SETIT 2009
5th International Conference: **S**ciences of **E**lectronic, **T**echnologies of **I**nformation and **T**elecommunications
March 22-26, 2009 – TUNISIA

New Variant of the Growing Hierarchical Self Organizing Map GH-DeSieno-SOM for Phoneme Recognition

Chiraz JLASSI*, Najet AROUS* and Noureddine ELLOUZE*

* *École Nationale d'Ingénieurs de Tunis, BP 37, Belvédère 1002 Tunis, TUNISIE*

Chiraz_jlassi@yahoo.fr

Najet.Arous@enit.rnu.tn

N_Ellouze@enit.rnu.tn

Abstract: The Growing Hierarchical Self-Organizing Map (GHSOM) is a network of neurons whose architecture combines two principal extensions of SOM model, the dynamic growth and the tree structure.
This paper presents a variant of the growing GHSOM. The proposed variant is like the basic GHSOM. However, it is characterized for each neuron of each map level by a conscious term which takes into consideration how many a unit has won a competition and multiple prototype vectors (general centroid vector and more than a mean vector). In this paper, we study the learning quality of the proposed GHSOM variant and we show that it is able to reach good phoneme recognition rates.
Key words: Neural network, growing hierarchical self-organizing map, conscious term, prototype vector, phoneme recognition.

INTRODUCTION

In speech recognition applications many neural networks architecture have been used successfully, we cited: multi-layer perceptrons, radial basis functions, and self-organizing maps (SOM) [JEN 91], [KAN 94], [TAD 95], [VES 97]…

The SOM, is a structured non hierarchical network of simple laterally interconnected computing units. This means that all neurons are of the same kind and they are all connected to their nearest neighbors with respect to the structure of the network.

Among the large number of research publications discussing the SOM [KAS 98], [OJA 03] different variants and extensions have been introduced. Some of the extensions of the SOM algorithm and architecture address the disadvantages of fixed size and missing hierarchical representation. One of the SOM based models implementing an algorithm dealing with both issues is the Growing Hierarchical Self-Organizing Map (GHSOM) [DIT 00], [DIT 01], [DIT 02], [DIT 05].

The GHSOM is a neural architecture combining the advantages of two principal extensions of the self-organizing map, dynamic growth and hierarchical structure. Basically, this neural network model is composed of independent SOMs (many SOM), each of which is allowed to grow in size during the training process until a certain quality criterion regarding data representation is met. This growth process is further continued to form a layered architecture such that hierarchical relations between input data are further detailed at lower layers of the hierarchy. Consequently, the structure of this adaptive architecture automatically adapts itself according to the structure of the input space during the training process. So far, the quantization error has been used as a measure to automatically guide the growth process of the architecture, both in terms of map and hierarchical growth. In other words, the single maps are allowed to grow until a certain quality criterion depending on the quantization error of a higher-layer unit is reached. Moreover, the expansion of the hierarchy into further layers also depends on the quantization errors of the single units on a map. Consequently, each layer deeper in the hierarchy contains maps that represent the data at a higher level of granularity. Depending on the main parameter that guides the training process, the resulting structure is either a flat hierarchy with rather large maps or a deep hierarchy with rather small maps.

In this paper, we are interested in phoneme recognition by means of a GHSOM variant where each unit is characterized by a conscious term and more than a prototype vector. The front-end preprocessor to the proposed competitive learning

algorithm is a matrix of real-valued 12-dimensional vectors of mel cepstrum coefficients. Each output unit of the proposed GHSOM variant is described by a general centroid vector and information relating to each phoneme class described by a mean vector, a label and an activation frequency.

In the following, we present the basic model of the SOM and the GHSOM. Thereafter, we present the proposed competitive learning algorithm. Then we explain the recognition strategy. Finally, we illustrate experimental results of the application of the proposed GHSOM variant to recognize phoneme of TIMIT database.

1. Growing hierarchical SOM

The SOM is a nonlinear, ordered, smooth mapping of highdimensional input data onto the elements of a regular, low-dimensional (usually 2D) array. The SOM consists of a set of i units arranged in a 2D grid with a weight vector mi attached to each unit, which may be initialized randomly. Input vectors x are presented to the SOM and the activation of each unit for the presented input vector is calculated using an activation function. Commonly, it is the Euclidian distance between the weight vector of the unit and the input vector that serves as the activation function. In the next step, the weight vector of the unit showing the highest activation (i.e. the smallest Euclidian distance) is selected as the "winner" c_k where

$$c_k = \arg\min \left\| x_k - m_i \right\| \qquad (1)$$

The weight vector of the winner is moved toward the presented input signal by a certain fraction of the Euclidean distance as indicated by a time-decreasing learning rate α. The learning rate α can be an inverse time, linear or power function. Thus, this unit's activation will be even higher the next time the same input signal is presented. Moreover, the weight vectors of units in the neighborhood of the winner are also modified according to a spatial-temporal neighborhood function ε. Similar to the learning rate, the neighborhood function ε is time-decreasing. Also, ε decreases spatially away from the winner. There are many types of neighbourhood function, and the typical one is Gaussian. The learning rule may be expressed as

$$m_i(t+1) = m_i(t) + \alpha(t) \cdot \varepsilon(t) \cdot [x(t) - m_i(t)] \qquad (2)$$

Where t denotes the current learning iteration and x represents the currently presented input pattern. This learning procedure leads to a topologically ordered mapping of the presented input data. Similar patterns are mapped onto neighboring regions on the map, while dissimilar patterns further apart.

The GHSOM enhances the capabilities of the basic SOM in two ways. The first is to use an incrementally growing version of the SOM, which does not require the user to directly specify the size of the map beforehand; the second enhancement is the ability to adapt to hierarchical structures in the data. Prior to the training process a "map" in layer 0 consisting of only one unit is created. This unit's weight vector is initialized as the mean of all input vectors and its mean quantization error (MQE) is computed. The MQE of unit i is computed as

$$MQE_i = \frac{1}{|U_i|} \sum_{k \in U_i} \left\| x_k - m_i \right\| \quad U_i = \{k / c_k = i\} \qquad (3)$$

Beneath the layer 0 map a new SOM is created with a size of initially 2×2 units. The intention is to increase the map size until all data items are represented well. A mean of all MQE_i is obtained as <MQE>. The <MQE> is then compared to the MQE in the layer above, <MQE>above. If the following, inequality is fulfilled a new row or column of map units are inserted in the SOM.

$$MQE \quad > \tau_1 \cdot \langle MQE \rangle_{above} \qquad (4)$$

Where τ1 is a user defined parameter. Once the decision is made to insert new units the remaining question is where to do so. In the GHSOM array, the unit i with the largest MQE_i is defined as the error unit. Then the most dissimilar adjacent neighbor, i.e., the unit with the largest distance in respect to the model vector, is selected and a new row or column is inserted between these. If the inequality (4) is not satisfied, the next decision to be made is if some units should be expanded on the next hierarchical level or not. If the data mapped onto one single unit i still has a larger variation, i.e.,

$$MQE_i > \tau_2 \langle MQE \rangle_{above} \qquad (5)$$

Where τ_2 is a user defined parameter, then a new map will be added at a subsequent layer.

Generally, the values for τ_1 and τ_2 are chosen such that $1 > \tau_1 >> \tau_2 > 0$. In [PAM 04] the GHSOM parameter, τ_1 and τ_2 are called "breadth"- and "depth"-controlling parameters, respectively. Generally, the smaller the parameter τ_1, the larger the SOM arrays will be. The smaller the parameter τ_2, the more layers the GHSOM will have in the hierarchy.

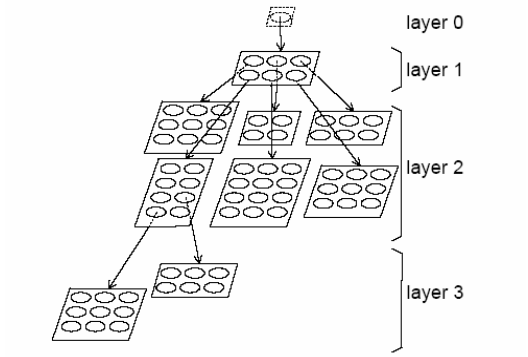

Figure1. *Growing Hierarchical Self-Organizing Map (GHSOM)*

2. The proposed competitive learning GH-DeSieno-SOM

The purpose of the proposed improvement learning is to form a better approximation of a probability density function of an input sample x. When an input vector is presented to the network, a competition is held to determine which neuron's weight vector w_i is closest in any metric to the input vector called BMU. The winning unit in the competition is not necessarily the element to have its weights reinforced. A bias is developed for each neuron based on the number of times it has won the competition. The weights of the unit winning this biased competition are adjusted as follows [DES 98]:

$$w_i(t+1) = w_i(t) + \alpha(t) h_{ci}(t)[x(t) - w_i(t)] - c\left(\frac{1}{n} - p_i(t)\right)$$

$$\forall i \in [1 .. n] \qquad (6)$$

where c is a positive scalar, n is the number of neurons, α(t) the learning rate at time t, $h_{ci}(t)$ the neighborhood kernel around the winner unit c and p_i is the likelihood of selection of unit i.

As a network is trained using this algorithm and processing elements start winning their share of the competitions, the bias terms have less impact and the process tends to revert to a simple Kohonen learning rule. The term conscience arises because a processing element that wins too often begins to "feel guilty" and prevents itself from winning excessively.

During training phase, when a sample input vector is attributed to a BMU unit, we save its corresponding vector, its label (since TIMIT corpus is labeled) and we update its frequency activation [ARO 01]. By this way, each unit of each Kohonen map is charaterized by:

- A general centroid vector (GCV): determined by means of Kohonen update rule.
- Information relating to each phoneme class attributed to a unit: a mean vector (MV), a label and an activation frequency. [ARO 03].

3. Recognition strategy

The recognition was performed at the frame level and the performance was evaluated by comparing each classified frame against the reference frame. Decision recognition is operated in two steps:

First step: for each test sample vector presented to the proposed competitive algorithm we search for the BMU among all general centroid vectors (GCV) of a map.

Second step: inside the selected BMU unit, we search for the best mean vector (MV) of different classes of the selected unit, in terms of minimal euclidean distance.

This process is repeated layer by layer using knowledge about the BMU of the frozen layer (l-1) in the search of the BMU on the next layer (l). For example the search of the BMU in the second layer is restricted into the map connected to BMU of the first layer. And when we are in the last level of the hierarchy, we look for the label of the last BMU.

4. Experimental results

We have implemented a variant of GHSOM network for continuous speech recognition. The realized system is composed of three main components: a pre-processor for phoneme sounds and producing mel cepstrum vectors. The sound input space is composed by 12 mel cepstrum coefficients each 16 ms frame. The second component is a competitive learning module based on a conscious term and multiple prototype vectors for each neuron. The third component is a phoneme recognition module.

The speech database used is the DARPA TIMIT acoustic-phonetic continuous speech corpus.

4.1. TIMIT Database

We have used the TIMIT corpus for the purpose to evaluate the proposed competitive learning algorithm for speaker independent continuous speech recognition. TIMIT contains a total of 6300 sentences, 10 sentences spoken by each of 630 speakers from 8 major dialect regions of the United States. A speaker's dialect region is the geographical area of the U.S. where they lived their childhood years. The data was recorded at a sample rate of 16 KHz and a resolution of 16 bits. In our experiments, we have used the New England dialect region (DR1) composed of 31 male and 18 female. The corpus contains 14 399 phonetic unit for training. Training has been made on phonemes for the seven macro classes of TIMIT database.

Table1. *Classification rates of the 7 macro-classes of TIMIT database (train test).*

	$\tau_1 = 0.7$ $\tau_2 = 0.02$		$\tau_1 = 0.8$ $\tau_2 = 0.02$	
	Basic GHSOM	**GH-DeSieno-SOM**	**Basic GHSOM**	**GH-DeSieno-SOM**
Affricates	65.13	75.15	59.23	71.11
Stops	39.50	70.36	43.00	60.15
Others	42.13	77.56	39.65	52.12
Nasales	53.12	80.11	41.95	58.46
Semivowels	59.47	71.56	40.20	65.23
Fricatives	60.22	84.23	52.47	69.12
Vowels	44.89	62.10	40.45	49.56
Mean classification rates	52.06	74.43	45.27	60.82

We have implemented cited above competitive learning algorithm. Two experiments were conducted. In the first, the parameter r_1 "breadth" which controls the actual growth process is equal to 0.7 and r_2 parameter "depth" which controls the minimum granularity of data representation is equal to 0.02, the generated models have more maps in the hierarchy (for example figure2) than the second experiment, when the parameter r_1 is equal to 0.8 and r_2 is equal to 0.02 (for example figure3). Table 1 shows a comparison of different recognition rates obtained by using respectively the different models cited above.

Table 1 shows a comparison of different classification rates of the macro class of TIMIT database obtained by using respectively, GHSOM based on a sequential learning (basic GHSOM) and GHSOM with a conscience term (GH-DeSieno-SOM). The implemented competitive learning algorithm provides good phoneme recognition rates accuracy. We should note that recognition rates depend on the parameter r_1 which controls the actual growth process. Also, we should note that the use of a conscious term proves notable recognition rates improvement comparing with the basic GHSOM. We suggest studying other GHSOM variants to the case study phoneme recognition of TIMIT database to improve the recognition rates.

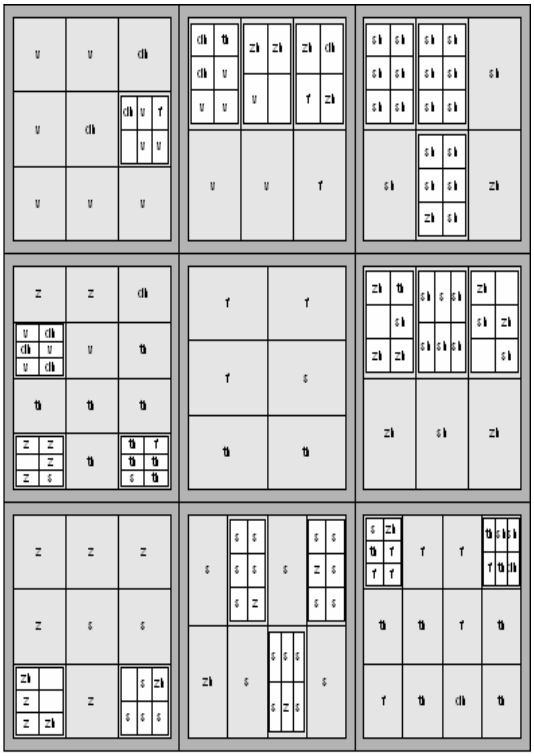

Figure 2. *GHSOM with 3 layers and 30 Maps (one map in the first layer, 9 maps in the second and 20 maps in the third) labelled by the fricatives of TIMIT database $\tau_1 = 0.7$ and $\tau_2 = 0.02$ (result of the GH-DeSieno-SOM variant).*

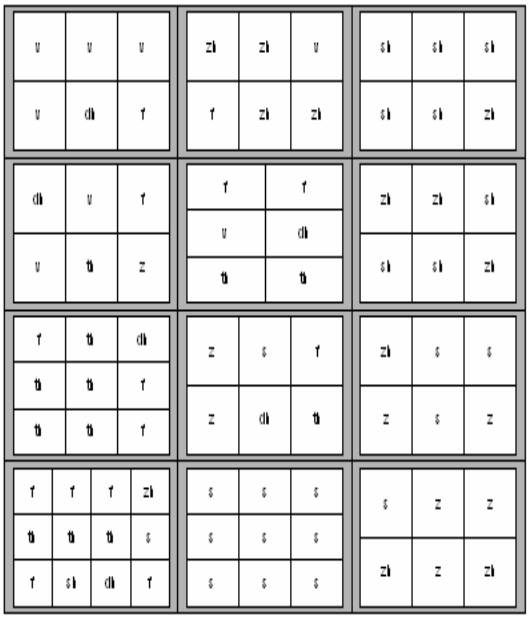

Figure 3. *GHSOM with 2 layers and 13 Maps (one map in the first layer and 12 maps in the second) labelled by the fricatives of TIMIT database $\tau_1 = 0.8$ and $\tau_2 = 0.02$ (result of the GH-DeSieno-SOM variant).*

5. Conclusion

In this paper, we are interested in phoneme recognition by means of a GHSOM variant where each neuron of each map level is characterized by a conscious term and more than a prototype vector, and we have study the learning quality of the competitive learning algorithm. The conscious term prevents each unit from winning excessively many competitions. We should note that the use of a conscious term prove notable recognition rates improvement than the basic GHSOM. On the other hand, the multiple prototype vectors (general centroid vector and more than a mean vector) describing a neuron determines precisely which reference vector the most resemble to a sample input vector in term of euclidean distance.

In the basic GHSOM and the proposed variant the "breadth" parameter τ_1 which controls the actual growth process, affects the hierarchy of the obtained models and the classification rates.

Finally, we suggest proposing other GHSOM variants for phoneme recognition in order to improve classification rates.

REFERENCES

[ARO 03] N. Arous and N. Ellouze. "Cooperative supervised and unsupervised learning algorithm for phoneme recognition in continuous speech and seaker-independent context". *Elsevier Science, Neurocomputing, Special Issue on Neural Pattern* Recognition, 51, pp. 225 – 235.

[ARO 01] N. Arous. "Speech Clustering by Means of Self Organizing Maps". Tunisan-German Conference on Smart Systems and Devices, Hammamet-Tunisia, 27-30.

[DIT 05] M. Dittenbach, A. Rauber and Polzlbauer. "Investigation of alternative strategies and quality measures for controlling the growth process of the growing hierarchical self-organizing map". Proceeding of the International Joint Conference on Neural Networks. (IJCNN 2005) pp 2954-2959. Canada.

[DIT 02] M. Dittenbach, A. Rauber and D. Merkl. "Uncovering the hierarchical structure in data using the growing Hierarchical self-organizing map". Neurocomputing. Vol48, no 1-4, pp199-216.

[DIT 01] M. Dittenbach, A. Rauber and D. Merkl . "Recent advances with the growing hierarchical self-organizing map" in Advances in Self-Organizing Maps.

[DIT 00] M. Dittenbach, A. Rauber and D. Merkl. "The growing hierarchical selforganizing map". Proceedings of the International Joint Conference on Neural Networks (IJCNN 2000).

[DES 88] D. DeSienov. "Adding a conscience to competitive learning". IEEE International Conference on Neural Networks, 117-124.

[FRI 95] B. Fritzke. "Growing grid – a self-organizing network with constant neighborhood range and adaptation strength". Neural Processing Letters, Vol. 2, pp. 9-13.

[JEN 91] O. B. Jensen M. Olsen and T. Rohde. "Automatic speech recognition & neural networks". Thesis in computer science at the Computer Science Department of Arhus University, Denmark.

[KAN 94] J. Kangas. "On the Analysis of pattern sequences by self-organizing maps". Thesis for the degree of Doctor of Technology Helsinki University of Technology.

[KAS 98] S. Kaski, J. Kangas and T. Kohonen. "Bibliography of self-organizing map (SOM) papers". Neural Computing Surveys, vol. 1, no. 3&4, 1–176.

[OJA 03] M. Oja, S. Kaski, and T. Kohonen. "Bibliography of self-organizing map (SOM) papers". addendum Neural Computing Surveys, vol. 3, 1–156.

[PAM 04] E. Pampalk and G. Widmer and A. "Chan A new approach to hierarchical clustering and structuring of data With self-organizing maps". *Intelligent Data Analysis Journal*. 8(2): 131-149.

[TAD 95] C. Tadj. " Méthodes connexionnistes de quantification vectorielle à apprentissage compértitif, Application à la détection de mots clés". Thèse de doctorat de l'Ecole Nationale Supérieure des Télécommunications Spécialité : Signal et Images, Paris.

[VES 97] J. Vesanto and May. "Data Mining Techniques Based on the Self-Organizing Map". Thesis of the degree of Master of Science in Engineering, Helsinki University of Finland 26