The Self-Organization by Lateral Inhibition Model: Validation of Clustering

Bin Tang, Malcolm I. Heywood and Michael Shepherd
Faculty of Computer Science
Dalhousie University
Halifax, Nova Scotia, Canada B3H 1W5
E-mail: {btang | mheywood | shepherd}@cs.dal.ca

Abstract- **An improved version of the Self-Organization by Lateral Inhibition Model (SOLI) has been applied to two synthetic data sets as well as the Breast Cancer and Liver data sets, two well-known benchmark data sets. The methodology developed combines the use of various validity indices with the SOLI model to discover the proper cluster structure within the data sets. In addition, the results explain why the Breast Cancer data set tends to be clustered so accurately while the Liver data set tends to be so difficult to cluster accurately.**

I. INTRODUCTION

Clustering is a fundamental problem for most computational intelligence tasks. Clustering may be used to discover the statistical regularity of an underlying data set for data mining applications, or it may be used together with supervised learning for feature detection, fuzzification and removing noise [1, 2].

Clustering, however, is very complex and there are a number of inter-related problems in discovering the clusters in a data set. The challenge is to find a solution that addresses all of these problems. In this paper, we present an improved version of the Self-Organization by Lateral Inhibition (SOLI) model, a new neural network model for clustering, as well as a methodology that appears to address a number of the inter-related problems.

The first problem of clustering is determining the proper number of clusters to be discovered in a dataset. There is no straight-forward solution.

The second problem is related to the clustering validity index, which measures the quality of the discovered clusters. There have been many validity indices proposed [3-11]. Unfortunately, none of them is perfect under all conditions and these indices do not always agree with each other, thus making their findings more inconclusive. A good validity index should not be tied to any specific implementation of clustering, rather it should be applicable universally.

The third problem is the robustness of the clustering algorithm used. Given similar training parameters, there is no guarantee that an algorithm will always converge to similar results (same number of clusters with similar locations). For example, it is well known that the performance of fuzzy c-means clustering depends not only on the proper number of clusters, but also on the random initialization of the clustering seeds [9].

The fourth problem is the interplay between the number of clusters discovered and the validity indices. These measures rely on each other. Unfortunately, there is a certain amount of uncertainty associated with both measures, making the evaluation of the clustering quality and determination of the proper number of clusters difficult; where the latter relies on good validity indices.

A related aspect of clustering is the discovery of the inherent hierarchical structure of a given data set. There are two major types of hierarchical clustering algorithms; agglomerative and divisive [12]. Agglomerative clustering starts by merging single objects and then repeatedly merging these lower level clusters into larger and higher level clusters. The merging criteria are rather subjective and heuristic and tend to be sensitive to outliers [13]. Divisive clustering algorithms partition the data set into smaller clusters. Divisive algorithms are usually efficient but the clustering result does not always correspond to natural clusters existing in the data [12]. The problem common to both classes of algorithms is that the merging or partitioning procedures are not guided by clustering quality measures such as cluster validity indices.

There have been many neural network models proposed for unsupervised learning tasks or clustering tasks. Among them, Kohonen's Self-Organizing Maps (SOM) is one of the most popular [14]. Inspired by SOM, a whole family of SOM-like neural network models have been proposed in recent years, with the intension of improving on certain aspects of the original SOM, as summarized in [15].

Recently, a new unsupervised neural network model for clustering, the SOLI model [16], was proposed and the potential of this model and its suitability for clustering tasks demonstrated. In this paper, we report recent improvements in the SOLI model and demonstrate the robustness of the enhanced SOLI model for clustering. We also present a methodology that combines the SOLI model and the use of various validity indices to discover the proper cluster structure within a data set.

In Section II, the family of SOM-like models is summarized while in Section III the enhanced SOLI model is described. The validity indices used in this research are described in Section IV. Sections V and VI describe the experiments and present the results, while Section VII presents our conclusions and points to future research.

0-7803-7898-9/03/$17.00 ©2003 IEEE

II. SUMMARY OF SOM-LIKE MODELS

Kohonen's SOM [14] is a neural network model, originally motivated by Hebbian learning, in which external stimuli are the sources of neuron excitation. Due to different degrees of excitation, neurons cooperate and compete through lateral connections (excitation and inhibition). In the standard SOM model, lateral connections are purely excitatory, approximated by a Gaussian kernel, which acts as a topology neighborhood function. This represents a very popular model on account of its' mathematical simplicity and visualization ability and has been applied in wide range of application areas.

Many SOM-like models have been proposed in recent years [15]. These models were intended to improve on the fixed map size, fixed dimensionality of the map and the topology preservation property of Kohenen's SOM. Some models have fixed dimensionality, while other models are incremental in nature. One of the fundamental differences among these models is how the neighborhood function is defined (details can be found in [15, 16]).

All of these models require a slow decrease of the learning rate or of the size of the neighborhood or both to guarantee convergence. Most of these models are of $O(n^2)$ time complexity. Many of these models require a number of training parameters (up to 7) to be specified by the user. To choose an optimal set of parameter settings for any given data set can potentially be very difficult [16].

III. THE SOLI MODEL

Lateral inhibition, though not heavily used in the neural network models for machine learning tasks, play an important role in computer vision and brain-style computing research [17-19]. Recent biological discoveries and artificial simulations shed new light on the relative role of lateral inhibition in the self-organization of biological neural networks [20, 21]. Contrary to previous belief, there exist a large number of short-ranged inhibitory lateral connections and long-range excitatory lateral connections among neurons. The local nature of inhibition contributes to the specialization of responses in the inferotemporal cortical neurons for complex stimulus [20]. In [21], the utilization of inhibitory connections and excitatory connections together provided for the modeling of advanced functionalities, such as perceptual grouping. It has been noticed that balancing between excitatory and inhibitory activation ensures a stable development of the network [21].

In the SOLI model [16], we explicitly incorporated inhibitory mechanisms. We have demonstrated that SOLI is suitable for the clustering problem under many conditions; data with arbitrary shape, arbitrary density distributions and density levels, and arbitrary dimensionality.

Comparing the SOLI model to SOM-like models, the SOLI model features reduced model complexity in terms of the training parameters and reduced time complexity, as demonstrated empirically in [16].

In the current SOLI model, the lateral connections among neurons are not implemented; thus, no neighborhood function is implemented. Therefore, SOLI is functionally related to vector quantization methods rather than to SOM-like networks. It is in the form of the Mexican hat function, implemented as DoGs (difference of Gaussians), and formulated as:

$h_{Ui}(x, u_i)=exp(-dist(x, u_i)^2/2\sigma^2)-exp(-dist(x, u_i)^2/2(\lambda\sigma)^2))/\lambda$

where $\lambda \in [1, \infty]$, σ is the given parameter relating to the scale of the cluster and controls the sharpness of the activation function in the shape of Mexican hat. and $dist(\cdot)$ denotes the Euclidian distance between pattern x and neuron u_i. Since h_{Ui} is further coupled with anti-Hebbian learning, the first part of this function dictates the local inhibition and the second part global excitation. λ controls the relative strength between excitation and inhibition. A larger λ value will cause flattened low degree global excitation. Too small a value will negate the influence of global excitation. A typical Mexican hat function is depicted in Fig 1. From our experience, λ=2 is used as the default value. It can be easily proven that the integration of this function over the range [0,∞] on distance, $dist(\cdot)$, is 0. This property serves as an internal balancing mechanism between excitation and inhibition.

The previous version of the SOLI model used a fixed network size and random uniform initialization of the network within the possible data range [16]. This initialization strategy does not scale well with data of high dimension or very sparse distributions. Such initialization for a high dimensional problem would require very high computational resources to fully cover the sparse high dimensional space. The enhanced SOLI model used in this research addresses these limitations.

The training algorithm for the enhanced model is as follows:

1. Initialize the network with α% of data points from the data set, $\alpha \in [1, 100]$ and set epoch=1;
2. Set learning rate, $\eta = \text{epoch}^{\beta}$, $\beta \in [0, -\infty]$;
3. Present an input pattern x at random;
4. Calculate node activation using the activation function $h_{Ui}(x, u_i)$ as defined above;
5. Update weight vector w_{Ui} of node u_i as $\Delta w_{Ui}=-\eta \times h_{Ui}(x, u_i)\times(x-w_{Ui})$;
6. epoch++;
7. Merge any nodes that are close enough, i.e., distance(u_i, u_j) < (ε%×σ), $\varepsilon \in (0, 100]$;
8. Return to step 2 until converge.

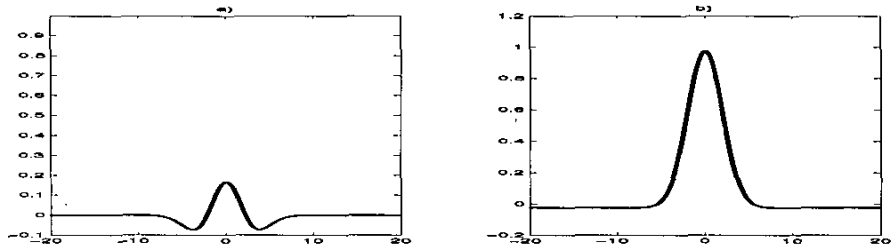

Fig. 1. The Mexican hat function. σ=2 for both graphs while for the graph on the left λ=1.2 and for the graph on the right λ=40.

This enhanced model has the following features:

1. The random initialization of the network in step 1 alleviates, to certain extend, the scaling problem in the case of data with high dimensionality or sparseness. The parameter α does not affect the final convergence of the network in terms of the final convergence epoch.
2. The learning rate (η) is decreased as the epoch increases, and β controls the rate of the decrease. This provides the convergence of the network.
3. Neurons may merge during training so that we do not need to pre-specify the network size and we can automatically represent the cluster centers with neurons at the convergence stage. ε controls the rate of merge.

In the current SOLI, we have four training parameters, α, β, ε and σ, among which only σ plays a decisive role in the final clustering result. α controls the initial network size and influences the quality of random initialization. β and ε together affect the speed of training. These two values should be compatible, i.e., the more negative β value should be companied by bigger ε value. β = -0.5, and ε = 50 are recommended as default values and used in this research. The recommended default values for these parameters are usually sufficient for many applications.

IV. Validity Indices

In general, there are three types of validity indices (VIs). The first type is based on statistical measures that test the degree of match between data and the discovered structures [10]. The second type is entropy based, in which case a good clustering should maximize the clustering entropy [5, 8]. The third type is based on heuristics, which include the DB index [6], Dunn's index and its variants [3, 10], Xie-Beni index [7], and the most recent VI in [11]. These heuristic-based VIs are biased towards those clustering algorithms that minimize intra-cluster scatter and at the same time maximize inter-cluster scatters.

Upon examination of these indices, the general conclusion is that there is no VI appropriate for all situations [9-11]. This may be caused by the complexity of clustering problem, i.e., the size, shape, density, degree of overlap of clusters, and the geometric arrangements among clusters [11]. The best strategy is to use many well-known indices together to have more confidence in the general results, obtained by a majority vote from different VIs [10].

In this research, four representative VIs were used, including three heuristic-based and one entropy-based VI. The heuristic-based VIs are the Xie-Beni index [7],, VI_{xb}, the DB index in [6], VI_{db}, and the index proposed in [11], VI_{ik}. The enropy-based VI was the Beni-Liu index [8], VI_{bl} . No statistical-based VIs were used as it has been found that the determination of the best value of the statistical-based indices tends to be subjective [10].

The proper number of clusters is found at the local minimum points on VI_{xb} and VI_{db}, and at the local maximum points on VI_{bl} and VI_{ik}. For this research, all the VI values were normalized to [0, 1]. The VI_{xb} and VI_{db} values were further inverted to $1-VI_{xb}$ and $1-VI_{db}$, respectively. Therefore, on the VI curve plots for this research, we are looking for the maximum values for all VIs.

V. Data Sets and Experimental Design

Two synthetic data sets, DATA1 and DATA2, and two benchmark data sets, the Breast Cancer data set (BR) and the Liver data set (LV) [22], were used to illustrate the efficacy of the SOLI model. DATA1 is designed to have significant overlap among data clusters. It is composed of four identical Gaussians, located at the three points of an equilateral triangle with the fourth at its center. DATA2 is designed to have two distinct levels of cluster structure. It is composed of four Gaussians with two well-separated subgroups, each of which is composed of two overlapping Gaussians. It is designed to test the ability of SOLI to discover the inherent structures at different levels (scale or size of clusters). Both DATA and DATA2 have 2000 data points with two attributes. Table I shows the means and the standard deviations of the original clusters for DATA1 and DATA2.

The BR data set has 699 data points with nine attributes and is known to be an easy data set in terms of classification accuracy [2]. The LV data set has 345 data points with six attributes but is known as a notoriously difficult data set to clustering accurately. The known best accuracy for the LV data set is not more than 70%. This research also

investigated the fundamental differences between these two data sets.

TABEL I
DESCRIPTION OF DATA1 AND DATA2

DATA1				DATA2			
μ_x	μ_y	σ_x	σ_y	μ_x	μ_y	σ_x	σ_y
-2.928	-0.037	1.035	1.013	-0.732	0.009	0.502	0.490
3.072	-0.037	1.035	1.013	0.768	0.009	0.502	0.490
0.072	5.159	1.035	1.013	0.018	6.009	0.502	0.490
0.072	1.695	1.035	1.013	0.018	7.509	0.502	0.490

For DATA1 and DATA2, α was set to 10, which is sufficient to prove our point. For BR and LV, α was set to 25 and 50 respectively. Since BR is an easy data set, we intentionally chose α to be smaller than that of LV data. We chose α=50 for LV to randomly initialize the network in the hope of covering most of the complexity within the LV data. For each data set, SOLI was run with different σ values, $\{\sigma_1, \ldots \sigma_m\}$, with the other parameters fixed. For each clustering result at σ_i, we calculate the values of the four different VIs.

Unsupervised clustering was performed for the BR and LV data sets. The classification accuracy (CA) is calculated as follows:

- Within each cluster, the class ID of the majority of the members is assigned to each member of this cluster.
- If there is a tie in the majority vote, the class ID is assigned as ND (non-determined).
- The newly assigned class IDs are compared to their original class IDs and any mismatch is counted as one error, including the mismatch with ND.

VI. RESULTS AND ANALYSIS

Experiment results for all four data sets are presented in Table II. In Table II, 'c' denotes the number of clusters discovered by our clustering method, and CA denotes the classification accuracy for the BR and LV data sets.

Fig. 2 and Fig. 3 present the result of clustering DATA1 and DATA2. In each figure, graph (a) is a plot of the clustering results (the cluster centers) and the original data used (in light grey), and graph (b) shows the VIs for that data set. The VI values plotted on part (b) of Fig 1, 2 are transformed values as mentioned in the previous section, in order to be plotted in a similar scale on the same plot.

For DATA1, the maximum plateaus are found on all VI curves, indicating c=4 to be the proper number of clusters for $\sigma \in [0.18, 0.43]$. An abnormal point is found at σ=0.225, c=6, where there is a sudden drop or "bulge" on the VI curves. This abnormality may be caused by the random initialization. Such dependency can be repressed or eliminated by repeated trials with an initialization at a given σ. The clusters discovered correspond correctly to the four Gaussians in DATA1. It is important to notice that the high degree of overlap among the clusters does not prevent SOLI from discovering the Gaussian centers correctly. Another relatively stable clustering structure is found at $\sigma \in$ [0.5, 0.8] with c=1, in which case the found cluster center approximates the geometric center of DATA1. At σ=0.45, SOLI found 3 clusters, an unstable structure as indicated by the VIs.

For DATA2, the most stable cluster structure is found at $\sigma \in$ [0.23, 0.75] with c=2. A less stable cluster structure is found at $\sigma \in$ [0.1, 0.15] with four clusters as indicated by the VIs. The stableness of the discovered cluster structure is indicated by VI values and the smaller relative range of σ and c. When c=2, the discovered cluster centers approximate the middle point between the two Gaussians within the same subgroup, while at c=4, cluster centers approximate the four Gaussian centers.

All of the VIs reach a consensus, with VI_{ik} being more sensitive to minor change in clustering results. The use of multiple VIs is also demonstrated to be an effective strategy as suggested in [10]. Experimental results on DATA1 and DATA2 demonstrate that SOLI is able to discover the correct clustering structure given a wide range of training parameters. SOLI is also able to discover cluster structures at different scales when used in combination with many VIs.

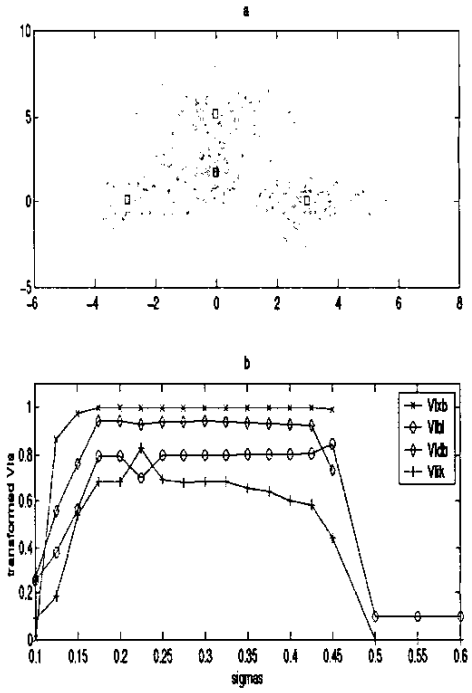

Fig. 2. DATA1 a. Data plot. Square denotes clusters centers at c=4, + denotes clusters centers found at c=1. b. Validity Indices plot.

TABLE II
EXPERIMENT RESULTS

DATA1		*DATA2*		*BR Data*			*LV Data*		
c	σ	c	σ	c	σ	CA	c	σ	CA
55	0.100	70	0.050	37	0.10	0.957	111	0.020	0.670
23	0.125	13	0.075	28	0.11	0.959	71	0.025	0.606
10	0.150	4	0.100	12	0.12	0.967	41	0.030	0.617
4	0.175	4	0.125	7	0.13	0.966	26	0.035	0.600
4	0.200	4	0.150	3	0.14	0.964	24	0.040	0.591
6	0.225	4	0.175	2	0.15	0.963	12	0.045	0.542
4	0.250	3	0.200	2	0.16	0.900	10	0.050	0.591
4	0.275	2	0.225	1	0.17	0.655	9	0.055	0.586
4	0.300	2	0.250	1	0.18	0.655	7	0.060	0.583
4	0.325	2	0.275				2	0.065	0.580
4	0.350	2	0.300				2	0.070	0.580
4	0.375	2	0.325				4	0.075	0.583
4	0.400	2	0.350				2	0.080	0.580
4	0.425	2	0.375				2	0.085	0.580
3	0.450	2	0.400				2	0.090	0.580
1	0.500	2	0.425				2	0.095	0.580
1	0.550	2	0.450				1	0.100	0.580
1	0.600	2	0.475				1	0.105	0.580

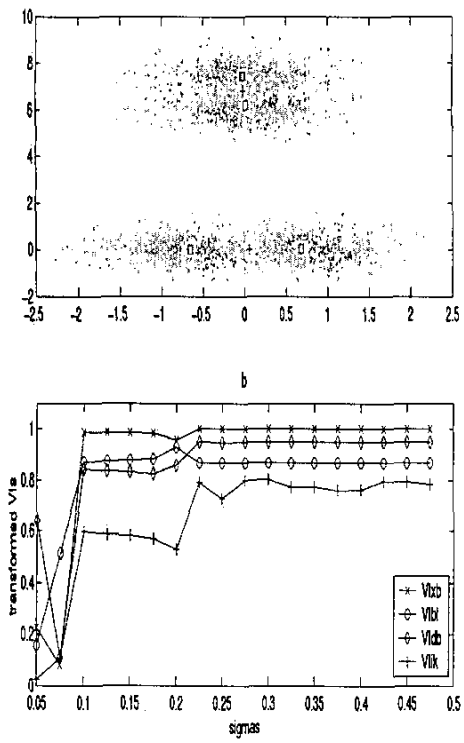

Fig. 3. DATA2. a. Data plot. Square denotes clusters centers at c=4, + denotes clusters centers found at c=2. b. Validity Indices plot.

In Fig. 4, we compiled the VI curves and CA curves for the BR and LV data. For BR data, in Fig 4a, a stable cluster structure of two clusters is found at $\sigma \in [0.15, 0.16]$, and validated by the VIs. The classification accuracy at c=2 reaches 96% accuracy in Fig 4c. A detailed check on the clustering result at σ=0.14, c=3, (a less stable clustering result), indicates that BR is composed of two major clusters. One large cluster is composed of 461 points, while the second cluster consists of two smaller clusters composed of 84 and 154 points, respectively. Reducing the size of σ (increasing the number of clusters) has no noticeable effect on CA. This leads us to believe that BR naturally has 2 major clusters, which are quite homogeneous in terms of their class IDs.

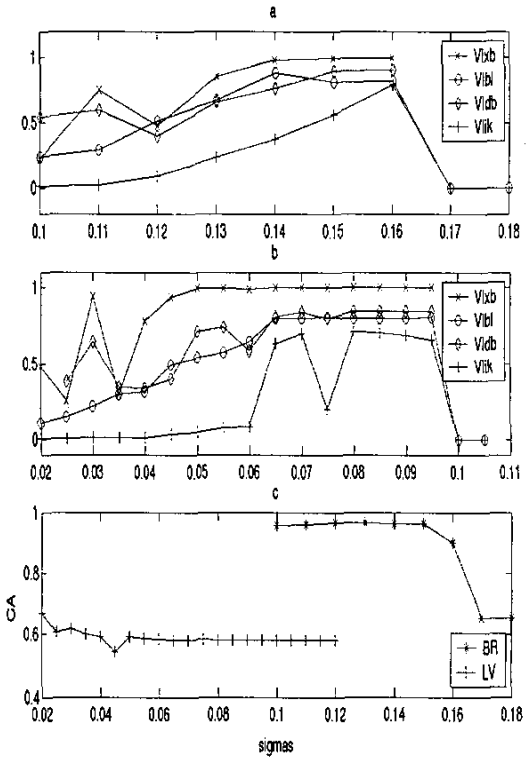

Fig. 4. a. VI plot for BR. b. VI plot for LV c. Classification accuracy plot of BR and LV.

For the LV data set, in Fig 4b, a stable cluster structure of two clusters is found at $\sigma \in [0.065, 0.095]$ and validated by the VIs, with the exception of an abnormal point at σ=0.075, c=4. The two clusters correspond to the two classes with relatively low CA values. Notice that CA remains constant across a wide range of c and σ. Only when the number of clusters is 111, i.e., the average cluster size is near three, does the CA reach 67% in Fig. 4c. This indicates that LV has two natural clusters but that within each natural cluster, the class boundaries are very complex. Only when the sub-

level cluster size is reduced to around three members on average, do the sub-level clusters become slightly more homogeneous in terms of their class IDs. This very complex class boundary may be the major factor for the LV data set being so notoriously difficult to classify.

Through all the experiments over all the data sets, it is noticeable that most of the VI curves are consistent and show very stable behaviors (a maximum plateau instead of a maximum point over all the VI curves). This is rather a reflection that the clustering result is stable and robust, considering the fact that such clustering results are obtained over a wide range of training parameters with random initialization of the network. This observation further points out the importance of a robust clustering algorithm over the choice of a VI for the clustering problem in general.

Using DATA1 and DATA2, we have demonstrated that SOLI is able to discover the hierarchical clustering structure that is inherent for the given data set. Such hierarchy recognition is robust even when the underlying data clusters have serious overlap. The discovery of such a hierarchy is guided and verified by the VIs, rather than formed by imposing some heuristic rules on the clustering algorithm.

VII. Conclusions and Future Directions

This research has shown that recent enhancements of the SOLI model results in a robust clustering method such that stable cluster structures can be discovered over a wide range of training parameters, even when there is serious overlap among the clusters. When combined with various validity indices, SOLI is able to discover stable clustering structures under different scales. This methodology can be used to discover the natural hierarchy of a given data set with an understanding of the scale of structure in each layer of the hierarchy. Using SOLI, we have discovered the homogeneous nature of the BR data set, and the highly heterogeneous nature of the LV data set. Such information is essential for some neural network designs that require this information to establish the number of neurons in the hidden layer, the proper number of hidden layers [1], and the proper number of experts in the Mixture of Experts model [2].

In SOLI, the network size is monotonically decreasing as nodes are allowed to merge during training. For a more dynamic environment, an incremental version of SOLI should be implemented. The random initialization scheme used in the current SOLI cannot guard against a bad initialization, even with a relatively large value of α. The potential for a bad initialization would be further reduced in an incremental version of SOLI. Future research also includes the application of SOLI to text mining in large text data sets.

References

[1] S. Haykin, *Neural networks, a comprehensive foundation.*, Macmillan College Publishing Company, New York. 1994.

[2] B. Tang, M.I. Heywood, and M. Shepherd, "Input partitioning to mixture of experts," *IJCNN '02. Proceedings of the 2002 International Joint Conference on Neural Networks*, vol. 1., pp. 227 -232, 2002.

[3] J.C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *J.Cybern.*, vol. 3, no. 3, pp.32-57, 1973.

[4] J.C Bezdek, "Cluster validity with fuzzy sets," *J.Cybern*, vol. 3, no. 3, pp. 58-72, 1974.

[5] J.C Bezdek, "Mathematic models for systematic and taxonomy" in *Proc. 8th Int. Conf. Numerical Taxonomy*, G.Estabrook, Ed., Freeman, San Francisco, CA, pp. 143-166, 1975.

[6] D. L. Davies and D.W. Bouldin, " A cluster separation measure," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 1, no. 4, pp. 224-227. 1979.

[7] X. L. Xie and G. A. Beni, "Validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 3, no. 8, pp. 841-846, 1991.

[8] G. Beni and X. Liu, " A Least Biased Fuzzy Clustering Method", *IEEE Transactions on Pattern Anal. Machine Intell.*, vol. 16, no. 9, pp. 954-960, 1994.

[9] N.R. Pal and J.C. Bezdek, "On clustering validity for the fuzzy c-means model," *IEEE Trans. Fuzzy Syst*, vol. 3, no. 3, pp. 370-379, 1995.

[10] J. C. Bezdek and N. R. Pal, "Some New Indexes of Cluster Validity," *IEEE Transactions on Systems, Man, and Cybernetics, Part B.* vol. 28, no. 3, pp. 301-315, 1998.

[11] S. Bandyopadhyay and U. Maulik, "Nonparametric Genetic Clustering: Comparison of Validity Indices," *IEEE Transactions on System, Man, and Cybernetics, Part C.* vol. 31, no. 1, pp. 120- 125, 2001.

[12] M. Halkidi, Y. Batistakis, M. Vazirgiannis, "Clustering algorithms and validity measures". Tutorial paper, *Proceedings of SSDBM Conference*, pp. 3-22, July 2001.

[13] E.Rasmussen, "Clustering Algorithms" in W.B.Frakes, R.Rocardo-Yates ed. *Information Retrieval, Data Strucutures & Algorithms.*, pp419-442. Eaglewood Cliff, NJ: Prentice Hall , 1992.

[14] T. Kohonen., "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, pp. 59-69, 1982.

[15] B. Fritzke, "Some competitive learning methods, Draft Doc." http://www.neurolin-formatik.ruhr-unibochum.de/ini/VDM/research/gsn/DemoGNG, 1998.

[16] B. Tang and M. Shepherd, "Achieving Self-Organization by Lateral Inhibition," *ICMLC02, Proceedings of the 2002 International Conference on Machine Learning and Cybernetics*, vol. 1., pp. 1935-1940, 2002.

[17] J. Sirosh and R. Miikkulainen, "Topographic Receptive Fields and Patterned Lateral Interaction in a Self-Organizing Model of the Primary Visual Cortex." *Neural Computation*, vol. 9, pp. 577-594, 1997.

[18] M. Pötzsch and C. von der Malsburg,"Self-organization of networks in the visual system," In G. Dorffner, K. Möller, G. Paaß, and S. Vogel, Eds, *Konnektionismus und Neuronale Netze*, vol 272 of GMD-Studien, pp. 299-315, Münster/Westf., 1995

[19] L. Itti, C. Koch, "A Comparison of Feature Combination Strategies for Saliency-Based Visual Attention Systems." *SPIE Human Vision and Electronic Imaging IV* (HVEI'99), San Jose, CA, vol, 3644, pp. 373-382, 1999.

[20] Y. Wang, I. Fujita, and Y. Murayama. "Neuronal mechanisms of selectivity for object features revealed by blocking inhibition in inferotemporal cortex," *Nature neuroscience.* vol. 3, no. 8, pp. 807-813, 2000.

[21] S. Grossberg and J. R. Williamson, "A Neural Model of how Horizontal and Interlaminar Connections of Visual Cortex Develop into Adult Circuits that Carry Out Perceptual Grouping and Learning", *Cerebral Cortex*, vol. 11, pp. 37-58, 2001.

[22] http://www.ics.uci.edu/~mlearn/MLRepository.html