IOSR Journal of Applied Chemistry (IOSR-JAC)
e-ISSN: 2278-5736.Volume 9, Issue 12 Ver. I (December. 2016), PP 25-34
www.iosrjournals.org

Clustering Thyroid Disease Using Kohonen Networks Approach

Mali H. Hakem Alameady

Abstract: *Clustering using Artificial Neural Network (ANN) is a technique used in data mining. In neural network models have two types: supervised neural networks and unsupervised neural networks. Kohonen Networks is an unsupervised neural networks that used to train and test historical data in order to build clustering a model. In this experiment, The main objective of the experiment is to cluster the Thyroid disease Data by using model of Kohonen Networks. In this experiment, neural connection software is used to run the clustering by Thyroid disease dataset supplied by Werner, which adopted from UCI Repository. Clustering techniques can be applied when we can't find class to predict but rather when we can divide the instances into natural groups.*

Keywords: *Thyroid Disease Data; Clustering; Supervised & Unsupervised models; Kohonen Network*

I. Introduction

In 1982, Teuvo Kohonen introduced The Self-Organizing Map (SOM) , The SOM algorithm is an artificial neural network algorithm, it is a best known. There is a contrast between the SOM and other neural networks using supervised learning ,The SOM is a very well kind of neural network based on unsupervised learning, a topology preserving mapping is constructed by the SOM from the high dimensional space on map units in a way that the relative distances are preserved between data points. map units, or neurons, form two dimensional regular lattice where the semantic information are carried by the location of the map. So the SOM can serve as a clustering tool of the high dimensional data. Because it has a two dimensional shape, it is very easy to visualize.

A variety of networks has been invented by Teuvo Kohonen. The phrase "Kohonen network" refers to one of the following types of networks:

- Vector Quantization is a competitive network which it can be viewed as unsupervised density estimators or autoassociators (Kohonen, 1995).
- Self Organizing Map is a competitive network which provides from the input space to the clusters A topological mapping (Kohonen, 1995).
- Learning Vector Quantization is a competitive network used for supervised classification (Kohonen, 1998, 1995).

Kohonen Networks can provide an objective way for clustering data by utilizing self organizing network of the artificial neurons. Each neuron can store a weight vector (the array of weights), each of them which corresponds to the input in the data. When it presented with a totally new input pattern, each neuron starts to calculate its activation level which it is based on the followings and the definition .the w_i is the ith element in the weight vector and the p_i is the ith element in the pattern of input. neuron in a lower activation level meaning the weight of it is closer in Euclidian space to the new pattern of input ,it is allowed to adjust the weights so it is much closer to the input pattern, some of the nodes where it is near of it, the number that is determined while the algorithm runs, at the beginning of the nodes and linearly decreasing throughout the process of training.

The amount which every single neuron changes the weight vector is basically determined by the definition:

$$\delta w_i = -\alpha(w_i - p_i)$$

α is a learning rate, that begins by the user and it decreases to 0 while the algorithm is running , δw_i is the change in w_i . and this change is carried out for all the elements of the weight vector.

the Kohonen network is characterized as the self organizing map which it is used for recognition the pattern.

At First, the application of the input vector to network will cause an activation in output neurons: the neuron that has a highest value, it will represent the classification. Second, network will be trained via non supervised learning techniques . This will pose an interesting problem. once the training has been done without target vector, we can't tell a priori which one of the output neuron can be associated with the input vector's given class. when training has been completed, the idea behind the Kohonen network is for setting up the interconnected processing unit's structure that compete for all the signals. While the map's structure may be arbitrary, the package will support rectangular and linear maps only, mapping can be done easily by testing network with the input vectors.

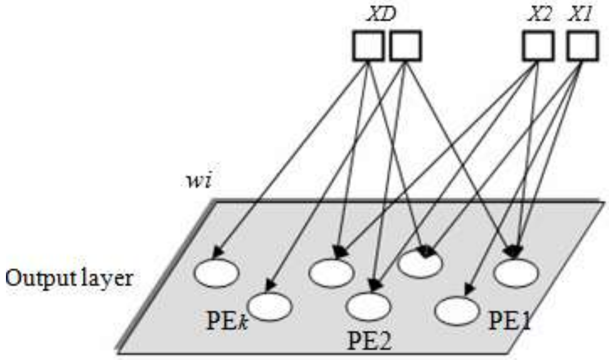

Figure 1: Connection in Kohonen Networks

The n connection weights in the neuron are basically treated as the vector in n dimensional space. Before the training, vector has to be initialized with any random values, then all the values will be normalized for making the unit's vector length in the weight space. All The input vectors that are in the training set are always normalized. The algorithm for training the Kohonen network is summarized as the following steps:

- Insert the input vector (X) to network.
- Calculate Dj the distance between X and weight vectors Wj of all the neurons.
- The neuron which has closest weight vector to X is always declared the winner.we have to Use this weight vector Wc as center of the weight vector's group that lie with the distance from Wc to d
- Train the vector's group according to all the weight vectors with the distance d from Wc.
- Perform all the steps 1 through 4 ,for all the input vectors.

while training proceeds, d values are reduced. by Kohonen you can start near 1 and then reduce to 0.1, but d may start as the large distance and then reduce to only single neuron.,the training's number cycles have to be nearly to 500 times of the output neuron's number for ensuring the statistical accuracy. Once the input and weight vectors have been normalized they will be viewed as the points in the unit hyper sphere's surface. The training algorithm will adjust the weight vectors that are surrounding winning neuron in order to be more similar to the input vector. That means the algorithm will tend to cluster weight vectors all around input vectors .this kind of adaptive units are organized in the layer in order to produce the feature map. the feature map can be defined as a nonlinear method used for representing space of original signal also for resembling topographic maps found in many areas of the brain. The feature map is produced by the unsupervised training of the adaptive units that are developed to spatially organized array of the feature detectors where the excited units signal's position statistically important features of input signals. so, more of the frequently occurring stimuli can be represented by a very larger areas in map than the infrequently ocurring stimuli.

The Kohonen maps and the unsupervised learning are only one way of the networks of training connectionist. But, if both of the inputs pattern and the corresponding outputs pattern they are known as a priori, the supervised learnings are used. usually The common supervised learning algorithms are the backpropagation algorithm that used to train generic PDP network.

II. Objectives

Generally, the objective of this project is to clustering Thyroid Diseases data, based on the change in the learning rate, using neural networks approach. Specifically, the objectives of this study are:

- ✓ *To identify an alternative techniques in clustering large number of data.*
- ✓ *To learn a pattern of data, by using their own characteristic.*

III. Methodology

3.1 Data Acquisition

The Thyroid data set is obtained from the UCI Repository of Machine Learning Databases. The donor of the database is Randolf Werner (evol@uniko.uni-koblenz.de). The data is available for public usage from October (1992). The dataset is used to determine if the patient is referred to clinic, it is hypothyroid. so three classes have to be built, it is normal hyper function and subnormal function. Because of the percent of the not hyperthyroid patients is 92, so good classifier have to be more better than 92 percent. All these data were tabulated in the Microsoft Excel format for preprocessing prior training and testing.

3.2 Data description

This data set includes descriptions of hypothetical samples corresponding to 22 species of thyroid diseases. Each species is identified as definitely normal (not hypothyroid), hyper function and subnormal functioning. The thyroid disease dataset has a total 3772 instances, 21 attributes, and no missing attributes values. The data is all presented in numbering form. The descriptions of the data sets are as following:

- Instances: 3772
- Classes: 3
- Attributes: 21 (15 binary attributes and 6 continuous attributes).

The purpose of the Neural Networks is to classify the thyroid disease's attributes and learn from their combination whether a thyroid disease is normal, hyper function and subnormal functioning (3 classes).

3.3 Thyroid Disease Data Attributes

Attribute Name	Possible Values	Attribute Name	Possible Values
age:	continuous.	lithium:	0, 1.
sex:	M (0), F(1).	goitre:	0, 1.
on thyroxine:	0, 1.	tumor:	0, 1.
query on thyroxine:	0, 1.	hypopituitary:	0, 1.
on antithyroid medication:	0, 1.	psych:	0, 1.
sick:	0, 1.	TSH:	continuous.
pregnant:	0, 1.	T3:	continuous.
thyroid surgery:	0, 1.	TT4:	continuous.
I131 treatment:	0, 1.	T4U:	continuous.
query hypothyroid:	0, 1.		
query hyperthyroid:	0, 1.		

3.4 Data cleaning and distribution

In this experiment, the cleaning process was done manually by getting through all the data, in which, any detected incorrect or missing values were corrected accordingly. However, since there is big gap between the classes, not all 3772 data were being used in this experiment. The highest number of class is 3428 and the lowest number of class is 93. In this experiment based on this information we have chosen 303 from thyroid disease dataset. Thyroid disease dataset was distributed into three different classes:

- ✓ **Class 1** = with No. of 93 instances. (30.70%)
- ✓ **Class 2** = with No. of 105 instances.(34.66%)
- ✓ **Class 3** = with No. of 105 instances. (34.66%)

Total = 303 instances.

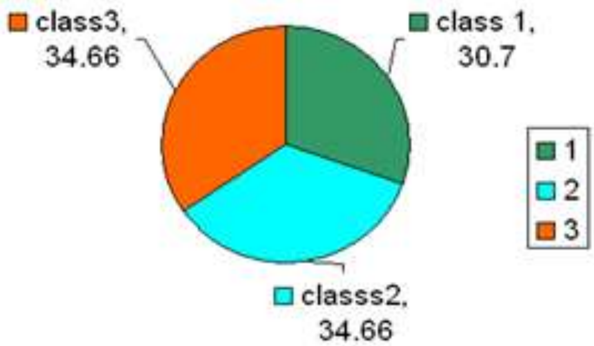

Figure 3.1: display the contribution of each value of classes

There are full amount of 303 instances that were used, and the all of the instances are processed in the Neural Connections program

3.5 Sample of Original Data

A part of original data is shown in Table 4.1 below:

```
0.76,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.00093,0.0208,0.142,0.109,0.129,3.
0.76,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.0011,0.02,0.092,0.095,0.097,3.
0.76,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.027,0.023,0.135,0.154,0.088,2.
0.5,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.058,0.024,0.025,0.121,0.02,1.
0.15,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0.145,0.017,0.019,0.113,0.017,1.
0.27,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.00002,0.029,0.118,0.152,0.078,3.
0.41,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0.00839,0.015,0.123,0.096,0.129,3.
0.43,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.0022,0.022,0.093,0.116,0.08,3.
0.25,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0.00009,0.022,0.181,0.123,0.147,3.
0.25,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.011,0.007,0.122,0.113,0.108,2.
0.25,1,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0.003,0.027,0.11,0.087,0.126,3.
0.37,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.0017,0.019,0.095,0.104,0.09,3.
0.79,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0.0013,0.009,0.102,0.095,0.107,3.
0.67,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.0032,0.019,0.091,0.079,0.115,3.
```

Table 3.1: Sample of Original Data

3.6 Setting a Target

The Thyroid Disease data was set in to tree, which are:
1= normal (not hypothyroid).
2= hyper function.
3= subnormal functioning.

3.7 Neural Network Tools

Neural Connection 2.0 is software that was used in this experiment. The software system allows us to build complex applications for solving many problems using neural computing and other techniques. Figure 3.2, illustrated the Kohonen Network model that going to be used in this experiment.

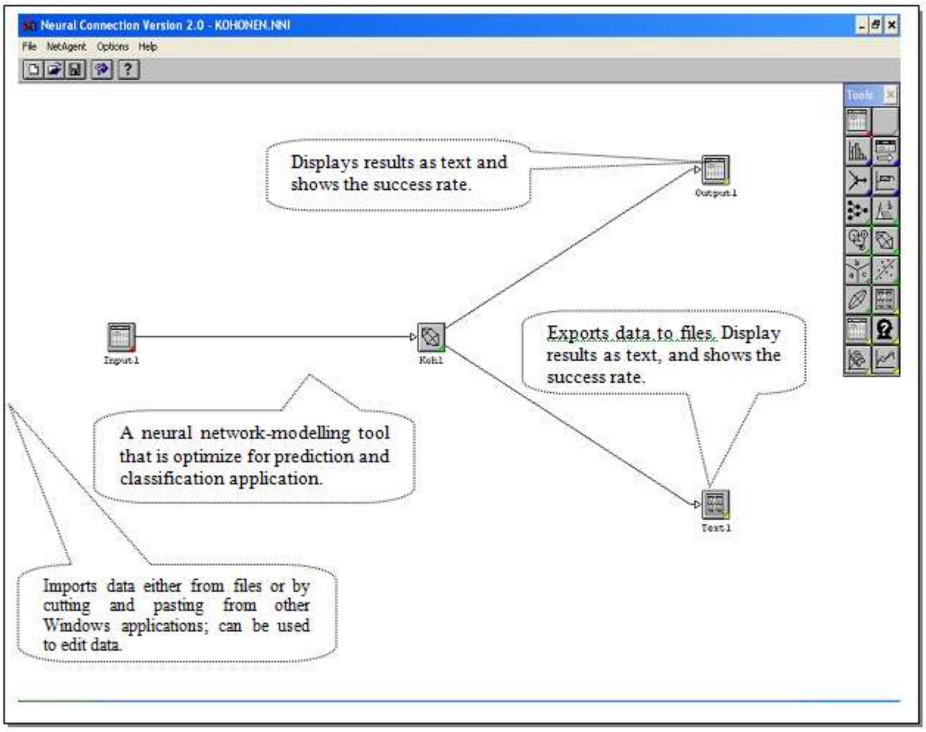

Figure 3.2: Kohonen Model Using Neural Connection 2.0

Tools in Neural Connection are independent of each other and make no assumptions about what other tools exist or precede or follow them. This makes the tools in Neural Connection totally interchangeable and makes Neural Connection extremely versatile program. Each type of tool that can be placed on the workspace has different function regarding through their usage. For this experiment, the following tools used for Kohonen model are described as above. Figure 3.2 below show the example of the data input tool that allow us to manipulate the input data directly in a transparent way and also allows a degree of freedom in specifying the usage of individual records and fields within data. A spreadsheet is a good way to display and manipulate data.

In spreadsheet, columns of data represent fields and records are represented by rows of data. These codes appear on the screen in the screen in the column from the left. They are listed below:

T = Training data, in cyan.
X = Test data, in yellow.
V = Validations data, in bright green.
R = Run data, in pale blue.

The codes listed below appear in the top row on the screen:

I = Input fields, in cyan.
T = Target fields, in yellow.
R = Reference fields in bright green.

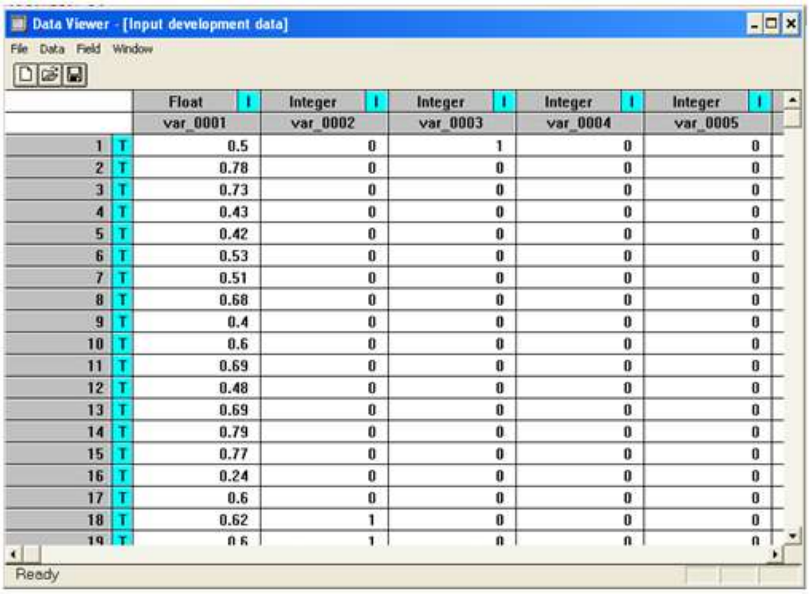

Figure 3.2: Example of data input window

A Perceptron has input and output layers. with no hidden layers. neurons in input layer are connected to neurons in output layer and the connections between the input layer and output layer are adjusted as the network has been trained. Figure 3.3 and figure 3.4 shows the example of MLP dialog box and training stage dialog box that were used to setting the training stage. The Two major learning parameters are always used to control training process are:

1. **The learning rate:** The specify whether NN is going to make major adjustment after each learning trial or it is only going to make minor adjustment.
2. **Side length:** The specify the number of artificial neurons or nodes, that the Kohonen is contains. The number of nodes is simply the side length raised to the power of the dimensionality of the Kohonen layer.

These two parameters will produce the impact on the training time and performance.

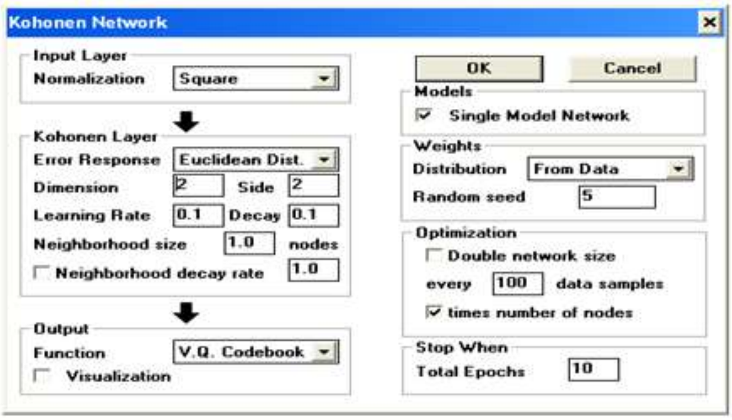

Figure 3.4: Kohonen Network Dialog Box

3.8 Data Allocation

In Multilayer Perceptron (MLP), the model needs target to do the training and testing process, but in Kohonen Network, they do not used the cluster and visualize data, all input are considered as the target. Before start the training process, the data should be randomized in order to get the better results.
Table 3.5 and figure 3.5 show that data allocation was set for this experiment as follow:

Table 3.5: Data Distribution

Data Distribution	Percentage	Amount
Training data	80%	242
Testing data	10%	31
Validation data	10%	30

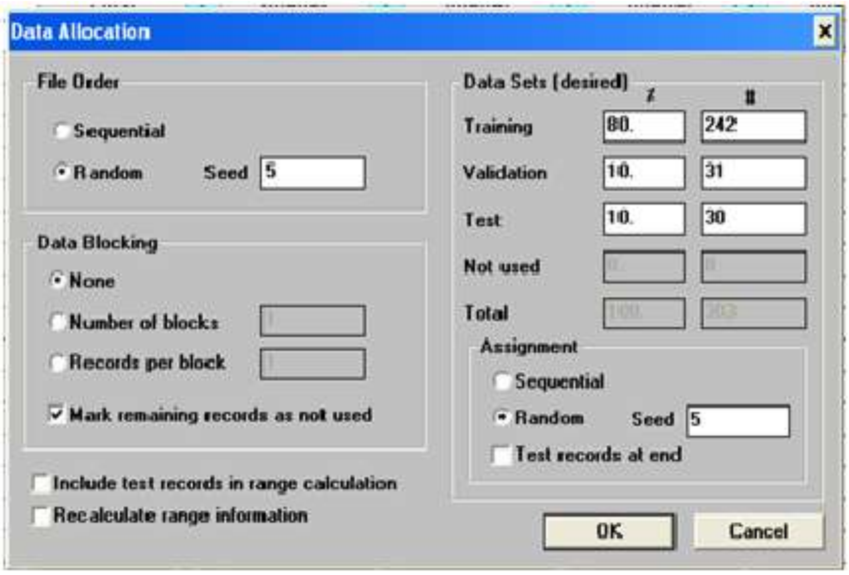

Figure 3.5: Data Allocation dialog box

IV. Rerults

In this experiment, the pattern of the clustering is affected by the learning rate. The experiment is carried out for side 2, which has 4, clusters ranges from (0 to 3), and side 3 with 9 clusters ranges from (0 to 8) and below is the result of all the experiment.

4.1 Side 2 : Table 4.1 below shows the result of training and testing that the parameter, learning rate has been set up from 0.1 to 1.0, Activation Function = V.Q.Codebook, and Training and Testing of Side 2

Table 4.1: Training and Testing for Side 2

LR	0.1		0.2		0.3		0.4		0.5		0.6		0.7		0.8		0.9		1.0	
Cluster	train	test	train	test	train	test	train	test	train	test	train	test	train	test	train	test	train	test	train	test
0	*15*	*0*	*14*	*0*	*160*	*19*	*1*	*0*	*1*	*28*	*237*	*28*	*237*	*28*	*237*	*28*	*2*	*0*	*2*	*0*
1	*134*	*15*	*7*	*0*	*7*	*11*	*3*	*2*	*237*	*2*	*1*	*2*	*1*	*2*	*1*	*2*	*1*	*2*	*2*	*2*
2	*13*	*2*	*14*	*2*	*73*	*0*	*2*	*0*	*2*	*0*	*2*	*0*	*2*	*0*	*2*	*0*	*2*	*0*	*1*	*0*
3	*80*	*13*	*207*	*28*	*2*	*0*	*236*	*28*	*2*	*0*	*2*	*0*	*2*	*0*	*2*	*0*	*237*	*28*	*237*	*28*

Table 4.2: Results of Training of Side 2

LR	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Cluster	Train	Train	Train	Train	Train	Train	Train	Train	Train	Train
Cluster 0	*15*	*14*	*160*	*1*	*1*	*237*	*237*	*237*	*2*	*2*
Cluster 1	*134*	*7*	*7*	*3*	*237*	*1*	*1*	*1*	*1*	*2*
Cluster 2	*13*	*14*	*73*	*2*	*2*	*2*	*2*	*2*	*2*	*1*
Cluster 3	*80*	*207*	*2*	*236*	*2*	*2*	*2*	*2*	*237*	*237*

5.1.1 Analysis

To analyze the effect of learning rate on the distribution on training pattern, the results exhibited in the table 4.2 are summarized in the following graphs. Table 4.2 shown that the total number of training data set 242 this employs that each cluster comprises of 60.500 .The description of data acceptable in table 4.1 shown that some of clusters are infected by LR ,for example LR 0.1 clusters the training bents in the clusters only 0,1, and 2. Since cluster 0 and 2 almost how the among the training .Therefore the description obtain is reasonable good, other LR that give to the cluster include 0.2, and 0.3.

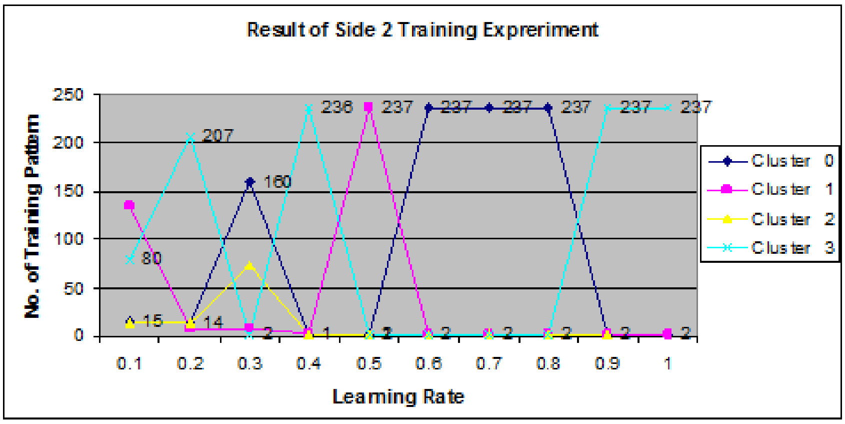

Figure 4.1

Two cluster will be obtained when LR is set on 0.1, 0.2 for analyze of duration of training bents shown that the points of clusters is closest when LR is 0.1, 0.2, 0.3. Indicated that LR 0.1 has produced a better distaining compare to, 0.2, 0.3 therefore 0.1 is considered to be best LR.From this graph we can extract that if two clusters are required the line graph indicate that LR 0.3 is the closest cluster points. Therefore LR 0.3 will be considered to be best LR. From the table 4.3, we are going to analyze the effect of the learning rate on the distribution of test patterns form cluster by cluster.

Table 4.3 : Results of Testing of Side 2

LR	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Cluster	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test
Cluster 0	*0*	*0*	*19*	*0*	*28*	*28*	*28*	*28*	*0*	*0*
Cluster 1	*15*	*0*	*11*	*2*	*2*	*2*	*2*	*2*	*2*	*2*
Cluster 2	*2*	*2*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
Cluster 3	*13*	*28*	*0*	*28*	*0*	*0*	*0*	*0*	*28*	*28*

The distribution of table 4.2 shown that LR 0.1 give rise to the cluster 1 and 3, for all LR cluster one comparers of like training patterns LR 0.4, 0.9 and 1 produce almost two cluster 1 and 3, if tree clusters are required for distribution LR 0.2 can be considered .However close valuation can indicate closer cluster points.

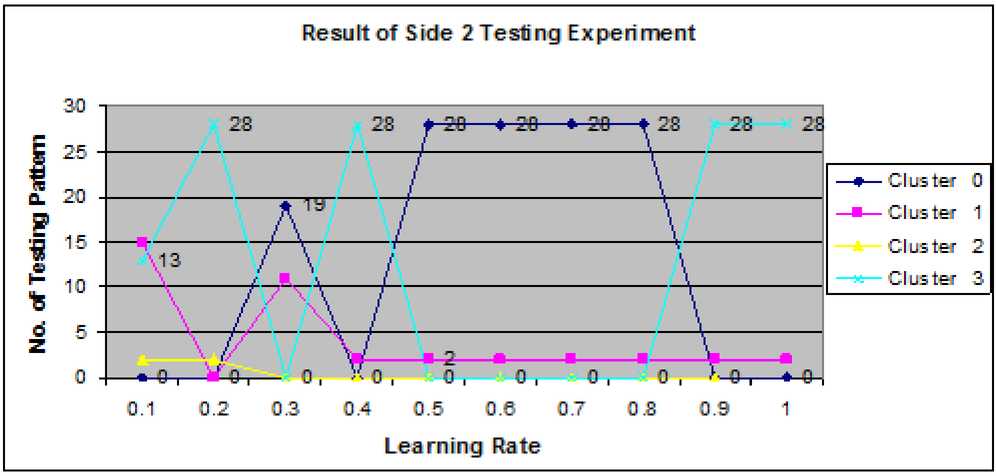

Figure 4.2

If two clusters are required the line graph indicates that LR 0.1 the closest cluster points. Learning rate 0.1 has been chosen to be the best learning rate to perform the clustering as it did not jeopardize the data distribution of all the clusters. Learning rate 0.1 also could produce all the 4 clusters (evenly distributed) compare to other learning rate that produce different distribution of clusters. As conclusion for the graph analysis of the Training and Test Side 2 using Kohonen Network, Learning Rate 0.1 gives the best data distribution which gives the target of 4 classes and hence it is chosen as the suitable learning rate to perform the clustering.

4.2 Side 3 :

The table 4.4 below shows that result of using Kohonen Network for Side 3 that has 9 clusters range from 0 to 8 with the learning rate has been set up from 0.1 to 1.0.

Training and Testing of Side 3:

Table 4.4: Training and Testing for Side 3

LR	0.1		0.2		0.3		0.4		0.5		0.6		0.7		0.8		0.9		1.0	
Cluster	train	test	train	test	train	test	train	test	train	test	train	test	train	test	train	test	train	test	train	test
0	11	2	1	0	1	0	1	0	14	0	139	13	1	0	5	0	2	0	8	1
1	71	4	11	2	7	0	2	0	2	0	5	0	2	0	2	0	5	0	5	0
2	13	2	13	2	13	2	13	2	137	13	77	12	5	0	77	12	1	0	1	0
3	7	0	118	12	2	0	2	0	70	12	3	2	17	2	1	0	199	23	139	15
4	40	9	65	11	65	11	65	11	3	2	2	0	3	2	3	2	17	2	77	10
5	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	5
6	8	1	8	1	8	1	8	1	8	1	8	1	8	1	8	1	8	1	2	0
7	15	0	14	0	14	0	14	0	1	0	2	0	199	23	139	13	3	2	2	0
8	72	10	7	0	127	14	132	14	2	0	1	0	2	0	2	0	2	0	3	2

Table 4.5: Results of Training of Side 3

LR	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Cluster	Train	Train	Train	Train	Train	Train	Train	Train	Train	Train
Cluster 0	*11*	*1*	*1*	*1*	*14*	*139*	*1*	*5*	*2*	*8*
Cluster 1	*71*	*11*	*7*	*2*	*2*	*5*	*2*	*2*	*5*	*5*
Cluster 2	*13*	*13*	*13*	*13*	*137*	*77*	*5*	*77*	*1*	*1*
Cluster 3	*7*	*118*	*2*	*2*	*70*	*3*	*17*	*1*	*199*	*139*
Cluster 4	*40*	*65*	*65*	*65*	*3*	*2*	*3*	*3*	*17*	*77*
Cluster 5	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*
Cluster 6	*8*	*8*	*8*	*8*	*8*	*8*	*8*	*8*	*8*	*2*
Cluster 7	*15*	*14*	*14*	*14*	*1*	*2*	*199*	*139*	*3*	*2*
Cluster 8	*72*	*7*	*127*	*132*	*2*	*1*	*2*	*2*	*2*	*3*

Table 4.5 shown that the total number of training data set 242 this employs that each cluster comprises of 30.250 .The description of data acceptable in table 4.5 shown that some of clusters are infected by LR, for example LR 0.2 clusters the training how the among the training .Therefore the description obtain is reasonable good, other LR that give to the cluster include 0.3, 0.4, 0.9 and 0.1.Three clusters will be obtained when LR is set on 0.1 0.3, 0.4, 0.9 and 0.1 for the analyze of duration of training bents shown that the points of clusters is closest when LR is 0.2 Indicated that LR 0.2 has produced a better distaining compare to 0.1 0.3, 0.4, 0.9 and 0.1, therefore 0.2 is considered to be best LR.

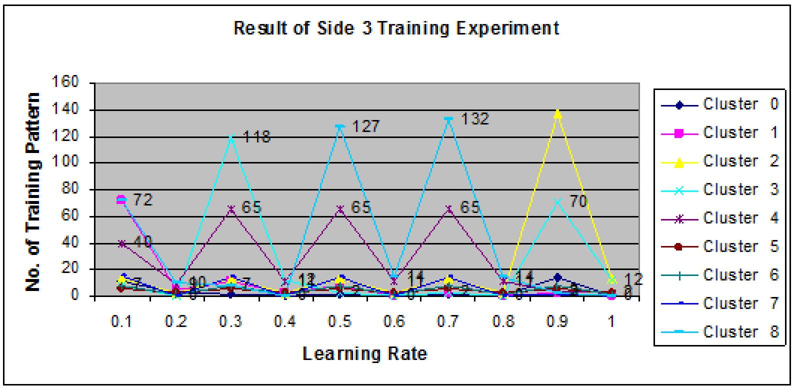

Figure 4.3

If three clusters are required the line graph indicate that LR 0.2 the closest cluster points
From the table 4.6, we are going to analyze the effect of the learning rate on the distribution of test patterns form cluster by cluster.

Table 4.6: Results of Testing of Side 3

LR	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1**
Cluster	Test	Test	Test	Test	Test	Test	Test	Test	Test	Test
Cluster 0	*2*	*0*	*0*	*0*	*0*	*13*	*0*	*0*	*0*	*1*
Cluster 1	*4*	*2*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
Cluster 2	*2*	*2*	*2*	*2*	*13*	*12*	*0*	*12*	*0*	*0*
Cluster 3	*0*	*12*	*0*	*0*	*12*	*2*	*2*	*0*	*23*	*15*
Cluster 4	*9*	*11*	*11*	*11*	*2*	*0*	*2*	*2*	*2*	*10*
Cluster 5	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*5*
Cluster 6	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*0*
Cluster 7	*0*	*0*	*0*	*0*	*0*	*0*	*23*	*13*	*2*	*0*
Cluster 8	*10*	*0*	*14*	*14*	*0*	*0*	*0*	*0*	*0*	*2*

The distribution of table 4.6 shown that LR 0.1 or 0.2 give rise to the clusters 2,5 and 6 ,for all LR clusters one comparers of like testing patterns LR 0.1, 0.2, 0.3, 0.4 …..,0.9 produce almost three clusters 2,5 and 6 ,if four clusters are required for distribution LR 0.1 and 0.2 can be considered .However close aviation can indicate closer cluster pointsIf three clusters are required the line graph indicate that LR 0.2 the closest cluster points

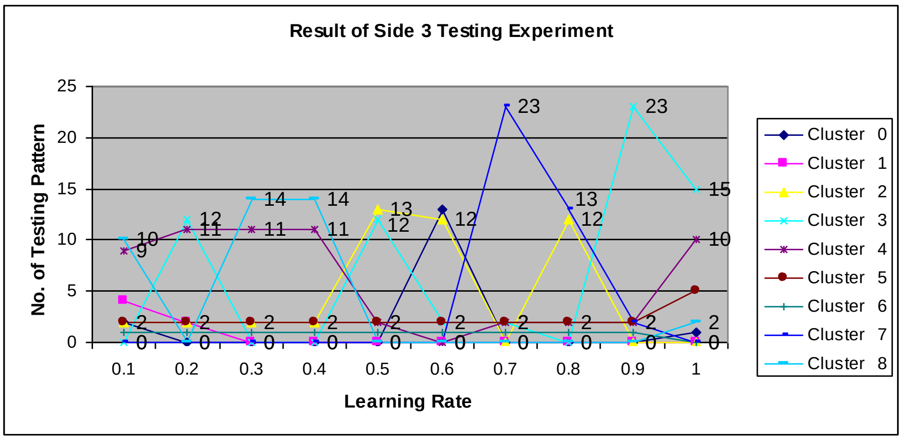

Figure 5.4

If three clusters are required the line graph indicates that LR 0.2 the closest cluster point.Learning rate 0.2 has been chosen to be the best learning rate to perform the clustering as it did not jeopardize the data distribution of all the clusters. Learning rate 0.2 also could produce all the 8 clusters (evenly distributed) compare to other learning rate that produce different distribution of clusters. As conclusion for the graph analysis of the Training and Test Side 3 using Kohonen Network, Learning Rate 0.2 gives the best data distribution which gives the target of 8 classes and hence it is chosen as the suitable learning rate to perform the clustering.

V. Conclusion

we can conclude from experimental results for each learning rate, the test and train having almost the same plot pattern. If the test is increase, the train value will increase too. The same condition happens if one of the values is decrease, another one will reduce too.

PARAMETER	**RESULTS**
Clustering	Kohonen Network
Learning rate	0.1 – 1.0
Activation Function	V.Q.Codebook
Side:	2, 3
Learning rate chosen to perform clustering	Side 2 : LR 0.1 Side 3: LR 0.2

Summary of Results

The train and test pattern for Side 2 and 3 affected by learning rate. The learning rate gives an affect to the plot pattern of the clusters, which some of the clusters gives drastic changes in their data distribution plot in the training and test pattern.Therefore, for each of the training and test side, one Learning Rate from 0.1 to 1.0 has been chosen to perform the clustering for all the clusters. This learning rate is chosen by analyzing its effect to the pattern in the clusters and whether it could give the best data distribution to produce these clusters. Learning rate that did not jeopardize the data distribution among the clusters and could give an evenly distributed pattern is considered to be the selected learning rate to perform the clustering.

Reference

[1] Klerfors, D. (1998).“*Artificial Neural Networks”.*School Of Business & Administration, Saint Louis University.

[2] Smith, L. (1996).“*An Introduction to Neural Networks.”* Department of Computing and Mathematics, University of Stirling.

[3] Mendelsohn, L. (1993). *“ Preprocessing Data for Neural Networks”.* California Avenue S.W., Seattle.

[4] Kohonen Network - Background Information. (2004).

[5] http://www.lohninger.com/helpcsuite/kohonen_network_-_background_information.htm.

[6] Tsoukalas, L.H. and Uhrig, R.E. (1997). “*Fuzzy and Neural Networks Approaches in Engineering.”* John Wiley and Sons, New York.

[7] Kohonen Networks: http:/ /www.cs.bham.ac.uk/ resources/ courses/ SEM2A2/ Web/ Kohonen.htm.

[8] Unsupervised Neural Networks http://soft.ee.umist.ac.uk/nigel/neural/SOM1.html

[9] Wiley, W.R. *(1997). “What is Artificial An Neural Networks?”* Environmental Molecular Sciences Laboratory, United State.

[10] *Malakooti, B., and Yang, Z. (1994). An Unsupervised Neural Network Approach for Machine-Part Cell Design. IEEE Communications. 665-670.*

[11] *Stergio, C., and Siganos, D. (n. d.). Neural Networks. Retrieved September 26, 2005, from http:// www.eurostat.org/MoreInfo.htm.*

[12] *UCI Machine Learning Repository, (1987, April 27) ' Thyroid Disease Database'.(Center for Machine Learning and Intelligent Systems),Available: https://archive.ics.uci.edu/ml/datasets/ Thyroid Disease (Accessed: 2016, February 10).*