Anadolu Üniversitesi Bilim ve Teknoloji Dergisi A- Uygulamalı Bilimler ve Mühendislik
Anadolu University Journal of Science and Technology A- Applied Sciences and Engineering

2018 - Volume: 19 Number: 1
Page: 144 - 152
DOI: 10.18038/aubtda.363606
Received: 07 December 2017 Revised: 02 January 2018 Accepted: 04 January 2018

A NOVEL APPROACH FOR LEARNING RATE IN SELF ORGINIZING MAP (SOM)

Emin GERMEN *

Electrical & Electronics Engineering Department, Engineering Faculty, Anadolu University, Eskişehir, Turkey

ABSTRACT

The performance of resultant topological structure of Kohonen Self Organizing Map SOM is highly dependent of the learning rate and neighborhood parameters. In literature there are plenty many different types of approaches to and proposals for those parameters. It has been investigated that in general the learning rate and neighborhood parameters are data independent and predefined before the training period. Here in this paper a novel approach has been proposed to change the learning rate parameter according to the interaction of neurons with data. During training, the worst matching neuron also tracked and used to trace the formation of topological structure of SOM. A slight modification on conventional learning rate with proposed method has a considerable influence on resultant topologies in a positive way. The effects of this approach have been tested with the real world problem and different synthetic data.

Keywords: Self organizing map, Kohonen, Learning rate, Classification

1. INTRODUCTION

Nowadays the classification and clustering algorithms have gained momentum in the shadow of deep learning and machine learning, and the popularity of neural networks has come to light one more time during its history. The Kohonen's proposed neural network model Self Organizing Map (SOM) [1] [2] is used in order to project high dimensional input data space to the two dimensional lattice in order to visualize the possible relationship between data. The visualization process is accomplished by lateral interactions of neuron reside on a planar surface. The unsupervised nature of training SOM is used to cluster the data on the other hand. The inspiration of SOM truly depends on the structural organization of the brain. It is observed that different parts of cortical map evolve to satisfy different types requirements such as somatosensory region, auditory region, visual map etc.

As the SOM performs both clustering and vector quantization, the resultant topology formation highly depends on the parameters of the learning algorithm. During the evaluation of theory, different types of learning rate and neighborhood functions have been proposed where those are the crucial parameter for the clustering and vector quantization and related with the formation of neurons in lattice structure. In [3] it has been introduced a statistical method in order to improve the convergence. In [4] both a new learning rate and neighborhood function has been introduced to track the variations in statistical nature of data. Similarly in [5] a new methodology has been introduced to effect the change of weights of a neuron according to the past history. The influence of learning rate and neighborhood functions has been scrutinized in [6].

The idea which is presented in this paper in order to improve the resultant quality of SOM depends on tracking the topographic organization of SOM during training process. The two basic quality measures of topology formation at the end of training process are quantization error and topographic error of the topologies [7]. During training, the formation of topology has tracked by checking the position of worst matching neuron (WMN) and according to its position, the amount of decrease in the learning rate has been changed. The tracking of WMN and using this in learning rate produces no extra computational

*Corresponding Author: egermen@anadolu.edu.tr

complexity however has considerable positive impact on the quality of the resultant mesh structure. This new proposed method has been applied to different kind of learning rate functions and for each learning rate function, considerable amount of improvement has been achieved.

In this paper, the SOM and its common learning rate and neighborhood functions has been addressed in Section 2. In section 3 the new proposed method has been explained. The simulation result of the proposal has been examined in section 4. Also in this section the two different quality estimation methods are also described. The paper is concluded in section 5.

2. SELF ORGINIZING MAP (SOM)

SOM composed of two dimensional organized artificial neurons of having a weight vector with the same dimension of input vectors. The unsupervised training algorithm depends on competitive process between neurons in order to find the most similar distance between the input vector and the weights of each neuron as shown in Figure 1.

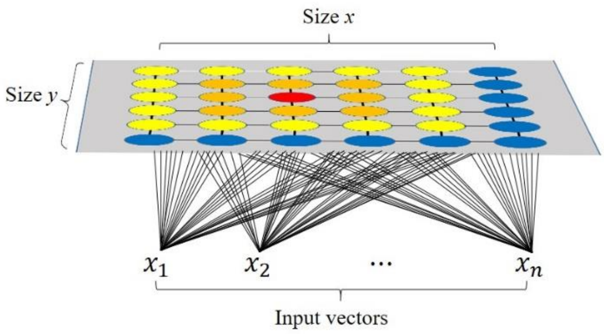

Figure 1. The SOM topological map. The neurons on the plane are connected in square lattice structure. After finding the Best Matching Neuron (BMN) the neurons around this BMN form the neighborhood in lattice structure.

The training algorithm depends on updating weights according to the presentation of each input vector. In the training process the distance (in general Euclidian distance) between input vectors $x(k)$ and the weight vectors $w_i(k)$ of each neuron is calculated as

$$d_i(k) = \|x(k) - w_i(k)\| \qquad i = 1,2,\cdots m \tag{1}$$

where m is the number of neurons in the lattice.

In the algorithm the crucial point is finding the best matching neuron (BMN) w_c using the distance measure given in (1) as :

$$w_c(k) = \arg\min_i [d_i(k)] \quad i = 1,2,\cdots m \tag{2}$$

All the neurons in the lattice are updated iteratively according to the equation (3)

$$w_i(k+1) = w_i(k) + \alpha(k) \cdot \beta(i,c,k) \cdot (x(k) - w_i(k)) \tag{3}$$

where $\alpha(k)$ is the learning rate parameter and the $\beta(i,c,k)$ is the neighborhood function that depends on iteration time k. If there is not a planar neighborhood relationship of neurons with each other, the algorithm can be considered as a vector quantization algorithm. However, the neighborhood function gives extra power to the model in order to estimate the relationships of the given input vectors. The learning rate is scalar value which can be considered as adaptation rate changing according to the iteration time and the neighborhood function determines the adaptation strength of the neurons around the BMN indexed by c. This functions defines the stiffness of the elastic surface around the winning neuron. Both the learning rate and neighborhood function have no explicit definition in literature, however, they have to be decreased by iteration time.

The most widely used learning rate rates are given as:

- Exponential learning rate [2]:

$$\alpha(k) = \alpha_{initial}\left(\frac{\alpha_{final}}{\alpha_{initial}}\right)^{\frac{k}{k_{max}}} \tag{4}$$

In (4) $\alpha_{initial}$ and α_{final} values are determined empirically which are truly problem dependent and k_{max} is the maximum iteration time set before the training process.

- Linear learning rate [1]:

$$\alpha(k) = \alpha_{initial} \cdot \frac{k_{max} - k}{k_{max}} \tag{5}$$

- Inverse learning rate [1]:

$$\alpha(k) = \frac{C \cdot \alpha_{initial}}{C + k} \tag{6}$$

where C is a problem dependent constant.

The most common neighborhood function described as :

$$\beta(i, c, k) = e^{-\left(\frac{{d_{i,c(k)}}^2}{R(k)^2}\right)} \tag{7}$$

where

$$R(k) = 1 + (R_{initial} - 1) \cdot \frac{k_{max} - k}{k_{max}} \tag{8}$$

and

$$d_{i,c}(k) = \left\| w_{i(topology)}(k) - w_{c(topology)}(k) \right\| \tag{9}$$

The neighborhood function is decreasing function around the BMN and wide in the beginning of training phase, however shrinks during training.

3. MODIFICATION OF LEARNING RATE

The main drawback of SOM is predefined learning rate and the amount of iteration which has to be chosen empirically without concerning the data. By scrutinizing the training process, it can be deduced that the whole period can be divided into two different phases if the rapidly decreasing learning rate functions are used such as exponential learning rate (4) or inverse one (6). In the first phase, relatively large initial learning rate values are used such as $\alpha = 1$ or $\alpha = 0.9$ in where the global ordering of neurons occurs. Progressing of training cause a sharp rapid decrease in learning rate (eg. $\alpha = 0.2$ or $\alpha =$

0.1) and the neuron movements getting slower. This phase is called fine tuning. The experiments showed that, this two phase kind of training scheme improves the quality of the resulting topology. This fact is studied in [8]. However, on the other hand, it has been observed that the effects of early presentations with large neighborhood are almost forgotten in the period of fine tuning. The very striking result of the research reveals that only the 20 percent of input data has an effect on the formation of final topology.

In order to eliminate the disadvantage of fast decreasing learning rate, batch version of the algorithm has been proposed [9]. In [10] it has been proposed random sample of batch processing instead of using the whole data set. Thus it is not difficult to guess that the order of presentation has a major influence on the final topology quality.

In this paper in order to associate the data with learning rate, a new approach has been proposed. During training period, either in global ordering or fine tuning phase, the data presented and organization of neurons should have to be scrutinized and according to the possible ordering or disordering, the learning rate has to be changed. In order to track the formation of neurons, during determination of BMN, also the Worst Matching Neuron (WMN) is found as:

$$w_z(k) = \arg \max_i [d_i(k)] \quad i = 1,2,\cdots m \tag{10}$$

where z is associated with the WMN index.

The WMN is also used to track the quality of the formation of topology. The main idea behind the new learning rate is checking the relation of WMN and the BMN. If the distance between them is not the possible maximum value from the point of view of topological ordering, it represents the disordered topology formation. In order to embed this phenomenon into the training sequence, the reduction in learning rate is sustained and the learning phase kept in rough training phase if there are problems in the topological structure. For the new learning rate the algorithm is defined as:

Algorithm for the new learning rate:

i. For each data point *x(k)* determine the WMU which is the furthest element in the map.
ii. Check the placement of the WMN on the planar surface.
iii. If WMN is the furthest element to the BMN on the surface decrement the learning rate as proposed in the algorithm.
iv. Otherwise, don't update the learning rate and use previous iteration time value instead.

The new proposed method can easily be adapted to all kinds of learning rates and at the same time will prevent the instant decrease of the learning rate if there are disorders in the topological structure. Especially in the rough training phase the constant decrease in the learning rate causes improper ordering according to the statistical nature of data if stochastic training method is applied instead of batch training. The proposed method has an effective role on solving this problem. Also the determination of WMN will not introduce an extra complexity to the algorithm because during determination of BMN, WMN can easily be calculated.

4. EXPERIMENTAL RESULTS

The introduced modification in learning rates has a considerable improvement in the final topologies. In order to analyze the related contribution, two different types of quality measurement methods are used. The Quantization Error (QE) is the mostly appreciated and widely accepted quality measurement method which is used in all types of quantization and the clustering algorithms. It is computed by measuring the average distance between the neurons and sample vectors represented by those. The properly placed centroids cause reduction in QE. Since the true nature of SOM also depends on the formation of topographically connected neurons, also it is necessary to check the topographic structure of the final formation. The Topographic Error (TE) is used check the topology preservation. The sample vectors are used to determine this quality measure. Here for each data sample, the best and the second-best neurons are determined respectively. If those are in close vicinity from the point of topographic orientation, it is tracked as OK. However, if those are not adjacent neurons, it is regarded as error. The total error is calculated and normalized between 0 and 1 where 0 is considered as perfect topology preservation.

In the experiments, two dimensional empiric data is used to trace the quality of topographic formation of SOM using two different learning rates. Also in order to test possible contribution, the real IRIS data set is tested with exponential learning rate. Both the QE and TE are calculated using the same amount of training data and results are shown in Table 1 and Table 2. The experiments are:

Experiment 1: In this experiment two dimensional 10 x 10 neuron map is trained with two dimensional uniformly spaced data set between 0 and 1 in order to visualize the results. Here the exponential learning rate is used. In order to track the effects of algorithm, the same initialized SOM has been trained using untouched and updated learning rate. The neurons are connected in hexagonal lattice. The resultant topologies are shown in Figure 2. In Figure 3 the effects of the new parameter on the learning rate can easily be traced. During training if there is a topographic misplacement, the learning rate doesn't update and remains the same, however after true settlement of the neurons, the learning rate decreases. The same phenomenon also can be seen in Figure 5 for Inverse learning rate.

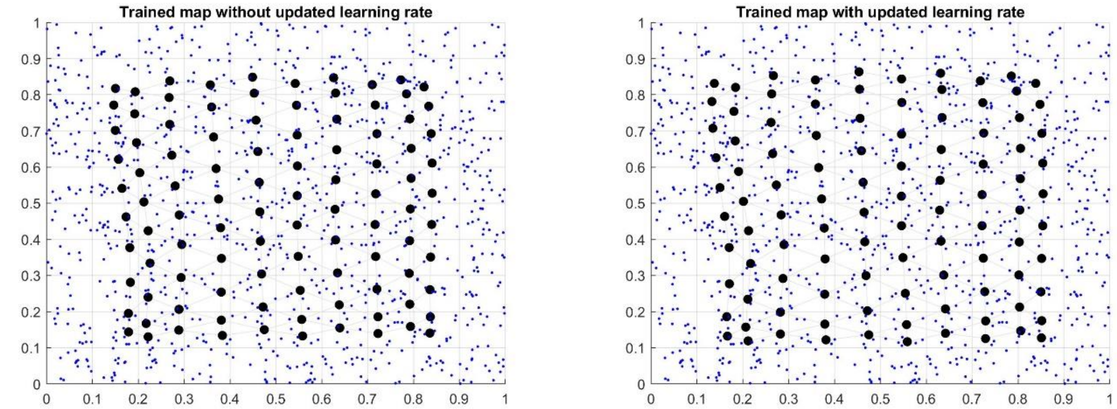

Figure 2. The comparison of topological formations of 10x10 neuron with exponential learning rate

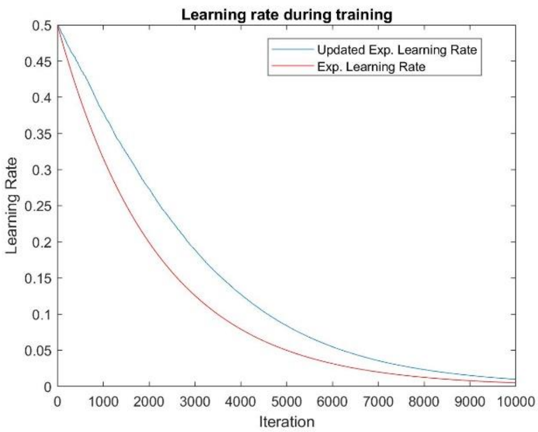

Figure 3. The change of exponential learning rate during training

Experiment 2: In this experiment two dimensional 10 x 10 neuron map is trained with two dimensional uniformly spaced data set between 0 and 1 in order to visualize the results. Here the inverse learning rate is used. In order to track the effects of algorithm, the same initialized SOM has been trained using untouched and updated learning rate. The neurons are connected in hexagonal lattice. The resultant topologies are shown in Figure 4.

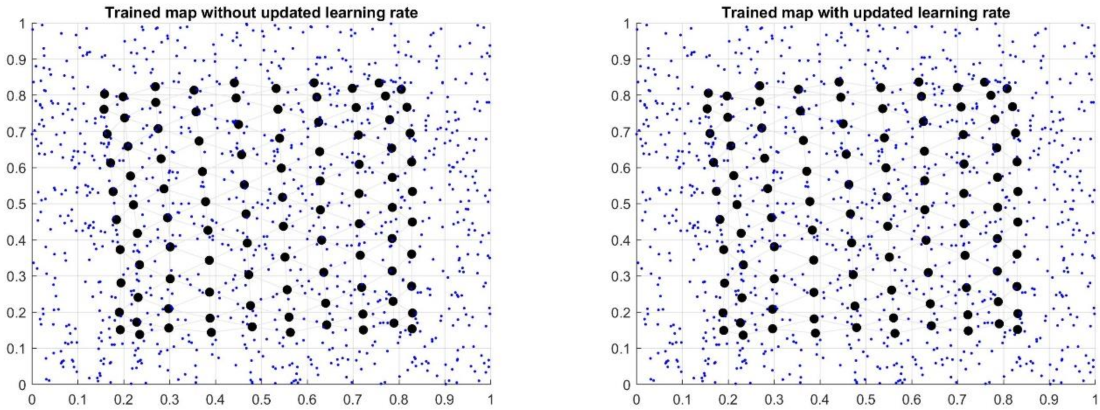

Figure 4. The comparison of topological formations of 10x10 neuron with inverse learning rate

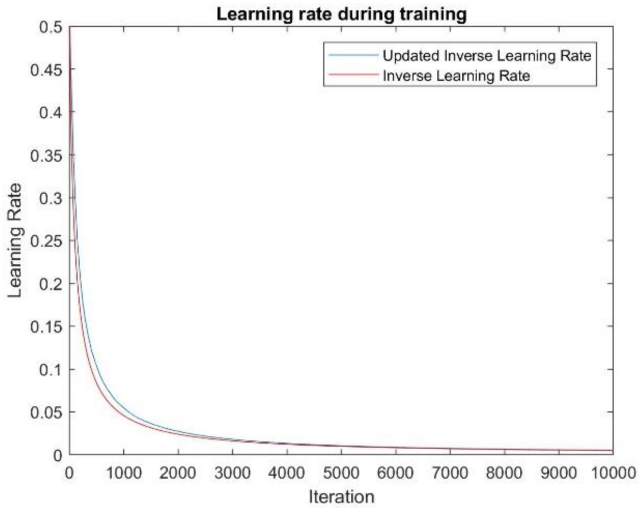

Figure 5. The change of inverse learning rate during training

Experiment 3: In this experiment two dimensional 10 x 10 neuron map is trained with two dimensional donut shaped data. The data set consists of 1000 samples uniformly distributed between radius 3 and radius 7. The exponential learning rate is used and the same data set is used to train the SOMs which is initialized with same kind of structure in where the neurons are connected in hexagonal lattice. The results are shown in Figure 6.

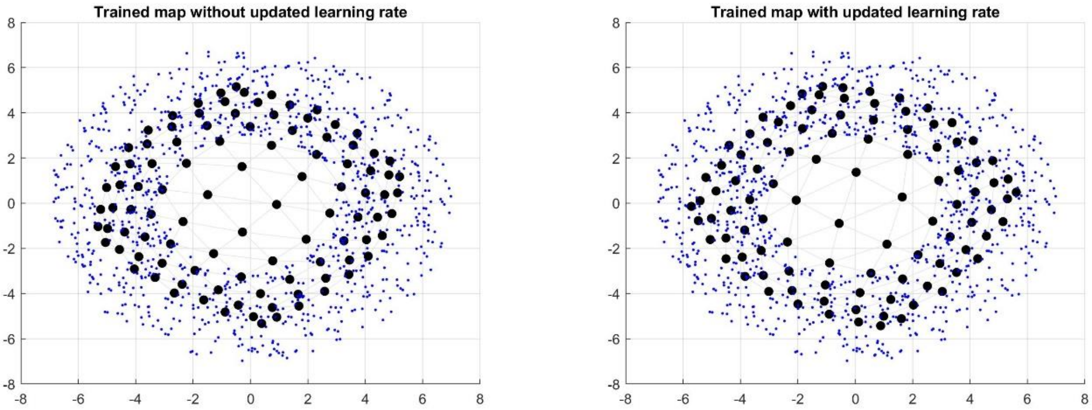

Figure 6. The comparison of topological formations of 10x10 neurons trained by donut data with exponential learning rate

Experiment 4: In this experiment in order to check the effect of modified learning rate parameter, the Iris data set is used. The data set is obtained from UCI Machine Learning Repository. [11] This set is very common in clustering area. The data set consists of 3 different species called "Setosa", "Virginica" and "Versicolor" where each has 50 instances of four dimensional data. Here the SOM with 10x10 neurons has been trained with exponential learning rate. The same experiment exactly repeated by modifying the learning rate parameters and both quantization and topographic errors are calculated. The resultant maps are shown in Figure 7. The classification of Setosa, Virginica and Versicolor flowers can be seen clearly. The size of the colors on the map which are actually called "hit histograms" define the amount of hit terms of the species. Those are found by calculating the amount of BMN's of each data sample from the map. The increasing size denotes the amount of data samples represented by the neuron. Here it is worth to note that some neurons can be hit with different species which means that the classification for those data is problematic. In Figure 7, "trained map without updated learning rate" the sizes of mixed up neurons (separation of Virginica and Versicolor is problematic) are greater than the "map with updated learning rate" which denotes a slight improvement in the classification of Virginica and Versicolor species.

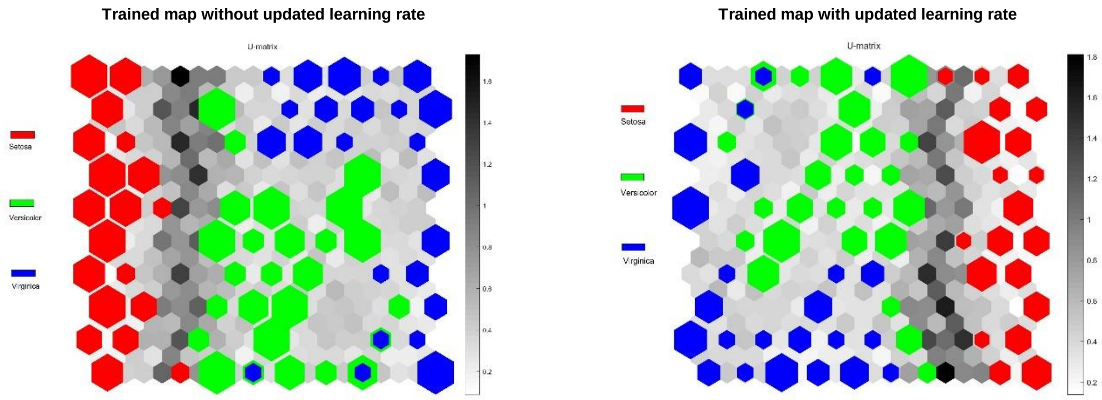

Figure 7. The comparison of topological formations of 10x10 neurons trained by iris data with exponential learning rate

The quantization error and topographic error values for all the experiments can be seen in Table 1 and Table 2.

Table 1. The Quantization Error values for the experiments

Experiment Number	Quantization Error without updating learning rate	Quantization Error with updating learning rate
Exp 1	0.0765	0.0582
Exp 2	0.0700	0.0680
Exp 3	0.7344	0.6690
Exp 4	0.3565	0.3370

Table 2. The Topographic Error values for the experiments

Experiment Number	Topographic Error without updating learning rate	Topographic Error with updating learning rate
Exp 1	0.0167	0.0
Exp 2	0.003	0.003
Exp 3	0.005	0.005
Exp 4	0.02	0.0067

By applying the new learning rate parameter, for each experiment, there is a considerable decrease in Quantization Error which points that keeping data into account to determine learning rate improves the quality of final settlement of the neurons. From the point of Topographic Error there still is a slight improvement on results.

5. CONCLUSION AND FEATURE WORK

Choosing proper learning rate for Kohonen SOM is always being a bit of state of art for data scientists. This paper has denoted how an improvement can be achieved by slight change in learning rate considering the influence of data on the neurons during training period. There must be a close relation between the formation of the network's topological structure and characteristic of data. SOM always suffers predefined learning rate and neighborhood functions which are not related to data itself. Here in this paper a data relevant learning rate modification has been issued and the results of the novel approach have been tested with different kinds of data and different learning rate parameters.

The analyses were performed on experimental two dimensional data and dataset IRIS. Those show that; a small but in-place change in learning rate has a very positive impact on the outcome without increasing complexity. The same idea can easily be applicable for neighborhood functions. Those preliminary results are encouraging and it has been thought that when the learning rate and neighborhood function are modified according to the data statistic, much better results will be obtained and the resulting topology can be taken a form which reflects the data in much better way.

REFERENCES

[1] Kohonen T, Self Organizing Map, Berlin: Springer Verlag, 1995.

[2] Kohonen T, The Self Organizing Map, Proceedings of IEEE, pp. 1464-1480, 1990.

[3] Mulier F, Cherkassky VS. Statistical analysis of self-organization,» Neural Networks, cilt 8, no. 5, pp. 717-727, 1995.

[4] Germen E, Bilgen S, Statistical Approaches to Determine the Neighborhood Function and Learning Rate in Self Organizing Map, Proceedings of the 1997 International Confererence on Neural Unfırmation Processing and Intelligent Information Systems, Dunedin, New Zealand, 1997.

[5] Germen E, Improving the Resultant Quality of Kohonen's Self Organizing Map Using Stiffness Factor, Wang L, Chen K, Ong YS. (eds) Advances in Natural Computation. ICNC 2005. Lecture Notes in Computer Science, vol 3610. Springer, Berlin, Heidelberg, 2005.

[6] Stefanovič P, Kurasova O. Influence of Learning Rates and Neighboring Functions on Self-Organizing Maps, Laaksonen J, Honkela T. (eds) Advances in Self-Organizing Maps. WSOM 2011. Lecture Notes in Computer Science, vol 6731. Springer, Berlin, Heidelberg, 2011.

[7] Villmann T, Der R, Hermann M. Martinetz TM. Topology preservation in self-organizing feature maps: exact definition and measurement, IEEE Transactions on Neural Networks, cilt 8, no. 2, pp. 256-266, 1997.

[8] Mulier F, Ph.D Thesis, University of Minnesota, 1994 .

[9] Lawrence RD, Almasi GS, Rushmejer HF. A scalable parallel algorithm for Self organizing maps wwith application to sparce data mining problems, Data Mining and Knowledge Discovery, cilt 3, no. 171, pp. 171-195, 1999.

[10] Golli A. Speeding up the self organizing map for dissimilarity data, Proceedings of International Symposium on Applied Stochastic Models and Data Analysis, Brest, France, 2005.

[11] UCI Machine Learning Repository, https://archive.ics.uci.edu/ml/datasets/iris.