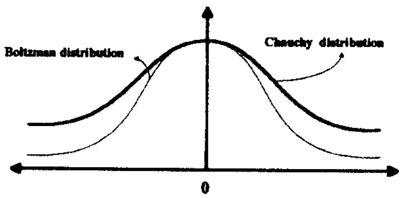


Fig. 7 The Cauchy probability distribution

The proposed algorithm uses lateral inhibition in order to form the feature map of output layer. And the range of lateral inhibition is defined by learning rate. When the learning rate is less than 0.01, the range of lateral inhibition is limited to the nearest cells and this value can be selected by experiments. SASOM algorithm is as follows;

Step 1. Initialize the T_0 and α ,

Step 2. Learning the network until its state converges the most stable for a learning rate,

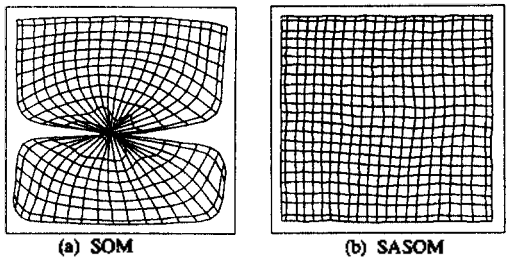
Step 3. Calculate the $T(t)$,

Step 4. Calculate the learning rate α,

Step 5. Terminate the learning if mean changing value of learning rate or artificial temperature is less than the critical value, otherwise repeat the step 2 to step 5.

V. Experiment

To evaluate the performance of proposed algorithm, we carry out three experiments as follows. The first is performed for the ordering ability with SOM and SASOM respectively. It is shown in Fig. 8 (a) that the state of network can't converge to global minimum point but a local minimum point, after ordering process of SOM algorithm. However, it is shown that the state of network converges to global minimum point, in the Fig. 8 (b), after ordering process of SASOM algorithm in identical environments.



(a) SOM (b) SASOM

Fig. 8 The weight vectors in ordering process, two dimensional array

The second is phoneme recognition experiment which is performed with all Korean phoneme classes; vowels /a/, /i/, /u/, /ɛ/, /o/, /ɨ/, /ə/, unvoiced stop /p/, /t/, /k/, /pʰ/, /tʰ/, /kʰ/, /p'/, /t'/, /k'/, voiced stop /b/, /d/, /g/, liquid and nasals /r/, /m/, /n/, /G/, fricatives /s/, /s'/, /h/, affricative /Č/, /Čʰ/, /Č'/, and final stop /p'/, /t'/, /k'/. The speech data is sampled at 10kHz, bandpass filtered and digitized with a 16 bits A/D converter. The 16th bandpass filter bank coefficients are used as a input feature vector. And the experiments is performed in speaker dependent. The number of tokens is 792 for learning, and 1256 for testing. The recognition rate is 88.08% for total phonemes, and as follows for each class.

Finally, we perform the simple continuous speech recognition experiments for Korean phonemes /k/, /t/, /p/, /a/, /i/, /u/, /a/, /o/, /i/. In training process, we use a raw data, directly, without segmentation, /ka/, /ta/, /pa/, /ki/, /ti/, /pi/, /ku/, /tu/, /pu/, /ka/, /ta/, /pa/, /ko/, /to/, /po/, /ki/, /ti/,

Table 1. The result of recognition experiment

task	#errors/ #tokens	%correct	total %
a	0/40	100	98.75
e	0/40	100	
o	3/40	92.5	
u	0/40	100	
ɨ	0/40	100	
i	0/40	100	
ɛ	0/40	100	
d	3/32	90.625	85.416
b	6/32	81.25	
g	5/32	84.375	
p	9/40	77.5	85
t	7/40	82.5	
k	2/40	95	
p^h	5/40	87.5	88.33
t^h	7/40	82.5	
k^h	2/40	95	

task	#errors/ #tokens	%correct	total %
p'	6/40	85	91.66
t'	2/40	95	
k'	2/40	96	
p'	9/40	77.5	86.66
t'	4/40	90	
k'	3/40	92.5	
r(l)	1/40(9/40)	97(77.5)	83.5
m	8/40	80	
n	10/40	75	
ŋ	5/40	88	
s	9/40	77	85
s'	7/40	83	
h	2/40	96	
č	10/40	75	73.33
$č^h$	16/40	60	
č'	6/40	85	

Total recognition rate (%) : 88.08

/pi/. The unit of input token to network is frame. It is shown that the recognition result is about 90% (77.7% ; consonant).

VI. Conclusion

In this paper, we propose SASOM algorithm that improves the problem of SOM algorithm which cannot make the state of network converged to the global minimum. We perform experiments that compare the ordering ability of SASOM with that of SOM, recognize all korean phonemes for estimating the discriminant ability of SASOM algorithm. Phoneme recognition in continuos utterances is performed using simple data for evaluating the applicable possibility of continuos speech.

In the result of the experiments, we can confirm the performance of SASOM is better than that of conventional SOM, and SASOM algorithm can be applied to recognize the continuous speech directly without segmentation.

References

[1]. T. Kohonen, "The self-organizing map", Proc. of the IEEE, 78, 9, 1464-1477, 1990.

[2]. T. Kohonen, "Self-organized and associative memory", Spring Verlag, 3d ed., New York, 312 p, 1989.

[3]. D. P. Morgan, C. L. Scofield, "Neural Networks and Speech Processing", Kluwer Acardemic, 1991.

[4]. T. Kohonen, "Self-organized formation of topo -logically correct feature maps", Biological Cybernetics, 43, 62, 1982.

[5]. P. D. Wasserman, "Neural Computing Theory and Practice", Van Nostrand Reinhold, 1989.

[6]. J. A. Freeman, D. M. Skapura, "Neural Networks", Addison-Wesley, 1991.

[7]. J. M. Zurada, "Artificial Neural Systems", West, 1992.

[8]. A. Papoulis, "Probability, Random Variables, and Stochastic Processes", McGraw-Hill, 3d ed., 1991.

[9]. P. D. Wasserman, "A Combined Back-Propagation/Cauchy Machine Network", Journal of Neural Network Computing, 1990.

[10].Y. K. Jeon, Soon Hyob Kim, "Korean Phoneme Recognition by Combining Self-Organized Feature Map whith K-means clustering algorithm", WESTPRAC, 1994.