Journal of Theoretical and Applied Information Technology
31st August 2015. Vol.78. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A NEW ALGORITHM FOR OPTIMIZING THE SELF-ORGANIZING MAP

BEN-HDECH Adil, GHANOU Youssef, EL QADI Abderrahim
Team TIM, High School of Technology, Moulay Ismail University, Meknes, Morocco
E-mail: adilbenhdech@gmail.com, youssefghanou@yahoo.fr, elqadi_a@yahoo.com

ABSTRACT

The selection of the architecture of Kohonen networks, associated with a given problem, is one of the most important aspects of neural network research; the choice of neurons number and the initial weights has a great impact on the convergence of learning methods trained. In this paper, we present a new method for optimizing the Self-Organizing Map. This approach consists to identify and to select the best map. The goal is to delete unimportant units in the kohonen map. To test our approach, we use two databases Iris and Seeds. The results obtained demonstrated the effectiveness of the proposed approach.

Keywords: *Self-Organizing Map, Optimization, Clustering, Classification, SVM*

1. INTRODUCTION

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. It is composed of a large number of units (neurons) interconnected working in unison to solve specific problems. The Artificial Neural Networks are a very powerful tool to deal with many applications [8][9], and they have proved their effectiveness in several research areas such as speech recognition [4], analysis and image compression [5], recognition of writing signal analysis, process control, robotics, feature selection and research on the Web .

Kohonen networks or Self-Organizing Maps (SOM) are a type of neural network. They were developed in 1982 by [6], which he has introduced the very interesting concept of self-organized topological feature maps, which are maps that preserve the topology of a multidimensional representation within the new one- or two-dimensional array of neurons. The concept of topology has become the essential feature of the Kohonen approach in neural network research. The Self-Organizing Maps is more interesting in many fields such as: pattern recognition, clustering [10], feature selection [11], speech recognition, medical diagnosis, etc

Kohonen network is one of the unsupervised learning model that will classify the units by the similarity of a particular pattern to the area in the same class. The results obtained by the Kohonen networks are acceptable, but the quality of clustering depends on the optimal parameters (numbers of units) of the networks.

One of the most important problems that neural network designers face today is choosing an appropriate neural architecture for a given application. Optimization of neural architectures and training method has an influence to the quality of networks measured by their performance. Finding a solution of the hybrid training and architecture optimization requires a solution of a mixed-integer non-linear optimization problem with a linear constraint. Such problems are NP-complete. The authors [12], [13], [14] used the genetic algorithm approach to select the architecture of Kohonen networks.

The mean propose with this work is to model this choice problem of neural architecture. The objective of our method is: to optimize the self-organizing map in order in order to determine the optimum number of neurons, increase classification rate and minimize the number of iterations. Our approach consists to identify and to select the best map. In the first step, we determine unimportant units and in the second step we delete this unit. We apply this approach for tow problems of classification.

This paper is organized as follows: section 2 presents the Self-organizing map; Section 3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

presents our proposed algorithm for optimizing the SOM. In section 4 we present the results obtained on different data sets and in the last section we conclude with future work .

2. SELF-ORGANIZING MAP (SOM)

Neural networks, in particular the self-organizing Kohonen maps are an important class of neural networks. They were particularly important since the work of Kohonen [1], [2]. A large variety of related algorithms have been derived from the first SOM model, which differs from one to another, but share the same idea that introduces the topological order between the different clusters. The structure of standard SOM is given in Fig.1.

For convenience, let us mention some notations: let N be the number of sample points in the data set, n be the number of features in the original feature set S, M be the size of the map units set U and w_j be the prototype of the j^{th} unit.

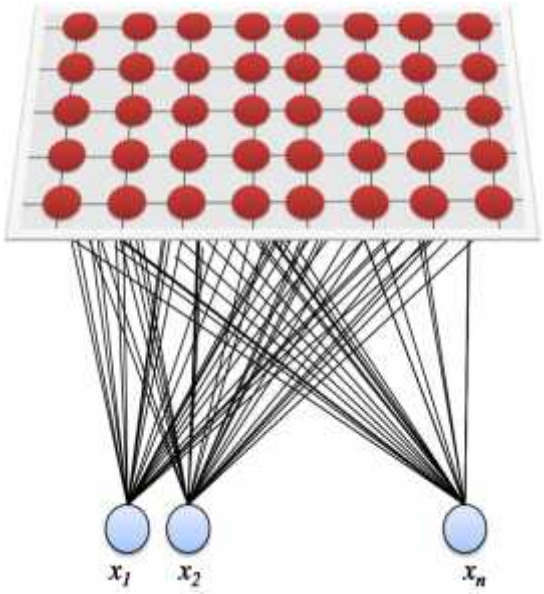

Figure 1: Kohonen Topological Map

1. *Initialization*: Choose random values for the initial weight vectors w_j and time $t = 0$ and Tmax
2. *Repeat* $t < T$
   - *Sampling*: Draw a sample training input vector x from the input space.
   - *Matching*: Find the winning neuron I(x) that has weight vector closest to the input vector:

$$D_j(x) = \arg\min_{j=1}^{M} \sum_{i=0}^{n} (x_i - w_{ij})^2 \quad (1)$$

   - *Updating*: In this phase the algorithm updates the weights of the winning neuron and of the neurons that lie in a user defined neighborhood as follows:

$$w_j(t+1) = w_j(t) + \eta(t) h_{i,j}(t) ||x_i - w_{ij}|| \; i \epsilon V_g(t)$$

   - Continuation: $t = t + 1$.

Where:

- $\eta(t)$: Adaptation coefficient;
- $V_g(t)$: The neighborhood function
- $h_{gi}(t)$: Neighborhood kernel centered on the winner unit:

$$h_{gj}(t) = \exp(-\|r_g - r_i\|/2\sigma^2(t)$$

Where r_g and r_i are positions of neurons g and i in the SOM map. Both $\eta(t)$ and $\sigma(t)$ decrease monotonically with time. There is also a batch version of the algorithm where the adaptation coefficient is not used [7].

Changing vectors associated with units made differently depending on the position of the units with respect to the winning unit. The winner will be the node whose vector undergoes the most changes, while the remote units will be less affected. The neighborhood function $V_g(t)$ will be maximum at $g = 0$ and decreases when r increases (when away from the winner node). A commonly used function is the Gaussian curve.

3. NEW ALGORITHM FOR OPTIMIZING THE SELF-ORGANIZING MAP

SOM has mort parameters such as the neurons, the weights and numbers of Iterations. For optimizing we remove the selected neuron after each iteration, and learning is done again with the remainder of variables. let N be the number of sample points in the data set, n be the number of features in the original feature set S, M be the size of the map units, w_j be the prototype of the j^{th} unit, and Cp denote the maps provided at each step, such as p the number of units in Cp. The principal stages of the algorithm proposed are:

Journal of Theoretical and Applied Information Technology
31st August 2015. Vol.78. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

1. *Initialization :* Initialized maps Cp, p = M
2. *Training :* Learning a map Cp with variables S using SOM algorithm defined above;
3. *Finding:* In this phase the algorithm searches and deletes the neuron j^{th} .
   - $a_j = \frac{c \quad (U_j)}{N}, 1 < j < p,$ *(4)*
   - $j = a \quad m \quad {}^{p}_{j=1}(E_j)\, w \quad \{j/ , a \quad m \quad {}^{p}_{j=1}(a_j)\}$ *(5)*
   - *Deleting the neuron* j^{th} , $p = p - 1$
4. If $E_T > S \quad o \quad (p) > n$ go to step 2;

Where:

- $card(U_j)$: Represents the number of samples assigned to the neuron j^{th} , where U_j is set the samples wining by unit j^{th} .
- E_j: Error for j^{th} unit $E_j = 1/card(U_j) \sum_{i}^{c} {}^{(U_j)} D_j$ (6)
- E_T: The estimated topological error.

4. RESULTS AND DISCUSSION OF EXPERIMENT

For evaluating the performance of our algorithm, we choose two known problems from the UCI repository of machine learning databases [3]: Databases "Iris" and "seeds". We developed the programs in JAVA environment and tested in the Intel (R) Core TM Duo CPU of processor 2.53GHz and RAM of 2Go.

4.1 Preprocessing

To illustrate the advantages of our method, we apply our algorithm to a normalized data set. After normalization, all features of these examples are between zero and one. By doing this, all features have the same contribution to classification. To this end, we use the following normalization:

$$x_{ij}^{k} = \frac{x_{ij}^{k} - \min(x_i)}{\max(x_i) - \min(x_i)} \quad (7)$$

4.2 Performance Evaluation

In this paper, we evaluate the performance of different classification methods using predictive accuracy, which can be defined as:

$$\text{Accuracy} = \frac{N_1 + N_2 + \cdots N_i}{N} \times 100$$

Here, $N_1, N_2, \ldots N_i$ respectively denote the correct classification numbers of the samples belonging to a corresponding class; N represents total sample numbers.

4.3 Experiment results

Let NLB (Number the learning base), CC (Data Correctly Classified) and MC: Data Misclassified.

4.3.1 Database Iris

This database contains 3 classes of 50 samples, where each class refers to a type of iris plant. The samples in this database are characterized with 4 variables. For testing our method, the data set is divided into two parts: the training set and test set. We used half of the samples (75 samples) for training, 75 samples for the test.

We conducted a series of experiments by reducing the size of the map, where we have applied our algorithm for instance of the database Iris. The TABLE I presents the obtained results. We remark that this method give the good results because, all the training data were correctly classified except five. In fact; these elements (misclassified) are one from Setosa class and two from the Virginica and Versicolor class.

Table I: Numerical Resultants Obtained By Our Approach Applied To Database Iris

	NLB	CC	MC	Accuracy (%)
Setosa	25	24	1	96%
Virginica	25	23	2	92%
Versicolor	25	23	2	92%
Total	75	70	5	93,33%

4.3.2 Database Seeds

This database represents measurements of geometrical properties of kernels belonging to three different varieties of wheat. A soft X-ray technique and GRAINS package were used to construct all seven, real-valued attributes. The examined group comprised kernels belonging to three different varieties of wheat: Kama, Rosa and Canadian, 70 elements each, randomly selected for the experiment. The samples in this database are characterized with 7 variables.

Journal of Theoretical and Applied Information Technology
31st August 2015. Vol.78. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

For testing our method, the data set is divided into two parts: the training set and test set. We used 135 samples for training and the rest samples for the test.

We realize a series of experiments by reducing the size of the map, where we have test our methods for instance of the database Seeds. The TABLE II presents the obtained results of training data. We observe the proposed method permits to well classify the training data. The results obtained by our algorithm are good, because most the training data were correctly classified except eight. These elements are three from Rosa, Canadian class and two from the Rama class.

Table Ii: Numerical Resultants Obtained By Our Approach Applied To Database Seeds

	NLB	CC	NCA	Accuracy (%)
Kama	45	43	**2**	**95,55%**
Rosa	45	42	**3**	**93,33%**
Canadian	45	43	**3**	**95,55%**
Total	135	128	**8**	**94,81%**

4.4 Comparisons and discussions

In order to evaluate the performance of the proposed method, we compared our method with another learning algorithm: Support Vector Machine (SVM) for the classification of Iris and Seeds data. The results are presented in TABLE III and TABLE IV.

Table Iii*: Comparisons Of Classifiers Data Iris*

	SOM	SVM	proposed method
NI	2000	5000	120
NCLB	4	3	5
NCTB	4	5	4
AcLS(%)	94.6	96	93.3
AcTS (%)	94,6	93.3	94.6
ER	0.053	0.053	0.060

Table Vi*: Comparisons Of Classifiers Data Seeds*

	SOM	SVM	proposed method
NI	1600	4000	100
NCLB	8	7	8
NCTB	3	4	3
AcLS(%)	94,81	95.5	94,81
AcTS (%)	96,0	93.3	96,0
ER	0.052	0. 052	0. 052

Where:

- NI: Number of iterations
- NCLB: Not classified as learning base.
- NCTB: Not classified as test base.
- AcLS: Accuracy for the learning set.
- AcTS: Accuracy for the test set.
- ER: Error Rate

Our classifier, SOM and SVM classifier have been applied to the same digital database. In order to consider their performance in the most objective manner possible, TABLE III and TABLE IV, show the best qualities of classification methods offered by each classifier model applied respectively a database Iris and seeds. The number of iterations is obtained by several tests

During our experiments, we observe on the basis of results illustrated in TABLE III, TABLE IV, Figure 3 and Figure 4, that our approach produces satisfactory to those obtained by SOM and SVM in terms of execution time.

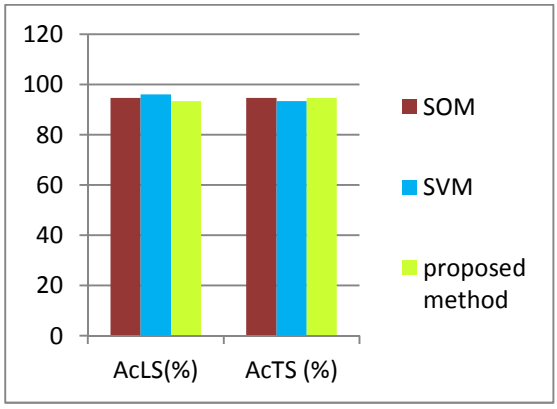

Figure 2. Comparisons Of Classifiers Data Iris

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

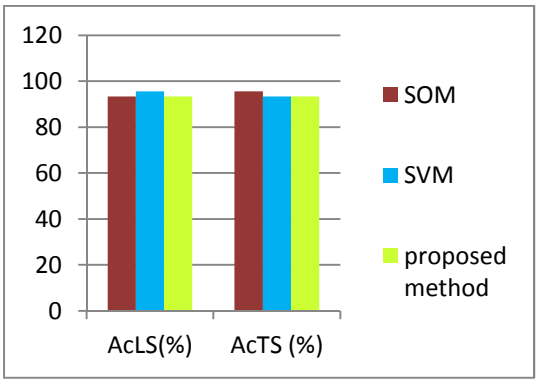

Figure 3. Comparisons Of Classifiers Data Seeds

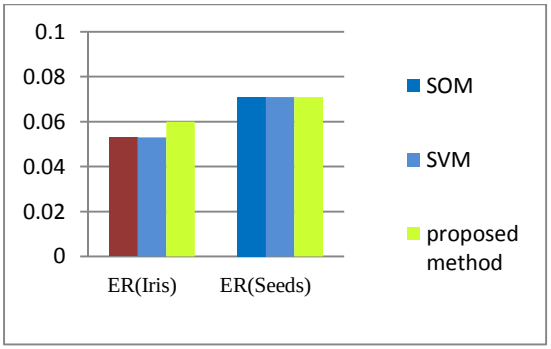

Figure 4. Comparisons Of Error Rate The Data Base Iris And Seeds

The results obtained by comparing our method with SOM and SVM our proposed give a good results with a number less iterations end neurons by report SOM and SVM. For example the database Iris our method has 120 iteration and 13 neurons and the SOM method has 2000 iterations and 36 neurons.

5. CONCLUSION

In this paper we proposed a method for optimizing the Kohonen Self-Organizing Map. This method has been applied two known problems from the UCI repository of machine learning databases. The obtained results demonstrate the performance of this method. We compared the performance of this approach with the SOM and SVM classifier; the results obtained are quite satisfactory.

In the application in classification, the proposed method can classify the data very well with less number of iterations relative to SOM.In future work we will try to adapt our algorithm for the problem of image classification and we will enhance our algorithm.

REFRENCES:

[1] T. Kohonen, Self-Organized Formation of topologically correct features maps, *Biological Cybernetics*, 1988.

[2] T. Kohonen, Self-Organising Maps, *Edition Springer*, second edition 199

[3] Blake, C.L and Merz, C.MJ. "UCI repository of machine learning databases".*Technical report. University of California, Department of information and Computer science, Irvine, CA*, 1998.

[4] M. K. Apalak, M. Yildirim, R Ekici, "Layer optimization for maximum fundamental frequency of laminated composite plates for different edge conditions", *composites science and technology (68)*, 2008, pp. 537-550.

[5] M. Ettaouil, Y. Ghanou, K. El Moutaouakil, M. Lazaar, "Image Medical Compression by A new Architecture Optimization Model for the Kohonen Networks". *International Journal of Computer Theory and Engineering* vol. 3, no. 2, 2011, pp. 204-210.

[6] T. Kohonen, Self-Organized Formation of Topologically Correct Feature Maps, *Biological Cybernetics*, vol. 46, pp. 59–69, 1982..

[7] T.Kohonen,"Self-Organizing Maps". Berlin/Heidelberg, Germany: *Springer*, vol. 30, 1995.

[8] G. Dreyfus, J. Martinez, M. Samuelides, M. B. Gordon, F.Badran, S. Thiria, L. Hrault. "Réseaux de neurones Méthodologie et applications". *Eyrolles 2ième édition*

[9] M. Ayache, M. Khalil and F. Tranquart. "Artificial Neural Network for Transfer Function Placental Development: DCT and DWT Approach". *IJCSI International Journal of Computer Science Issues*, Vol. 8, Issue 5, No 3, September 2011

[10] J. Vesanto, E. Alhonierni "Clustering of the self-organizing map" *IEEE Trans. Neural Networks* 11 (3) (2000) 586–600.

[11] K. Benabdeslem and M. Lebbah, "*Feature selection for self organizing map, " in International Conference on Information Technology Interface-ITI* 2007, June 25-28, 2007, Cavtat-Dubrovnik, Croatia, 2007, pp. 45-50.

[12] Ettaouil M. and Ghanou Y. 2009. Neural architectures optimization and Genetic algorithms. *Wseas Transactions On Computer* 8 (3). 526-537.

[13] Ettaouil M., Ghanou Y, Elmoutaouakil K., Lazaar M. 2011. *A New Architecture*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Optimization Model for the Kohonen Networks and Clustering. Journal of Advanced Research in Computer Science (JARCS), 3 (1), 14 - 32.

[14] M. Ettaouil, Y. Ghanou, K. El Moutaouakil and M. Lazaar, "*Image Medical Compression by A new Architecture Optimization Model for the Kohonen Networks"*. International Journal of Computer Theory and Engineering, Vol. 3, No. 2, April 2011