

Efficient Neighborhood Function and Learning Rate of Self-Organizing Map (SOM) for Cell Towers Traffic Clustering

Haider Kadhim Hoomod

Computer Science Dept.

Al-Mustansiriyah University

drhjnew@gmail.com

Intisar Al-Mejibli

Informatics Systems Management Dept.

University of Information

Technology and communication

dr.intisar.almejibli@gmail.com

Abbas Issa Jabboory

Informatics Institute for

Postgraduate Studies

Iraqi Commission for

Computer & Informatics

it_abbas@yahoo.com

Recived : 8\10\2017

Revised : 23\10\2017

Accepted : 30 \10\2017

Abstract:

The self-organizing map (SOM) neural network is based on unsupervised learning, and has found variety of applications. It is necessary to adjust the SOM parameters before starting learning process to ensure the best results. In this research, three types of data represent high and low traffic of specific cell tower with subscriber positions distribution in central of Iraq are investigated by self-organizing map (SOM). SOM functions and parameters influence its final results. Hence, several iteration of experiments are performed to test and analyze Bubble, Gaussian and Catgass neighborhood functions with three learning rates (linear, inverse of time and power series) and they were evaluated based on the quantization error. The experiments results show that Bubble function with linear learning rate gives the best result for clustering cell tower traffic.

Keyword: SOM, Traffic. Cell Tower, Learning rate, Neighborhood functions.

1. Introduction

The increase in the number of mobile phone users with services demand which provided by mobile operators in urban and rural areas requires network specialists to expand their coverage to all locations in the wireless system [1]. Telephone traffic is one of the most important factors in the planning of wireless networks. Each cellular network needs a cell planning with a specific final target to give sufficient scope and call quality. The good cell planning process must be continuously updated depending on the mobile network scenario to ensure good quality of service and coverage [2, 3]. In addition to the telephone traffic, coverage, topography, propagation features and system capacity are important factors in the design of the cellular network [4].

The self-organizing map (SOM) neural network is the most popular network based on unsupervised learning, as it does not need for human intervention during the learning process and it is only needs for a little information about the characteristics of the input data [5]. It has proven to be an effective clustering method in many applications as it is widely used in various applications. It describes the formation of a representative sample in multi-layer perceptron training. Additionally, it is effective in pattern recognition systems and future prediction in

Satellite Imagery systems [6, 7]. For example in [1] SOM used for cell towers distribution optimization to find optimal number of cell towers to cover the specified area.

This research aims to find the most appropriate neighborhood function with learning rate for three type of data represent cell towers traffic and subscriber position distribution in capital city of Iraq. Three neighborhood functions (Bubble, Gaussian and Catgass) and three learning rates (Inverse of time, Linear and Power series) are applied to analyze the quality of SOM by computing quantization error[8, 9].

The rest of this paper is organized as follows: The section II describes the data collection .SOM and its parameters are presented in section III. Section IV presents the proposed experiment design. In section V, simulation and obtained result are analyzed. Last section discusses the conclusion.

2. Applied Data

Three types of data were collected to be applied in these experiments; first data type represents high traffic of cell tower in rural environment within the borders of Baghdad. Second data type represents low traffic of cell tower in rural environment within the center of Baghdad. Third data type represents location coordinates of wireless telephone subscribers for homes, offices and medical clinics[11, 12]. This research is provided with data of 11 days 24 hours from ASIA CELL Telecom [10].

3. Self-Organization Map (SOM)

This neural network, or called Kohonen is a popular neural network, which is based on unsupervised learning. The structure of SOM is a single feedforward, it has a set of input (x_1, x_2, \dots, x_n) elements and a set of output unit (w_1, w_2, \dots, w_n) where each input element is connected to all output neurons [13]. SOM network uses competition concept that tries to find the most similar distance between the input vector (X_i) and neuron with weight vector (W_i), the architecture as shown in figure (1).

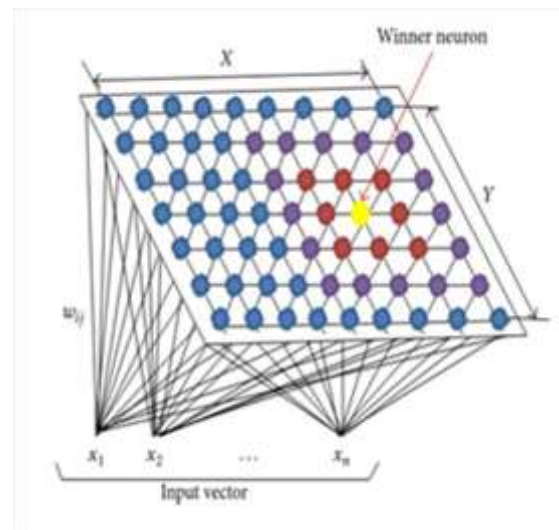


Figure 1: SOM Network Structure output space (two-dimensional) and input space (one-dimension).

The learning process includes the following steps:

- 1- Define and initialize number of parameters, which are network size, learning iteration (t), radius of neighborhood (d0), learning rate, and total number of iterations (T). At this stage, the user experience plays a major role in defining these parameters to achieve the best results.
- 2- Initialize all node's weight (wij) with small random values in the same dimensionality as the training patterns.
- 3- Find the winner node or called the best matching unit BMU by computing Euclidian distance between of them[7, 8].

$$\text{Dist} = \sqrt{\sum_{i=1}^{i=n} (X_i - W_i)^2}$$

(1)

Where X is the current input data vector and W is the weight of nodes.

- 4- Update all the unit weights depending on the distance in the output space between each unit and the BMU, the distance in the input space between the unit and the training pattern [1].

The following formulas are used to update weight vectors and radius of both the winner neuron j and its neighbors as [1]:

$$\sigma(t) = \sigma_0 \exp(-t/\lambda) \quad (2)$$

$$W_{(t+1)} = W_{(t)} + L_{(t)} (V_{(t)} - W_{(t)}) \quad (3)$$

Where σ_0 represents the width of the grid at time t_0 and the λ denotes a time constant, t is the current time-step. L is a small variable called the learning rate, L and $\sigma(t)$ are exponentially decreases with time .

Both the learning rate and the neighborhood radius should converge to zero to reach stable solutions in SOM. Further, both parameters must be updating after presenting each individual data pattern to the network (iteration).

3.1 SOM Parameters

Before the SOM training process, many factors related to SOM structure that affect the result must be define. These include: the size, topology, and shape of the used map. In addition to, the training parameters (number of iterations, initial learning rate, learning function, initial neighborhood radius and neighborhood function) [9].

3.1.1 Size and Dimension of the Map

The size of the problem at hand determines the SOM size and dimension, which makes this definition mainly an empirical process. SOM's size must also take into consideration the size of the dataset of training patterns and the number of units should be equal to the number of expected clusters. Therefore each cluster should be represented by a single unit. Each unit is a cluster centroid.

The SOM output space dimensionality reflects the number of units used in each dimension (x, y, z, etc.). For instance in a two-dimensional SOM, using x equal to 10, y equal to 10 produces a network with 100 units[9].

3.1.2 Topology, shape and initialization

SOM uses two basic types of topology square and hexagonal. In square topology, each unit is connected to its four neighbors, while in hexagonal topology each unit is connected to its six neighbors. Figure (2) shows the two type of topology where the dark gray units represent neighbors of the black units.

Unit vector must be initialized before the training phase , proper initialization allows SOM converges faster to a good solution[9].

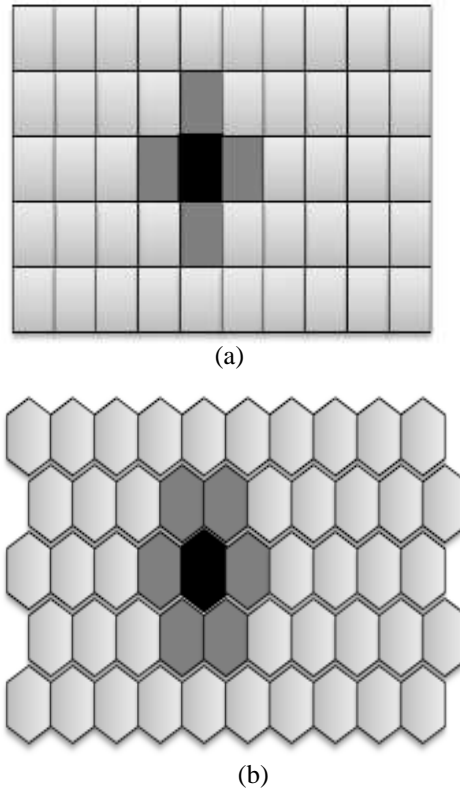


Figure 2: SOM topologies a) Square topology with four neighbors b)

3.1.3 Number of iterations

The number of iteration (usually called epochs) is important parameter that defines how long training phase in SOM. One epoch is achieved when all the training patterns were presented once to the network. There is no rule about selecting optimal number of iterations to use. However the training phase of the SOM should be large enough to fit with the dataset to reach a good solutions [8].

3.1.4 Learning rate and learning functions

Learning rate is the initial value given by the user which is usually between (0, 1), and gradually decreases during the training phase. Several different functions can be used to control the decreasing of learning rate. Linear, power of series and inverse of time (INV) are major learning rate functions[8, 14] as shown in figure (3).

$$\text{Linear } \alpha(t) = \alpha_0 \left(1 - \frac{t}{T}\right) \quad (4)$$

$$\text{Power } \alpha(t) = \alpha_0 \left(\frac{0.005}{\alpha_0}\right)^{\frac{t}{T}} \quad (5)$$

$$\text{INV } \alpha(t) = \frac{\alpha_0}{\left(1 + \frac{100t}{T}\right)} \quad (6)$$

Where T is the training length and α_0 is the initial learning rate.

Different decrease functions of learning rate will influence the network mobility in adapting to the input patterns.

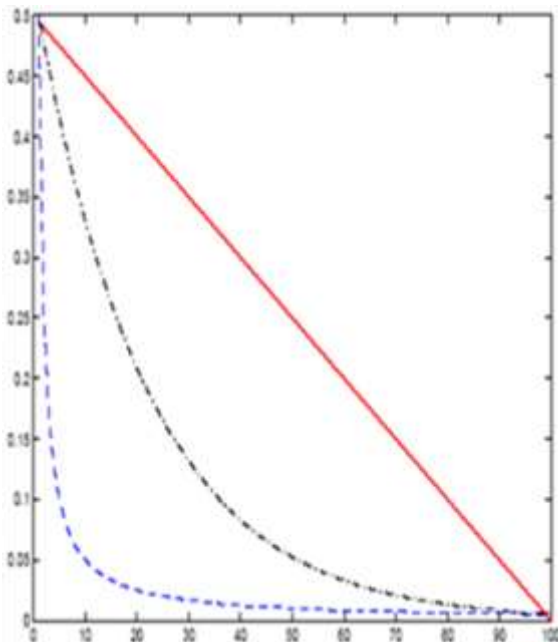


Figure 3: Learning rate functions.

3.1.5 Neighborhood radius and neighborhood functions

Neighborhood function (h) is a value between (0, 1), and it is a position function of two units (BMU_t and another unit) with a given radius r, it usually decreases with time. BMU h has a high value and decreases with distance increase. Neighborhood radius (r) can have value between 0 (when only BMU updated) and maximum size of the network (when all units will be updated).

There are several neighborhood functions, such as Bubble, Gaussian and Cutgass. Figure (4) shows the three neighborhood functions (Bubble, Gaussian and Cutgass respectively) in one and two dimensions [9].

From the left:

$$\text{Bubble} \quad h_{ci}(t) = 1(\sigma_t - d_{ci}) \quad (7)$$

$$\text{Gaussian} \quad h_{ci}(t) = e^{\frac{-d_{ci}^2}{2\sigma_t^2}} \quad (8)$$

$$\text{Cutgass} \quad h_{ci}(t) = e^{\frac{-d_{ci}^2}{2\sigma_t^2}} 1(\sigma_t - d_{ci}) \quad (9)$$

Where σ_t is the neighborhood radius at time t, and $d_{ci} = \|r_c - r_i\|$ is the distance between units c and i in the output space.

The neighborhood function and the number of units effect the accuracy of the resulting mapping. High values of the neighborhood function and high neighborhood radius make maps more rigid [8, 14].

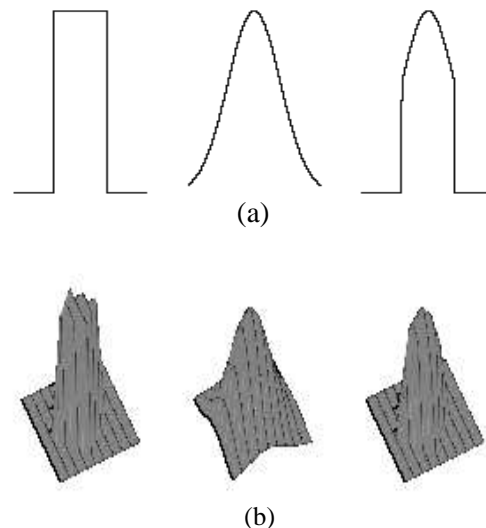


Figure 4: The three neighborhood functions: Bubble, Gaussian and Cutgass

a) Neighborhood function using a one-dimension output space

b) Neighborhood function using a two-dimension output space and a neighborhood radius $\sigma_t = 2$

3.2 Quality of the SOM

The important factor in the SOM analysis process is quality assessment. The major type of quality measurement is quantization error. It is given by the average distance between the unit and training input patterns mapped to it, *i.e.*, all the input data patterns that share it as BMU [8, 9].

$$qe = \frac{\sum_{k=1}^N \|x_k - w_{BMU}\|}{N} \quad (10)$$

Where x_k is the training pattern, w_{BMU} is the best-matching unit for the training pattern, and N is the number of existing training patterns. The SOM with low quantization error is more accurate than a SOM with higher quantization error [8, 9].

4. Proposed Experiment Design

Three types of data used as the input vector for SOM. First type represents high traffic of cell tower with matrix of (11 days x 24 hours) and eleven neuron represent the output of SOM. Second type represents low traffic of cell tower with matrix of (11 days x 24 hours) and eleven neuron represent the output of SOM. Third type represents subscriber coordinate to cell tower distributed problem with matrix of (2 x 344) and four neuron represent the output of SOM. Twenty-seven experiments are applied to find the appropriate parameters of SOM algorithm. The experiments applied using Bubble, Gaussian and Catgass functions with three learning rates (linear, inverse of time and power series). To find the appropriate parameter for all aforementioned cases, the SOM algorithm is used as shown in figure (5).

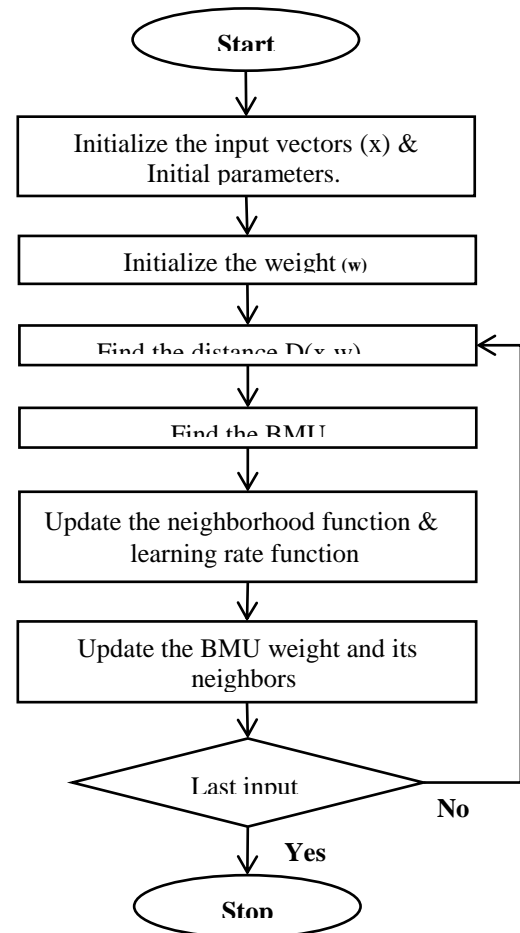


Figure 5: Flowchart of proposed model

Following steps represent the used SOM algorithm:
 SOM algorithm steps:

Step 1: Initialize the input vectors in form of matrix with the size of three cases, and initial SOM parameters.

Step 2: Initialize the weight of three cases by small random value.

Step 3: Compute the distance for all input data and weight by using (equation 1).

Step 4: Find the BMU.

Step 5: Update the neighborhood function using neighborhood equations (7,8,9) and learning rate function by using learning rate equations (4,5,6), to use on update the BMU weight and his neighbors by using (3).

Step 6: Repeat steps 3-5 for all input data.

Step 7: Compute the quantization error using (10).

5. Simulation and Result

The research is implemented using simulator (MATLAB 2016) because of its ease of interface. The initial parameters of SOM are shown in table 1.

The table 2, table 3, table 4, table 5 and figure (6) shows the results of twenty seven experiments with three neighborhood functions and three learning rates. These experiments applied to cluster three types of data that represent high and low traffic of cell towers with subscriber coordinate distribution.

Table 1: SOM Parameters

Parameters	Value
Initial learning rate	0.06
Initial weight vector	random
Initial radius	5
Learning iteration	1-1000

Table 2: Quantization Error Average of High Traffic in Center of Baghdad

Learning rate	Bubble	Gaussian	Cutgass
Linear	0.9114	2.3687	3.5461
Inverse of time	0.8849	3.0695	3.3010
Power series	0.9050	3.0003	3.3886

Table 3: Quantization Error Average of Low Traffic in Border of Baghdad

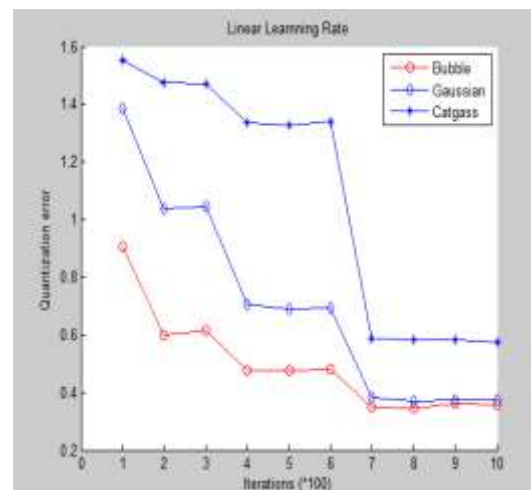
Learning rate	Bubble	Gaussian	Cutgass
Linear	0.5154	0.7539	1.2547
Inverse of time	0.5213	0.8065	1.3690
Power series	0.5201	0.7607	1.2623

Table 4: Quantization Error Average of Subscriber Coordinate Distribution.

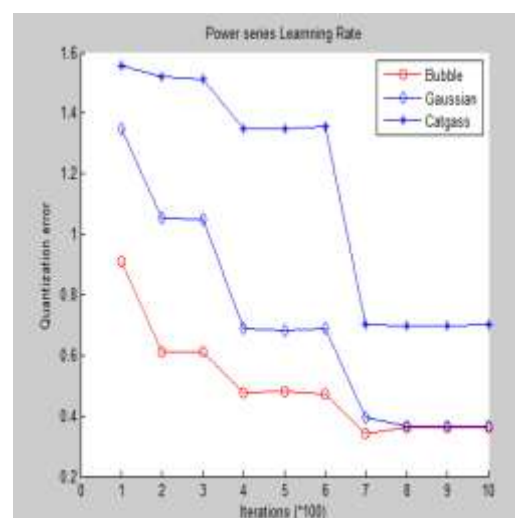
Learning rate	Bubble	Gaussian	Cutgass
Linear	583.7148	583.7333	583.7495
Inverse of time	593.8229	593.8243	593.8243
Power series	591.1260	591.1348	591.1484

Table 5: Running Time in Second of all Experiments in Three Type of Data.

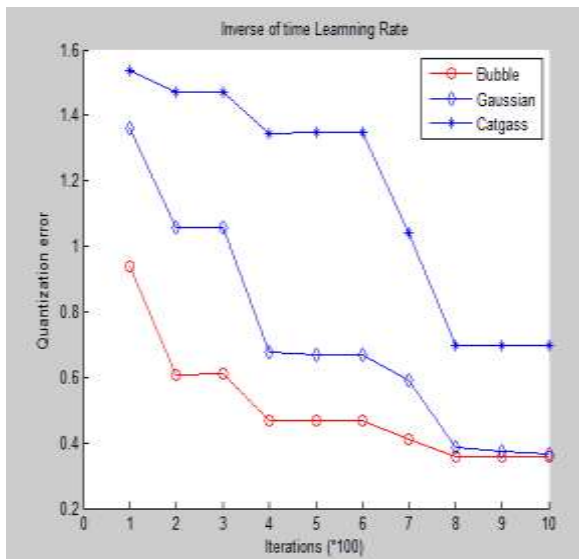
Learning rate	High Traffic	Low Traffic	Subscriber coordinates
Linear-Bubble	47.49	50.15	50.73
Linear-Gaussian	49.10	49.20	46.58
Linear-Catgass	51.36	48.01	47.29
Inverse of time-Bubble	53.01	55.22	52.29
Inverse of time-Gaussian	50.12	50.14	49.82
Inverse of time-Cutgass	49.76	48.05	51.64
Power series-Bubble	53.16	53.88	46.05
Power series-Gaussian	49.89	51.42	47.98
Power series-Catgass	51.75	46.60	48.04



(a)



(b)



(c)

Figure 6: The Quantization Error for Low Traffic Data of Cell Tower, (a) Linear Rate, (b) Power Series, (c) Inverse of Time Learning Rate

Tables 2, 3 and 4 show the quantization error of Bubble is less than others functions in all investigated cases. SOM gives the best performance with small quantization error when using Bubble function with linear and inverse of time learning rates. SOM provides the best result when the number of iteration is increasing. Therefore, it can be concluded that Bubble function is appropriate for all cases of learning rates while linear learning rate gives the smallest quantization errors in low traffic and subscriber coordinates. Further, when increase the number of iteration, the Bubble function results in better outcomes. Table 5 shows that there is little difference in running time between trials.

Figure 7 shows the number of required towers in Al-Rasheed Sub-district which regarded as rural area. The result was 8 towers with less quantization error when applying linear (as learning rate) and Bubble (as neighborhood function).



Figure 7: Number of Required Towers in Al-Rasheed Sub-district (Rural area)

6. Conclusion

This research presented applying various parameters to the SOM neural network for clustering problem. The proposed experiments were successfully implemented on three types of data; represent high traffic, low traffic and subscriber position distribution. Apply SOM to cluster a given data requires initializing important parameters such as learning rate in addition to defining appropriate neighborhood functions. In this research, three types of data, which are high and low traffic of specific cell tower with subscriber positions distribution in central of Iraq, are investigated using self-organizing map (SOM). SOM is applied with Bubble, Gaussian and Catgass neighborhood functions. All neighborhood functions are tested with three learning rates (linear, inverse of time and power series) and evaluated by the quantization error. Analyzing results were done by comparing the average of quantization error for each learning rate with all neighborhood function and number of iteration to three types of data. The results shows that Bubble function with linear learning rate gives the best performance with the smallest quantization error for data type of low traffic and subscriber position distribution, and with invers of time in high traffic. Additionally, there is a little difference in running time between all trials.

Acknowledgement:

The authors would like to acknowledge ASIA CELL Telecom for their technical support and cooperation.

Reference

- [1] D. H. K. Hoomod, D. I. Al-Mejibli, and A. I. Jabboory, "Optimized SOM for Cell Towers Distribution," in NTICT2017 IRAQ Baghdad: IEEE, 2017.
- [2] N. M. Shehu, "Coverage and Capacity Improvement in GSM Network," International Journal of Novel Research in Electrical and Mechanical Engineering, vol. 2, 2015.
- [3] J. J. HAMAD-AMEEN, "Cell Planning in GSM Mobile," WSEAS TRANSACTIONS on COMMUNICATIONS, vol. 7, 2008.
- [4] A. Awasthi and N. Arora, "An Approach to BTS Localization using Optimization Techniques," in International Journal of Engineering Research and Technology, 2014.
- [5] A. A. Pastukhov and A. A. Prokofiev, "Kohonen self-organizing map application to representative sample formation in the training of the multilayer perceptron," science direct, pp. 134-143, 2016.
- [6] M. S. S. K and D. H. R. S. Reddy, "Kohonen neural network based Kannada numerals recognition system," International Journal of Scientific & Engineering Research, vol. 5, 2014.
- [7] P. Sharma, G. Bhardwaj, G. Sharma, and V. Pandey, "Implementation of Artificial Neural Network (SOFM) for future prediction in Satellite Imagery," International Journal of Emerging Technology and Advanced Engineering, vol. 5, 2015.
- [8] W. Natita, W. Wiboonsak, and S. Dusadee, "Appropriate Learning Rate and Neighborhood Function of Self-organizing Map (SOM) for Specific Humidity Pattern Classification over Southern Thailand," International Journal of Modeling and Optimization, vol. 6, 2016.
- [9] R. A. P. Henriques, "Artificial Intelligence in Geospatial Analysis: applications of Self-Organizing Maps in the context of Geographic Information Science," in Information Systems. vol. Doctor of Philosophy in Information Systems Lisboa, Portugal: Universidade Nova de Lisboa – ISEGI, 2010.
- [10] A. C. TELECOMS, "Cell Tower Traffic," RF, Ed. Baghdad, 2017.
- [11] M. o. Planning, "Areas of districts and sub-districts," c. s. organization, Ed., 2015.
- [12] M. o. Planning, "Number of housing units," c. s. organization, Ed., 2009.
- [13] R. LASRI, "Clustering and Classification Using a Self-Organizing MAP," SAI Computing Conference, 2016.
- [14] S. Hasan and S. M. Shamsuddin, "Multistrategy self-organizing map learning for classification problems," Computational intelligence and neuroscience, vol. 2011, p. 1, 2011.

وظيفة الجوار ومعدل التعلم الفعال الخاص بخريطة التنظيم الذاتي (SOM) لتجميع حركة المرور في الأبراج الخلوية

عباس عيسى الجبوري	انتصار شديد المجبلي	حيدر كاظم حمو
معهد الدراسات العليا للمعلوماتية	قسم ادارة أنظمة المعلوماتية	قسم علوم الحاسبات
الهيئة العراقية للحاسبات والمعلوماتية	جامعة تكنولوجيا المعلومات والاتصالات	الجامعة المستنصرية
it_abbas@yahoo.com	dr.intisar.almejbli@gmail.com	drhjnew@gmail.com

المستخلص :

تستند الشبكة العصبية المسماة الخريطة الذاتية التنظيم (SOM) على التعلم غير الخاضع للرقابة، ولها مجموعة متنوعة من التطبيقات. من الضروري ضبط معلمات (SOM) قبل البدء في عملية التعلم لضمان أفضل النتائج. في هذا البحث، هناك ثلاثة أنواع من البيانات تمثل حركة المرور العالية والمنخفضة لبرج خلوي معين مع توزيع مواقع المشتركين في وسط العراق يتم التحقيق فيها باستخدام خريطة التنظيم الذاتي (SOM). وظائف (SOM) والمعلمات تؤثر على نتائجها النهائية. وبالتالي، يتم إجراء العديد من تكرار التجارب لاختبار وتحليل وظائف الجوار (Bubble, Gaussian and Catgass) مع ثلاثة معدلات التعلم (linear, inverse of time and power series)، وتم تقييمها بالاعتماد على (quantization error). نتائج التجارب تبين أن وظيفة الجوار Bubble مع معدل التعلم Linear يعطي أفضل نتيجة لتجميع حركة البرج الخلوي.