Review of current Online Dynamic Unsupervised Feed Forward Neural Network classification

[Roya Asadi*, Haitham Sabah Hasan, Sameem Abdul Kareem]

Abstract— **Online Dynamic Unsupervised Feed Forward Neural Network (ODUFFNN) classification is suitable to be applied in different research areas and environments such as email logs, networks, credit card transactions, astronomy and satellite communications. Currently, there are a few strong methods as ODUFFNN classification, although they have general problems. The goal of this research is an investigation of the critical problems and comparison of current ODUFFNN classification. For experimental results, Evolving Self-Organizing Map (ESOM) and Dynamic Self-Organizing Map (DSOM) as strong related methods are considered; and also they applied some difficult datasets for clustering from the UCI Dataset Repository. The results of the ESOM and the DSOM methods are compared with the results of some related clustering methods. The clustering time is measured by the number of epochs and CPU time usage. The clustering accuracies of methods are measured by employing F-measure through an average of three times performances of clustering methods. The memory usage and complexity are measured by the number of input values, training iterations, clusters; and densities of clusters.** ***(Abstract)***

Keywords*—Neural Network (NN) model, Feed Forward Unsupervised Classification, Training, Epoch, Online Dynamic Unsupervised Feed Forward Neural Network (ODUFFNN)** ***(key words)

I. Introduction

In data mining, neural network has the best features of learning and high tolerance to noisy data, as well as their ability to classify data patterns on which they have not been trained. Neural networks are suitable for extracting rules, quantitative evaluation of these rules, clustering, self-organization, classification, regression feature evaluation, and dimensionality reduction [1-4].

Roya Asadi,

Haitham Sabah Hasan

Sameem Abdul Kareem

Department of Artificial Intelligence, Faculty of Computer Science and Information Technology, University of Malaysia, 50603. Tel:+603-79676300, Fax : +603-79579249

Neural networks are able to dynamically learn the types of input values based on their weights and properties. A feed forward neural network is a popular tool for statistical decision making and is a software version of the brain. The neural network is flexible algorithm that allows us to encode nonlinear relationships between input and desirable outputs [5-7]. The online dynamic unsupervised classification must solve several problems such as a huge data with high dimensions which causes huge memory usage, high processing time and low accuracy [2, 8, 9]; losing data details because of dividing input values in a few clusters [8, 10, 11]; defining the principles, number, densities and optimization of the clusters [12-14].

In the next sections, we will explain some unsupervised feed forward classification methods which are used as the base patterns for current ODUFFNN methods: Neural Gas (NG) model [15], Growing Neural Gas (GNG) model [16] and Self-Organization Map (SOM) [6]. Then, we will present some strong current ODUFFNN models with their advantages and disadvantages such as Evolving Self-organizing Map (ESOM) [17], Enhanced Self Organizing Incremental Neural Network for online unsupervised learning (ESOINN) [18], Dynamic Self Organizing Map (DSOM) [9], and Incremental Growing with Neural Gas Utility parameter (IGNGU) [19].

II. Unsupervised Feed Forward Neural Network classification

During clustering analysis, the data are divided into meaningful groups for a special goal with similarity inside of groups and dissimilarity between groups. Clustering is applied as preprocessing for other models, or data reduction, summarization, compression and vector quantization. Clustering methods can be categorized to: partitioning methods, hierarchical methods, model-based methods, density based methods, grid-based methods, frequent pattern-based, constraint-based and link-based clustering [2,11,20,21]. Clustering methods are implemented in two modes: training and mapping [6]. Training creates the map by using input values. High dimension of input values creates high relation complexity and dimension reduction method maps high dimensional data matrix to lower dimensional sub-matrix for effective data processing with high speed [2, 7].

K-means [21] is one type of the partitioning methods. The main problem of the partitioning methods is the definition of the number of clusters and an undefined initialization step [12, 14]. K-means clustering is efficient by good initialization in large datasets [8, 14]. Linde et al. [22] (LBG) is an algorithm for Vector Quantization (VQ) design. The VQ is powerful for using in large and high-dimensioned data. Since data points are represented by the index of their closest centroid, commonly occurring data have low error. The VQ is used for modelling of probability density functions by the distribution of prototype vectors. Each group is represented by its centroid point such as Self-Organization Map (SOM) and Growing Neural Gas (GNG) model [9]. These unsupervised feed forward neural network classification models are used as base patterns by current ODUFFNN methods. Kohonen's Self-Organizing Map (SOM) [6] maps multidimensional data onto lower dimensional subspaces where geometric relationships between points indicate their similarity. The SOM generates subspaces with an unsupervised learning neural network and a competitive learning algorithm. Neuron weights are adjusted based on their proximity to "winning" neurons. The neurons most closely resemble a sample input. The main advantage of using the SOM is that the data is interpreted easily. The SOM can handle several types of classification problems while providing a useful, interactive, and an intelligible summary of the data is considered. The SOM can cluster large data sets and solve their complexity in a short amount of time. The major disadvantage of the SOM is that it requires necessary and sufficient data in order to develop meaningful clusters. Lack of the data, noisy and unclean data affect on the accuracy of clustering. The weight vectors are based on data that can successfully group and distinguish inputs but initialization of weights is at random. Another problem of the SOM is difficulty to obtain a perfect mapping where groupings are unique within the map. LBG, NG and GNG are other kinds of current similar clustering models which cannot solve the main problems of clustering [18, 23]. Neural Gas (NG) [15] is similar the SOM method based on vector quantization and data compression. It dynamically partitions itself like gas and initializes the vectors of the weights at random. The neural gas algorithm is faster and gains smaller deformation errors but does not delete or add a node. Growing Neural Gas (GNG) [16] model is able to follow dynamically distributions by creating nodes and deleting them when they have a too small utility parameter. First, two random nodes are initialized and network competition is started for the highest similarity to the input pattern. Local errors are computed during the learning to determine where to add new nodes, and a new node is added close the node with the highest accumulated error. The number of nodes is increased to get input probability density [24] and maximum number of nodes and thresholds are predetermined. Therefore these methods are not suitable for online or lifelong learning.

III. Online Dynamic Unsupervised Feed Forward Neural Network (ODUFFNN) classification

The traditional neural network is powerful in solving artificial intelligence problems. However there are several problems in real and online environment that neural network model must be improved and be flexible such as online dynamic learning, control and intelligent agents. Some necessary properties of flexible, dynamic and online neural network models are as follows [2, 13]:

- Fast learning in one epoch from huge and high dimensional data
- To handle new data or noise in online mode immediately and dynamically
- The model should not be rigid and must be ready to change the structure, nodes, connection and etc
- To be able to accommodate and prune data, rules and etc incrementally
- To be able to control time, memory space, accuracy and etc efficiently
- To learn the number of clusters

Incremental learning refers to the ability of training in repetition to add or delete a node in lifelong learning without destroying the outdated prototype patterns [9, 18]. In this study, we consider some efficient current methods of Online Dynamic Unsupervised Feed Forward Neural Network Classification:

A. *Evolving Self-Organizing Map (ESOM)*

Evolving Self-organizing Map (ESOM) [17] is based on the SOM and the GNG methods. The ESOM begins without nodes and during training in one epoch, the network updates itself with on-line entry, and if necessary, it creates new nodes. The same SOM method, each node has a special weight vector and the strong neighbourhood relation is determined by the distance between connected nodes. If the distance is too big, it creates a weak threshold and the connection can be pruned.The ESOM is based on an incremental network quite similar to the GNG that creates dynamically based on the measure of the distance of the winner to the data, but the new node is created at the exact data point instead of the midpoint as in the GNG. The ESOM is a model based on Gaussian or normal distribution and VQ in its own way and creates Gaussian sub clusters across the data space. The ESOM is sensitive to noise nodes and prunes weakness connection and isolated nodes based on competitive Hebbian [25] learning. The ESOM works directly with prototype nodes in the memory and with entrance of each online input value, it checks all nodes as a neighbourhood or special cluster in the memory for adding or updating nodes of a network which takes long CPU time and high memory usage but during just one epoch. The ESOM is unable to control growing of the number of clusters and size of the network. The ESOM is sensitive to the first entrance of input value that is poor adapt to input vector values. Initialization of the parameters for training is based on trial and error experimentation and after several performances of the clustering model and checking the results, the best amounts of parameters can be recognized. Subsequently the model is not scalable and has different results in each performance. The time complexity of the ESOM model is $O(n^2.m)$ and its memory complexity is $O(n^2.m.\ s_m)$. The parameters n, m, s_m are the number of nodes, attributes and size of each attribute. The number of epochs does not consider in time and memory complexities because clustering process is just during one epoch.

B. Enhanced Self Organizing Incremental Neural Network for Online Unsupervised Learning (ESOINN)

Furao and Hasegawa introduced Self Organizing Incremental Neural Network for online unsupervised learning (SOINN) in 2006. The SOINN is suitable for the initial codebook, and the report of the number and density of the clusters. The method performs in two layers, the first layer for the generation of a topological structure of the input pattern or prototype, and the second layer uses the identified nodes of the first layer and reports the number of the clusters and their density distribution. However there are some problems in the SOINN [23]. The user must know the end of the learning in the first layer and the start of the learning in the second layer by using identified nodes of the first layer as input nodes. The SOINN cannot separate the clusters in high density and will have problem of overlapping between clusters. Another problem, the SOINN need many parameters for the second layer for insertion within classes. Therefore this method is not suitable for online clustering. The authors proposed the Enhanced Self Organizing Incremental Neural Network for online unsupervised learning (ESOINN) in 2007. The ESOINN [18] method has one layer and solves the problems of the SOINN. In this model, it is necessary that very old learning information is forgotten. The ESOINN model finds the winner and the second winner of the input vector, and then if it is necessary to create a connection between them or to remove the connection. The density, the weight of the winner and the subclass label of nodes will be updated in each epoch and the noise nodes dependent on the input values will be deleted. After learning, all nodes will be classified into different classes. The input vectors are not stored during learning. In next epoch, new data will be learned. If the distance between the new data and the winner or second winner is greater than the similarity threshold distance, the network will grow. If the distance between the new input vector and the winner or second winner is less than the similarity threshold, the new data have been learned well, and the network is not changed. Therefore learning of a new input does not destroy the last learned knowledge.

The disadvantages of the ESOINN model are: very outdated learning information is forgotten; new learned patterns are lost and only the old input pattern is represented and the topological structure of the incremental online data cannot be represented well; initialization of the parameters for training is based on trial and error; and there is relearning in several epochs [19]. The time complexity of the ESOINN is $O(c.n^2.m)$ and the memory complexity is $O(c.n^2.m.s_m)$. The parameters n, m, s_m are the number of nodes, attributes and size of each attribute and parameter c is the number of epochs.

C. Dynamic Self Organizing Map (DSOM)

Dynamic Self Organizing Map (DSOM) [9] is based on the SOM but it is suitable for incremental online dynamic unsupervised classification. In order to update the weights of neighbourhood nodes, the time dependency is removed, and the parameter of the elasticity or flexibility is considered which is free. The optimal parameter of the elasticity must be learned by using trial and error; if it is too high, the DSOM does not converge; and if it is too low, it may prevent the DSOM to occur and is not sensitive to the relation between neighbour nodes. If no node is close to the input values enough, other nodes must learn according to their distance to the input value. The DSOM method is not parameter free and the initialization of some parameters is done by trial and error. For clustering process, initializing the weights is done at random. Relearning during several epochs is another disadvantage of the DSOM. The time complexity of the DSOM is $O(c.n.m^2)$ and its memory complexity is $O(c.n.m^2.s_m)$.

D. Incremental Growing with Neural Gas Utility parameter (IGNGU)

Hebboul and Hacini et al. (2011) proposed an online unsupervised classification method as Incremental Growing with Neural Gas Utility Parameter (IGNGU). The introduced model is based on the GNG model. IGNGU [19] does not have any restraint and control on network structure and uses competitive learning method of Hebbian [25]. The structure of the IGNGU contains two layers of learning: first layer creates a suitable structure with low space of input and noise, and allow computing threshold; and second layer uses the output of the first layer in parallel and create the final structure of the clusters. The threshold of first layer must be greater than the within cluster distance and less than the between cluster distance. During clustering training, first layer considers one part of the dataset everytime and then moves to the second layer to process active nodes and disable all nodes and their parameters in the first layer. The same time first layer train another part of dataset. The IGNGU can solve some problems with disabling nodes. They do not crush learned data in the first layer and the second layer can learn this data again.

Some disadvantages of the IGNGU are: the parameters are determined experimentally by trial and error ; and some data and prototypes are lost for high speed clustering [19]. The time complexity of the IGNGU is $O(c.n^2.m)$ and its memory complexity is $O(c.n^2.m.s_m)$.

IV. Comparison

The goal of this study is an investigation and analyzing current ODUFFNN methods and identify their limitations and problems. We compared the ESOM and the DSOM as strong methods for implementation which are based on the SOM. We will discuss about them in section of discussion. All methods are implemented in Visual C#. Net. The mentioned methods are applied for clustering of seven datasets from UCI Irvine

Machine Learning Database Repository [26] with a comparison of several related methods. For experimental results the clustering time is measured by the number of epochs and CPU time; the clustering accuracy is measured by employing F-measure with an average of accuracies of 3 times performances; and the memory usage and complexity are measured by the amount of input values, training iterations and clusters, and densities of clusters. Also the accuracy of every method is measured by the number of clusters and the quantity of Correct Classified Nodes (CCN) which show total nodes with the correct class in the correct related clusters in all created clusters of the model; and the computed density of CCN by total of nodes in the dataset. The correct classified nodes are equal true positive and true negative.

A. *Breast Cancer Wisconsin Data Set*

Breast Cancer Wisconsin (Original) data set is selected from UCI Repository. As mentioned in the UCI Repository, the dataset characteristic is multi variable, the attributes characteristic is an integer, the number of instances is 699 and after cleaning 683, the number of attributes is 10 from life area. There are two classes: benign and malignant. The ESOM and the DSOM are compared with testing this dataset. Some related clustering methods are brought in the Table I just for better comparison. Table I shows the densities of correct classified nodes in clustering methods, the accuracies, the number of epochs and CPU times for Breast Cancer Wisconsin data set.

TABLE I. THE CORRECT CLASSIFIED NODES, ACCURACIES, CPU TIMES AND EPOCHS OF CLUSTERING METHODS FOR BREAST CANCER WISCONSIN DATA SET

Methods	CNN	Density of CCN %	Accuracy by F-measure %	Epoch	CPU TIME
SOM	660	97	-	20	-
K-Means	657	96	-	20	-
Neural Gas	657	96	-	20	-
GNG	477	70	-	5	-
ESOM	638	93	95	1	2" and 912 milliseconds
DSOM	459	67	75	700	4' , 22" and 165 milliseconds

Table I shows the CCN of the SOM for Breast Cancer Wisconsin dataset is 660, density of CCN is 97% with 20 epochs of training. The correct classified nodes of K-Means and Neural Gas methods are 657, the density of CCN is 96% with same 20 epochs [27]. The correct classified nodes of GNG method is 477, the density of CCN is 70% with 5 epochs [28]. The iteration of the ODUFFNN model of ESOM is one epoch with 638 CNN, 93% density of CCN and 95% F-measure accuracy during 2" and 912 milliseconds. The iterations of an online clustering model of DSOM are 700 epochs with 459 CNN, 67% density of CCN and 75% F-measure accuracy during 4', 22" and 165 milliseconds. All clustering models show two clusters for this dataset.

As the Table I shows the performance of the ESOM is better than the DSOM in the accuracy and CPU time and epochs. For better understanding the results of current ODUFFNN models, Standard Back Propagation Network [29] as a popular and standard supervised feed forward neural network classification is implemented. The SBPN can learn this dataset after 1000 epochs with an accuracy of 99.28% by using F-measure.

B. *Iris Data Set*

Iris data set is selected from UCI Repository. As mentioned in UCI Repository, the dataset characteristic is multi variable, the attributes characteristic is real, the number of instances is 150, the number of attributes is 4 from life area. There are three classes: Iris Setosa, Iris Versicolour and Iris Virginica. The methods of the ESOM and the DSOM are compared with testing this dataset. Some related clustering methods are brought in the Table II just for better comparison. Table II shows the densities of the correct classified nodes in clustering methods, the accuracies, the number of epochs and CPU times for Iris data set.

TABLE II. THE CORRECT CLASSIFIED NODES, ACCURACIES, CPU TIMES AND EPOCHS OF CLUSTERING METHODS FOR IRIS DATA SET

Methods	CNN	Density of CCN %	Accuracy by F-measure %	Epoch	CPU TIME
SOM	123	82	-	20	-
K-Means	134	89	-	20	-
Neural Gas	139	93	-	20	-
GNG	135	90	-	5	-
ESOM	146	97	97	1	36 milliseconds
DSOM	135	90	90	700	39" and 576 milliseconds

Table II shows the CCN of the SOM for Iris dataset is 123 and 82% density of CCN with 20 epochs of training. The number of correct classified nodes of K-Means is 134 and the density of CCN is 89% with 20 epochs. The correct classified nodes of the Neural Gas method is 139 and the density of CCN is 93% with same 20 epochs [27]. The correct classified nodes of GNG method is 135 and 90% density of CCN with 10 epochs [30]. The iteration of online clustering model of ESOM is one epoch with 146 CNN, 97% density of CCN and 97% F-measure accuracy during 36 milliseconds. The iterations of the online clustering model of DSOM are 700 epochs with 135 CNN, 90% density of CCN and 90% F-measure accuracy during 39" and 576 milliseconds. All clustering models show three clusters for this dataset.

As the Table II shows the performance of the ESOM is better than the DSOM in accuracy, CPU time and epochs. For comparison with a supervised feed forward neural network classification, the SBPN can learn this dataset after 14 epochs with the accuracy of 94% by using F-measure.

C. *Spambase Data Set*

Spambase Data Set is selected from UCI Repository. As mentioned in UCI Repository, the dataset characteristic is multivariable, the characteristics of the attributes are integer-real, the number of instances is 4601 and the number of attributes is 57 from computer area. There are two classes: Spam and Non-Spam. The methods of the ESOM and the DSOM are compared with testing this dataset. Some related clustering methods are brought in the Table III just for better comparison. Table III shows the densities of correct classified nodes in clustering methods, the accuracies, the number of epochs and CPU times for Spambase data set.

TABLE III. THE CORRECT CLASSIFIED NODES, ACCURACIES, CPU TIMES AND EPOCHS OF CLUSTERING METHODS FOR SPAMBASE DATA SET

Methods	CNN	Density of CCN %	Accuracy by F-measure %	Epoch	CPU TIME
SOM	1210	26	-	20	-
K-Means	1083	24	-	20	-
Neural Gas	1050	23	-	20	-
GNG	967	21	-	5	-
ESOM	2264	49	58	1	14' , 39" and 773 milliseconds
DSOM	2568	56	63	700	33' , 27' and 90 milliseconds

Table III shows the CNN of the SOM for Spambase dataset is 1210 and 26% density of CCN with 20 epochs of training. The CCN of K-Means is 1083 and 24% density of CCN with 20 epochs. The CCN of NG is 1050 and 23% density of CCN with same 20 epochs [27]. The CCN of GNG is 967 and 21% density of CCN with 5 epochs [28]. The iteration of the ESOM is one epoch with 2264 CNN, 49% density of CCN and 58% accuracy by using the F - measure during 14', 39" and 773 milliseconds. The number of iterations of the DSOM is 700 epochs with 2568 CNN, 56% density of CCN and 63% accuracy by using the F - measure during 33', 27' and 90 milliseconds. All clustering models show two clusters for this dataset.

As the Table III shows the performance of the DSOM is better than the ESOM in accuracy but with higher CPU time usage and epochs. For comparison with a supervised feed forward neural network classification, the SBPN can learn this dataset after 2000 epochs with an accuracy of 80% by F-measure.

D. *Spect Heart Data Set*

Spect heart Data Set is selected from UCI Repository. As mentioned in UCI Repository, the dataset characteristic is multivariable, the attributes characteristic is categorical, the number of instances is 267 and the number of attributes is 22. There are two classes: normal and abnormal. The methods of the ESOM and the DSOM are compared with testing this dataset. Some related clustering methods are brought in the Table IV just for better comparison. Table IV shows the densities of correct classified nodes in clustering methods, the accuracies, the number of epochs and CPU times for Spect Heart data set.

TABLE IV. THE CORRECT CLASSIFIED NODES, ACCURACIES, CPU TIMES AND EPOCHS OF CLUSTERING METHODS FOR SPECT HEART DATA SET

Methods	CNN	Density of CCN %	Accuracy by F-measure %	Epoch	CPU TIME
ESOM	177	66.3	79	1	197 milliseconds
DSOM	225	84	90	700	1' , 16" and 993 milliseconds

Table IV shows the iteration of the ESOM is one epoch with 177 CNN, 66.3% density of CCN and 79% F-measure accuracy during 197 milliseconds. The iterations of the DSOM are 700 epochs with 225 CNN, 84% density of CCN and 90% accuracy by using the F - measure during 1', 16' and 993 milliseconds. All clustering models show two clusters for this dataset.

As the Table IV shows the performance of the DSOM is better than the ESOM in accuracy but with higher CPU time and epochs. For comparison with a supervised feed forward neural network classification, the SBPN can learn this dataset after 25 epochs with an accuracy of 87% by F-measure and the SBPN by using Principal Component Analysis (PCA) can learn this dataset after 14 epochs training with an accuracy of 73% by F-measure. PCA [31] is known as dimension reduction technique and is a classical multivariate data analysis method that is useful in linear feature extraction and data compression [7, 32].

E. *Spectf Heart Data Set*

Spectf heart Data Set is selected from UCI Repository. As mentioned in UCI Repository, the dataset characteristic is multivariable, the attributes characteristic is an integer, the number of instances is 267 and the number of the attributes is 44 from life area. There are two classes: normal and abnormal. The methods of the ESOM and the DSOM are compared with testing this dataset. Some related clustering methods are brought in the Table V just for better comparison. Table V shows the densities of correct classified nodes in clustering methods, the accuracies, the number of epochs and CPU times for Spect Heart data set.

TABLE V. THE CORRECT CLASSIFIED NODES, ACCURACIES, CPU TIMES AND EPOCHS OF CLUSTERING METHODS FOR SPECTF HEART DATA SET

Methods	CNN	Density of CCN %	Accuracy by F-measure %	Epoch	CPU TIME
ESOM	165	61.8	77.8	1	342 milliseconds
DSOM	210	78.65	86.5	700	1' , 38" and 165 milliseconds

Table V shows the iteration of the ESOM is one epoch with 165 CNN, 61.8% density of CCN and 77.8% accuracy by F-measure during 342 milliseconds. The number of the iterations of the DSOM is 700 epochs with 210 CNN, 78.65% density of CCN and 86.5% accuracy by F-measure during 1' , 38" and 165 milliseconds. All clustering models show two clusters for this dataset.

As the Table V shows the performance of the DSOM is better than the ESOM in accuracy but with higher CPU time and epochs. For comparison with a supervised feed forward neural network classification, the SBPN can learn this dataset after 25 epochs with accuracy by F-measure around 79% and the SBPN by using PCA can learn this dataset after 14 epochs training with 75% accuracy by F-measure.

F. *Musk1 Data Set*

Musk1 Data Set is selected from UCI Repository. As mentioned in UCI Repository, the dataset characteristic is multivariable, the attributes characteristic is an integer, the number of instances is 476, and the number of the attributes is 168 from physical area. There are two classes: musks or non-musks. The methods of the ESOM and the DSOM are compared with testing this dataset. Some related clustering methods are brought in the Table VI just for better comparison. Table VI shows the densities of correct classified nodes in clustering methods, the accuracies, the number of epochs and CPU times for Spect Heart data set.

TABLE VI. THE CORRECT CLASSIFIED NODES, ACCURACIES, CPU TIMES AND EPOCHS OF CLUSTERING METHODS FOR MUSK1 DATA SET

Methods	CNN	Density of CCN %	Accuracy by F-measure %	Epoch	CPU TIME
ESOM	202	42	48	1	1" and 1 milliseconds
DSOM	202	42	48	700	4' , 52" and 562 milliseconds

The Table VI shows the iteration of the ESOM is one epoch with 202 CNN, 42% density of CCN and 48% accuracy by F-measure during 1" and 1 milliseconds. The number of the iterations of the DSOM is 700 epochs with 202 CNN, 42% density of CCN and 48% accuracy by F-measure during 4', 52" and 562 milliseconds. All clustering models show two clusters for this dataset.

As the Table VI shows the performance of the DSOM and the ESOM are equal in accuracy but the DSOM takes higher CPU time and epochs. For comparison with a supervised feed forward neural network classification, the SBPN can learn this dataset after 100 epochs with 75% accuracy by F-measure.

G. *Musk2 Data Set*

Musk2 Data Set is selected from UCI Repository. As mentioned in UCI Repository, the dataset characteristic is multivariable, the attributes characteristic is an integer, the number of instances is 6598, and the number of attributes is 168 from physical area. There are two classes: musks or non-musks. The methods of the ESOM and the DSOM are compared with testing this dataset. Some related clustering methods are brought in the Table VII just for better comparison. Table VII shows the densities of correct classified nodes in clustering methods, the accuracies, the number of epochs and CPU times for Musk2 data set.

TABLE VII. THE CORRECT CLASSIFIED NODES, ACCURACIES, CPU TIMES AND EPOCHS OF CLUSTERING METHODS FOR MUSK2 DATA SET

Methods	CNN	Density of CCN %	Accuracy by F-measure %	Epoch	CPU TIME
ESOM	4657	71	56.4	1	28" and 1 milliseconds
DSOM	3977	60	41.4	700	41' , 1" and 633 milliseconds

Table VII shows that the iteration of online clustering model of ESOM is one epoch with 4657 CNN, 71% density of CCN and 56.4% accuracy by F-measure during 28" and 1 milliseconds. The number of the iterations of the DSOM is 700 epochs with 3977 CNN, 60% density of CCN and 41.4% accuracy by F-measure during 41', 1" and 633 milliseconds. All clustering models show two clusters for this dataset.

As the Table VII shows the performance of the ESOM is better than the DSOM in accuracy, higher CPU time and epochs. For comparison with a supervised feed forward neural network classification, the SBPN can learn this dataset after 100 epochs with 67% accuracy by F-measure.

V. Discussion

In this section, we investigate and analyse current ODUFFNN methods and identify their limitations and problems. Table VIII shows some bold advantages of current online dynamic unsupervised feed forward neural network classification models.

TABLE VIII. SOME BOLD ADVANTAGES OF CURRENT ONLINE DYNAMIC UNSUPERVISED FEED FORWARD NEURAL NETWORK CLASSIFICATION MODELS

	ESOM	ESOINN	DSOM	IGNGU
Authors / Year of publication	(Deng and Kasabov 2003)	(Furao, Ogura et al. 2007)	(Rougier and Boniface 2011)	(Hebboul, Hacini et al. 2011)
Base patterns	SOM, GNG	GNG	SOM	GNG, Hebbian
Some bold features (Advantages)	Begin without any node	Control the number and density of each cluster	Improve the formula of updating the weights	Train by two layers in parallel
	Update itself with online input value	Initialize codebook	Elasticity or Flexibility property	Control density of each cluster and size of the network
	The nodes with weak thresholds can be pruned	Prune for controlling noise and weak thresholds		Control noise
		Input vectors are not stored during learning		Fast training by pruning
		New input does not destroy last learned knowledge		
Time Complexity	$O(n^2.m)$	$O(c.n^2.m)$	$O(c.n.m^2)$	$O(c.n^2.m)$
Memory Complexity	$O(n^2.m.s_m)$	$O(c.n^2.m.s_m)$	$O(c.n.m^2.s_m)$	$O(c.n^2.m.s_m)$

Table VIII shows some feed forward neural network clustering methods such as the SOM and the GNG which are used as base patterns and improved by the authors for proposing current ODUFFNN classification models. Therefore the models inherit the properties of the base patterns but by improving their structures; consequently the ODUFFNN methods obtain new properties. For example, the DSOM earns the property of elasticity by improving the formula of updating the weights of the SOM. The DSOM can control the size of the network, the number of clusters and their densities through elasticity property. The experimental results show the accuracy of the DSOM is often better than the ESOM in large dataset. The ESOM is based on the SOM and the GNG methods. The bold properties of ESOM are such as starting without any node; updating the clusters by online input values; and pruning weak connection but by losing clustering accuracy. The ESOM trains during one epoch and have better CPU time usage for clustering but the CPU time usage for one epoch is too long based on the time complexity of the ESOM. The ESOINN is based on the GNG with some suitable properties such as initial codebook, controlling noise but with losing accuracy. The IGNGU is based on the GNG and the Hebbian learning rule [25] models with some abilities such as parallel training in two layers; controlling noise and densities of clusters. All current related methods show the general problems as follow:

- Losing data details [8, 10, 11, 19]
- Unable to handle and manage clustering tasks such as pruning data and rules immediately [2, 8, 9]
- Inflexible and rigid model in structure, nodes, connection and etc [9, 13]
- Using random weights and random parameters for controlling the clustering tasks [2, 19]
- The number of clusters and density of each cluster are not clear and cannot be easily learnt [12-14, 18]
- Relearning during several epochs [11, 19]
- High dimensional data and big dataset create data complexity [7, 17, 33, 34]
- Unable to handle new data or noise [2, 13]
- Sensitive to the order of input [2, 11, 14]

The general problems are high CPU time and high memory usage with low accuracy during clustering [8, 9, 17, 19]. We illustrated some sources of the mentioned problems. Kasabov (1998) and Han and Kamber (2006) explained some necessary properties of flexible, dynamic and online neural network models but needless to say, the general problems remain and current ODUFFNN methods have poor solutions for solving them.

VI. Conclusion

Online dynamic feed forward neural network clustering is a valuable subject to research because of applying in different environments such as email logs and networks; although this field is new and there are some critical problems. The goal of this study is a comparison of the ODUFFNN methods and finding critical problems. Also we illustrated and analysed some sources of the problems like relearning, using random parameters and weights which affect the accuracy of clustering, CPU time and memory usage. For experimentation, the ESOM and the DSOM as strong ODUFFNN methods are implemented in C#.Net environment by applying seven datasets from the UCI Repository; and their results are compared with the results of several related methods in the scope of this study.

References

[1] Andonie, R. and B. Kovalerchuk, Neural Networks for Data Mining: Constrains and Open Problems. Computer Science Department, Central Washington University, Ellensburg, USA, 2007.

[2] Han, J. and M. Kamber, Data mining: concepts and techniques. 2006: Morgan Kaufmann.

[3] Bengio, Y., et al. Special Issue on Neural Networks for Data Mining and Knowledge Discovery. 2000: Institute of Electrical and Electronics Engineers.

[4] Mitra, S., S.K. Pal, and P. Mitra, Data mining in soft computing framework: A survey. Neural Networks, IEEE Transactions on, 2002. 13(1): p. 3-14.

[5] Demuth, H., M. Beale, and M. Hagan, Neural Network Toolbox TM 6. User's Guide, The Math Works. Inc., US, 2008.

[6] Kohonen, T., Self-organizing maps, volume 30 of Springer Series in Information Sciences. Springer, Berlin, 1997. 4: p. 22-25.

[7] Van der Maaten, L., E. Postma, and H. Van den Herik, Dimensionality reduction: A comparative review. Preprint, 2007.

[8] Jain, A.K., Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 2010. 31(8): p. 651-666.

[9] Rougier, N. and Y. Boniface, Dynamic self-organising map. Neurocomputing, 2011.

[10] Jain, A.K., M.N. Murty, and P.J. Flynn, Data clustering: a review. ACM computing surveys (CSUR), 1999. 31(3): p. 264-323.

[11] Pavel, B., Survey of clustering data mining techniques. Accrue Software Inc, 2002.

[12] Andrews, N.O. and E.A. Fox, Recent developments in document clustering. Computer Science, Virginia Tech, Tech. Rep, 2007.

[13] Kasabov, N. ECOS: Evolving connectionist systems and the ECO learning paradigm. 1998: Citeseer.

[14] Mercer, D.P., Clustering large datasets. Linacre College, 2003.

[15] Martinetz, T.M., S.G. Berkovich, and K.J. Schulten, Neural-gas' network for vector quantization and its application to time-series prediction. Neural Networks, IEEE Transactions on, 1993. 4(4): p. 558-569.

[16] Fritzke, B. A growing neural gas network learns topologies. 1995: Citeseer.

[17] Deng, D. and N. Kasabov, On-line pattern analysis by evolving self-organizing maps. Neurocomputing, 2003. 51: p. 87-103.

[18] Furao, S., T. Ogura, and O. Hasegawa, An enhanced self-organizing incremental neural network for online unsupervised learning. Neural Networks, 2007. 20(8): p. 893-903.

[19] Hebboul, A., M. Hacini, and F. Hachouf. An incremental parallel neural network for unsupervised classification. 2011: IEEE.

[20] Moore, A.W., Very fast EM-based mixture model clustering using multiresolution kd-trees. Advances in neural information processing systems, 1999: p. 543-549.

[21] Goebel, M. and L. Gruenwald, A survey of data mining and knowledge discovery software tools. ACM SIGKDD Explorations Newsletter, 1999. 1(1): p. 20-33.

[22] Linde, Y., A. Buzo, and R. Gray, An algorithm for vector quantizer design. Communications, IEEE Transactions on, 1980. 28(1): p. 84-95.

[23] Furao, S. and O. Hasegawa, An incremental network for on-line unsupervised classification and topology learning. Neural Networks, 2006. 19(1): p. 90-106.

[24] Hamker, F.H., Life-long learning Cell Structures--continuously learning without catastrophic interference. Neural Networks, 2001. 14(4-5): p. 551-573.

[25] Hebb, D.O., The organization of behavior: A neuropsychological approach. NewYork: John Wiley & Sons. Hinton, GE (1989). Deterministic Boltzmann learning performs steepest descent in weightspace. Neural Computation, 1949. 1: p. 143-150.

[26] Murphy, P.M., Repository of Machine Learning and Domain Theories. 1997.

[27] Camastra, F. and A. Verri, A novel kernel method for clustering. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2005. 27(5): p. 801-805.

[28] Bouchachia, A., B. Gabrys, and Z. Sahel. Overview of some incremental learning algorithms. 2007: IEEE.

[29] Werbos, P.J., Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. 1974.

[30] Costa, J.A.F. and R.S. Oliveira. Cluster analysis using growing neural gas and graph partitioning. 2007: IEEE.

[31] Jolliffe, I., Principal component analysis. . 1986, Springer, New York.

[32] Daffertshofer, A., et al., PCA in studying coordination and variability: a tutorial. Clinical Biomechanics, 2004. 19(4): p. 415-428.

[33] Hinton, G.E. and R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks. Science, 2006. 313(5786): p. 504.

[34] Kohonen, T., et al., Self organization of a massive document collection. Neural Networks, IEEE Transactions on, 2000. 11(3): p. 574-585.

About Author (s):

Roya Asadi received the Bachelor degree in Computer Software engineering from Electronics and Computer Engineering Faculty, Shahid Beheshti University and Computer Faculty of Data Processing Iran Co. (IBM), Tehran, Iran. She obtained Master of Computer Science in Database System from UPM university in Malaysia. She is a research student of Ph.D. of Computer Science in Artificial Intelligence (Neural Network) in University of Malaya. Her professional working experience includes 12 years of service as Senior Planning Expert 1. Her interests include Artificial Neural Network modeling, Medical Informatics, Image Processing.

Haitham Sabah Hasan is a PhD student at the faculty of Computer Science , University of Malaya, Malaysia. He holds a master degree in Computer Science from Oxford University, UK. He is now working on pattern recognition field and specifically on hand gesture recognition using Artificial Neural Network. He has published 3 ISI journal papers.

Assoc. Prof. Datin. Dr. Sameem Abdul Kareem: received the B.Sc. in Mathematics (Hons) from University of Malaya, in 1986, the M.Sc. in Computing from the University of Wales, Cardiff, UK, in 1992, and the Ph.D. in computer science from University of Malaya, in 2002. She is currently an Associate Professor of the Department of Artificial Intelligence, Faculty of Computer Science and Information Technology, University Of Malaya. Her interests include Medical Informatics, Information Retrieval, Data Mining and Intelligent Techniques. She has published over 80 journal and conference papers.