A NOVEL SELF-ORGANIZING NETWORK DETERMINES THE PROPER CLUSTER STRUCTURE AUTOMATICALLY

Bin Tang, Malcolm Heywood, Michael Shepherd
Dalhousie University, Faculty of Computer Science
Halifax, Canada
{btang, mheywood, shepherd}@cs.dal.ca

Abstract – *In this paper, we describe a novel self-organizing neural network model, the self-organization by balanced excitatory and inhibitory input model (SOBEII). Using balanced excitation and inhibition and anti-Hebbian learning strategy, SOBEII is capable of automatically determining the proper cluster structure of given datasets in a robust manner. This is demonstrated using both synthetic and real datasets. SOBEII results match those of EM, however, SOBEII is not sensitive to adverse initialization conditions or outliers in contrast to many conventional clustering methods.*

Key words – self-organization, balanced excitation and inhibition, automatic model selection

1 Introduction

Cluster analysis is a fundamental and generic tool used to find natural groupings inherent in multivariate datasets, which has found a wide range of applications in diverse fields. Clustering can be viewed as a mode-seeking procedure, i.e., to find a suitable number of data dense regions (clusters) into which to partition the data. Most conventional clustering methods, such as the *k*-means, FCM, and EM algorithms [1, 3, 4], assume that the suitable number of clusters (*k*) for the given dataset is known (rarely true), and only partition the data into *k* regions. There are other well-known issues associated with these conventional methods, namely, sensitivity to initialization, noise and outliers.

In practice, for most conventional clustering algorithms, the mode seeking for the suitable number of clusters (*k*) is determined in a two-step process. First, we need to build a series of candidate models for each potential *k* within a large range of *k*, and then select the best model based on some model selection criteria. Some of the commonly used criteria include Akaik's information criterion (AIC), Baysian inference criterion (BIC), and the consistent Akaik's information criterion (CAIC) [7]. Due to the sensitivity issues of the conventional methods, we often need to perform multiple runs of clustering at each given *k* to obtain a stable clustering result, which makes the total mode-seeking procedure computationally expensive or even potentially prohibitive.

Kohonen's self-organizing feature map (SOM), inspired by the organizational principles of biological neural systems, has become a standard tool for cluster analysis, data mining and visualization over a wide range of applications [2, 8, 9]. The incremental variants of SOM, e.g., GNG and RGNG algorithms [5, 10], have shed new light on configuring the appropriate number of clusters in an automatic fashion. However, the final number of nodes (clusters) is somewhat influenced by the network parameters. Although differing in the details, all SOM-like algorithms

share common principles, i.e., a winner-take-all network dynamic and a Hebbian learning strategy with collaboration within a near neighborhood (usually modeled by a Gaussian).
The Rival Penalized Competitive Learning (RPCL) algorithm, an extension of SOM, is capable of automatically finding the suitable number of clusters [13]. Contrary to the second principle of SOM, RPCL introduces local competition and utilizes anti-Hebbian learning during neighborhood adaptation. RPCL has been incorporated into the Baysian Ying-Yang Harmony Learning framework (BYY), with many improvements on robustness [7, 14, 15].
Recently, a novel self-organizing model, Self-organization by Lateral inhibition (SOLI) has been proposed as a further extension of SOM and RPCL. However, SOLI is distinctively different [11, 12]. The first fundamental difference from SOM lies in that SOLI is a winner-less competitive network. Each node in SOLI is associated with an activation function (functionally similar to the neighborhood function in SOM), which is defined by a balanced Mexican Hat (MH) function and implemented as the difference of two Gaussians. For a balanced MH function, there are areas naturally under positive/negative influences due to different degrees of excitation/inhibition. In reality, for a given input data point x, some nodes get positively activated (excited) while others negatively activated (inhibited). Thus, there is no need to decide winners or loser(s) explicitly. The second major difference from SOM and RPCL is that SOLI uses anti-Hebbian learning to update the receptive fields for all nodes. It has been demonstrated empirically that SOLI is capable of robustly mimicking the data dense areas using identical Gaussian clusters; when combined with model selection modules (clustering validity indices), SOLI is capable of discovering suitable cluster structures [11, 12]. However, such a discovery is not fully automatic. The limitation comes from the fact that the shapes and sizes of the Mexican Hat function for all nodes remain identical and unchanged during the learning process of SOLI.
In this paper, we extended the SOLI model to a new model, properly named as, the Self-Organization by Balanced Excitatory and Inhibitory Input model (SOBEII). During learning, SOBEII is allowed to learn not only the receptive fields, but also the exact shapes and sizes of the MH functions of all nodes. The learning of MH functions allow the nodes to model exact sizes of Gaussian clusters, not as defined parameters as in SOLI. In the course of learning, different nodes in SOBEII may be drawn into the same data dense area and become almost identical in terms of defining parameters, which will be merged together. Randomly starting from a large number of nodes, the learning process of SOBEII is driven solely by the density distribution of the given dataset. Eventually, all SOBEII nodes converge to a few local density centers (true number of clusters). Currently, SOBEII is designed to automatically determine the proper number of clusters in a robust manner, if data are generated by a mixture of spherical Gaussian components. Compared to most conventional methods, SOBEII is not sensitive to initialization or outliers.
This paper is organized as follows. In Section 2, we describe the implementation details of SOBEII. In Section 3, we demonstrate how SOBEII performs the automatic model selection task in a robust manner and how it compares with the conventional two-phase model selection approach using EM. We conclude and point out the future research in Section 4.

2 Introduction to SOBEII

Recent biological findings show that balancing inhibitory and excitatory inputs plays a fundamental role in shaping the selectiveness of the neurons and formation of topologically ordered organization of biological neural systems [16, 17]. Both SOLI and SOBEII are inspired by such principles. The core computational component for both SOLI and SOBEII is the balanced MH function, which is the incarnation of balanced excitatory and inhibitory inputs principle.

In terms of structure, SOBEII is composed of a set of neurons, with no lateral connections among them. Each neuron is defined by its receptive field u_i and a distinct activation function, MH_i, in the form of a Mexican Hat. For given input x, MH_i is defined as

$MH_i(x, u_i)=exp(-dist(x, u_i)^2/2\sigma_i^2)-exp(-dist(x, u_i)^2/2(\lambda_i\sigma_i)^2))/\lambda_i$,

where $\lambda_i \in (1, \infty]$, σ_i corresponds to a spherical covariance matrix associated with a Gaussian cluster, while $dist(\cdot)$ denotes a distance function (Euclidean distance is used in this research). It can be easily proven that the integration of MH_i over the range $[-\infty, \infty]$ on distance, $dist(\cdot)$, is 0. This property serves as an internal balancing mechanism between excitation and inhibition from inputs.

During learning, we update u_i, λ_i and σ_i for each node i after each input x using the following rules.

$$\Delta u_i=-\eta_\mu \times MH_i(x, u_i)\times(x- u_i); \quad (1)$$

$$\Delta\sigma_i=-\eta_\sigma \times MH_i(x, u_i); \quad (2)$$

$$\Delta\lambda_i=\eta_\lambda \times MH_i(x, u_i); \quad (3)$$

We define three global learning rates, η_μ, η_λ and η_σ respectively. In general, η_μ is much bigger than η_σ. In practice, we recommend the following values (also used in this work), η_μ=0.1, η_σ=0.01 and η_λ is set to be $\eta_\sigma^{1.5}$.

A pseudocode of the training procedure is defined below:

1. Initialize the network with m data points selected from the data set, $m >> k^{ideal}$,
2. Rough training of network parameters, mainly to stabilize u_i and converge to k-nodes
   while *maxEpoch* is not reached,
      within each epoch, for each input x, update network using rules (1, 2, 3) in sequence;
      merge nodes i and j if dist(u_i, u_j) < 0.1σ, where σ=min(σ_i, σ_j);
      η_μ reduced as η_μ*exp(-1/maxEpoch);
      IF no merging nodes occuring in the current epoch,
         $\Delta U = \arg\max_i \left(dist\left(\boldsymbol{u}_i^{current\_epoch}, \boldsymbol{u}_i^{previous\_epcoh} \right) \right)$;
         IF (ΔU <0.01) THEN (break);
      END
   end
3. Fine tuning of the network parameters
   Partition data into corresponding clusters;
   For each subset of data, X_i, that belongs to cluster i,
   while 1, // parameters not converged,
      run one epoch training of the network parameters using X_i,
      IF ($abs(\sigma^{current\_epoch}-\sigma^{previous\_epoch})/\sigma^{previous\_epoch} < 0.01 \& \lambda^{curren\_epoch} < 1.001$) THEN (break);
      end
4. Repeat step 2 and 3 once more to make sure of convergence and absorb outlier nodes.

3 Experimental Results

In this section, we demonstrate the relative robustness of SOBEII for the automatic determination of the proper number of clusters. We apply SOBEII to one difficult 2D synthetic dataset and two UCI real datasets. For the synthetic dataset, we compare SOBEII results against the known solution; in addition, values of different model selection criteria are computed. The overall results of SOBEII are further compared to that of EM algorithm. For the real datasets, besides the model selection task, we also investigate how we can utilize SOBEII to analyze the cluster organization for the data from each class. For all experiments, SOBEII converges within 250 epochs.

3.1 Synthetic data

We have experimented with SOBEII on many synthetic datasets of different degree of difficulties. Due to the space limit, here, we only report the results for one very difficult dataset, Data 1. Data 1 is composed of 19 spherical Gaussian clusters, arranged in a concentric three-ring shape with 19,600 points. It is designed with three grades of sizes of clusters (0.5, 1, and 2) and three grades of density levels (ρ), i.e., $\rho_{\text{inner-ring}} : \rho_{\text{middle-ring}} : \rho_{\text{outer-ring}} = 1 : 64 : 4$. The drastic density difference and geometric arrangements of clusters are designed to challenge the clustering algorithm.
For the synthetic dataset, we ran SOBEII 100 times, each starting with a different random initialization. We observe that SOBEII converges consistently to only a few possible cluster structures. We select the structure that SOBEII most frequently converges to, as the proper cluster structure recognized by SOBEII. For each structure converged to, we compute the model selection criteria values (AIC, BIC, CAIC) averaged over all the runs. We also ran EM 100 times for the given range of possible clusters (usually from 1 to $2k^{ideal}$). To determine the proper k value for EM, the three model selection criteria values are computed and averaged at each k.

Table 1. Synthetic data summary

Gaussian component parameters				SOBEI results		
x_1	x_2	σ	n_k	x_1	x_2	σ
0.000	0.000	2.00	400	0.096	0.081	1.27
3.750	6.495	0.50	1600	3.745	6.480	0.51
6.852	3.051	0.50	1600	6.886	3.056	0.52
7.336	-1.559	0.50	1600	7.369	-1.572	0.49
5.018	-5.574	0.50	1600	5.033	-5.557	0.50
0.784	-7.459	0.50	1600	0.764	-7.457	0.49
-3.750	-6.495	0.50	1600	-3.756	-6.538	0.47
-6.852	-3.051	0.50	1600	-6.868	-3.045	0.51
-7.336	1.559	0.50	1600	-7.338	1.558	0.49
-5.018	5.574	0.50	1600	-4.985	5.591	0.51
-0.784	7.459	0.50	1600	-0.800	7.459	0.49
4.592	11.087	1.00	400	4.598	10.956	0.96
11.087	4.592	1.00	400	11.053	4.664	1.02
11.087	-4.592	1.00	400	11.099	-4.666	0.94
4.592	-11.087	1.00	400	4.600	-11.089	0.82
-4.592	-11.087	1.00	400	-4.537	-11.157	0.90
-11.087	-4.592	1.00	400	-11.005	-4.659	0.96
-11.087	4.592	1.00	400	-11.148	4.547	0.98
-4.592	11.087	1.00	400	-4.611	11.169	0.96

Note: (x_1, x_2) denotes the centroid positions, σ represents the size of the cluster, n_k denotes the cardinality of the cluster.

Data 1 Results

The overall results are compiled in Fig 2. In Fig 2.a, SOBEII converges consistently to a 19-cluster structure with 87% convergence frequency. In Fig 1, we show a typical running result of 19 clusters with its numerical results reported in Table 1**.** We can see that the SOBEII results are very close to the theoretical Gaussian model parameters as shown in Table 1. In Fig 1, notice that among all the initial positions of neurons, many are located far from the data dense areas with some falling into "outlier" area or areas between data clusters. Even with such an initialization condition,

SOBEII is capable of converging to data dense area with the proper number of clusters. We observed similar behavior of SOBEII on many other synthetic datasets, which leads us to believe that SOBEII is not very sensitive to initialization conditions, outlier or noise in contrast to many conventional methods.

The three model selection criteria all agree that the 19-cluster structure obtained by SOBEII is the proper cluster structure for Data 1 as in Fig 2.b. Based on the EM results, none of the model selection criteria is able to recognize the ideal cluster structure in Fig 2.c. They all recognize the proper cluster structure to be about 21 to 23 clusters.

Checking the detailed convergent stages of EM, for the cases of $k > 19$, EM often converges to some trivial clusters or tries to split one real cluster into multiple sub-clusters. There are two contributing factors, first being EM's sensitivity to initialization and outliers, and the second is due to the partitional nature of EM. Therefore, for EM, simply removing bogus cluster centers will not solve all the problems that are rooted in the partitional nature of EM.

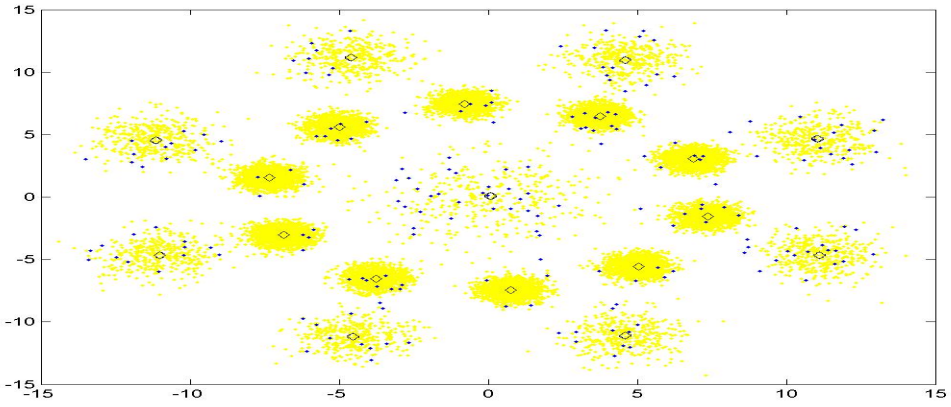

Fig 1. Data 1 running result of run 67. The light dots represent data points, the dark dots represent the initial nodes positions, and '◊' denotes the converged node positions.

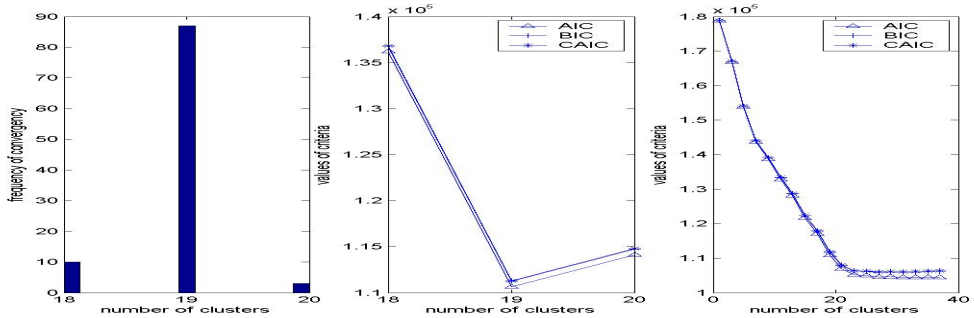

Fig 2. Results summary for Data 1. a). Histogram of converged number of clusters. x-axis: number of clusters, y-axis, frequency of convergence. b). Model selection for SOBEII results. x-axis: number of clusters, y-axis, values of model selection criteria. c). Model selection for EM results. x-axis: number of clusters, y-axis, values of model selection criteria.

3.2 UCI datasets

The two real datasets, Iris and Wine datasets are from the UCI datasets collection [18]. To understand the correlation between the obtained clustering structure and the known class distribution, the measure of classification accuracy is calculated as follows. For each run, we randomly split the data into training and testing sets with ratio of 9:1, we ran the clustering

algorithms on the training set, from which we assigned class labels to each cluster based on a majority vote of its members. The classification accuracy is computed only for the test data as the percentage of points whose class labels match the class label of its owner cluster (to reduce the bias from training set).

Iris data

The Iris data is a well-known dataset for cluster analysis. It has 3 classes (of 50 points per class), distributed in a 4-dimensional space. It is known that the one class can be linearly separated from the other two classes, and the latter two classes are not linearly separable. Traditionally, a 2-cluster structure has been identified as the most stable cluster structure [6].

The SOBEII results are reported in Fig 3 based on 150 runs. The most frequent convergent stage for SOBEII corresponds to a 2-cluster structure at (70% of the 150 runs). The model selection results in Fig 3.b support a 2-cluster proper cluster structure for SOBEII. Interestingly, the classification accuracies in Fig 3.c for 2 and 3 clusters remain almost the same. SOBEII results suggest that a 2-cluster structure is a good choice for proper cluster structure.

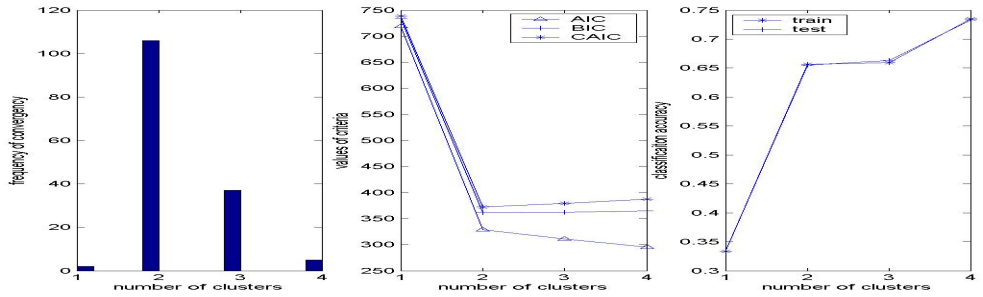

Fig 3. SOBEII result summary for Iris data. a). Histogram of converged number of clusters. x-axis: number of clusters, y-axis, frequency of convergence. b). Model selection for SOBEII results. x-axis: number of clusters, y-axis, values of model selection criteria. c). Classification accuracy. x-axis: number of clusters, y-axis, values of model selection criteria.

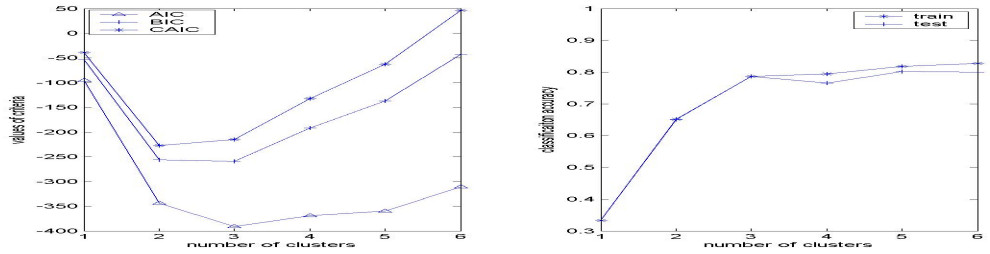

Fig 4. EM result summary for Iris data. a). Model selection for EM results. x-axis: number of clusters, y-axis, values of model selection criteria. b). Classification accuracy. x-axis: number of clusters, y-axis, values of model selection criteria.

The EM result is shown in Fig 4. Both AIC and BIC curves in Fig 4.a suggest 3 to be the proper number of clusters, while CAIC curve prefers 2 to be the proper value. However, the difference between a 2-cluster structure and a 3-cluster structure is not that obvious even for AIC and BIC. The classification results in Fig 4.b for the 3-cluster structure are significantly better than for the 2-cluster structure. Overall, the EM results prefer a 3-cluster structure.

The difference between EM results and SOBEII results can be explained by the inferior assumption of spherical covariance matrices used in SOBEII comparing to the arbitrary covariance matrices used in EM. This is confirmed by checking the detailed EM results for the 3-cluster solutions.

Wine data

The Wine dataset has 3 classes, distributed in a 13-dimensional space. In [10], GNG with model selection and Robust GNG (RGNG) obtained the proper cluster number to be between 3 and 4.

We ran SOBEII 250 times, and results are summarized in Fig 4. Fig 4.a suggests two proper cluster structures with almost equal likelihood (3 or 4). In Fig 4.b, the BIC and CAIC curves seem unable to determine the proper number of clusters, while the AIC curve suggests 3 or 4 to be the proper number of clusters. In Fig 4.c, a 4-cluster structure apparently provides a 5% absolute accuracy increase comparing to that of 3-cluster structure, on average.

The EM results are summarized in Fig 5. As with SOBEII, it is impossible to use model selection criteria to determine the proper number of clusters using EM results as shown in Fig 5.a. The classification accuracy results in Fig 5.b suggest a peak performance at 4 clusters.

Through the experiments, we notice that the EM performances are very unstable compared to SOBEII, which may be attributed to the sensitivity issues of EM. The unstableness of EM is reflected in the larger classification difference between training and testing sets and the drop of classification accuracy after 4-cluster as compared to that of SOBEII. We also noticed that the absolute values of classification accuracy of SOBEII are much better than that of EM.

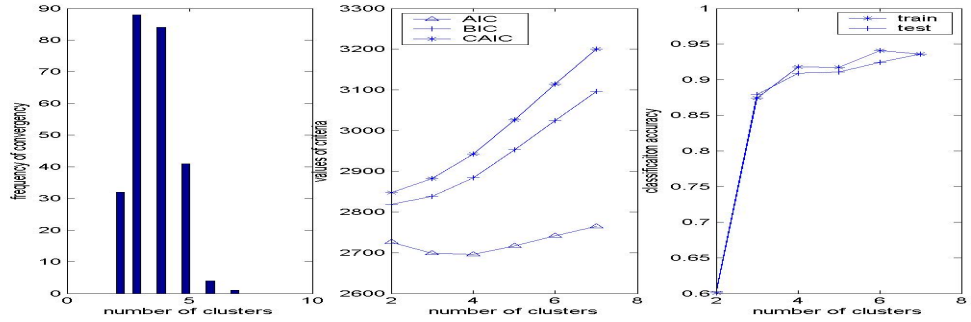

Fig 5. SOBEII result summary for Wine. (The full legends are the same as Fig 3, omitted here)

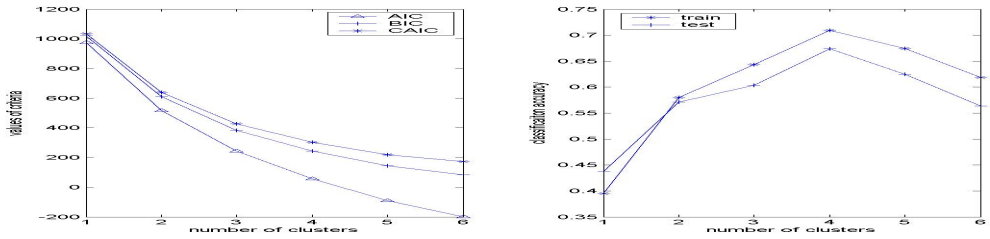

Fig 6. EM result summary for Wine data (The full legends are the same as Fig 4, omitted here)

4. Conclusion and future work

This work presents the SOBEII model, a self-organizing neural network model that is capable of automatically determining the cluster structure of a given dataset. Different from many SOM related models, it uses a balanced Mexican Hat function to model the activation of each node, while using anti-Hebbian learning to learn its receptive fields and the model parameters related to MH functions. Through experiments on both synthetic and real datasets, we have demonstrated that SOBEII can converge to the proper cluster structures with high probability in a robust manner. It is also observed that SOBEII is not sensitive to the initialization conditions and outliers in contrast to many conventional clustering methods. Compared to EM, the results of SOBEII can match those of EM for most cases. However, the performance of SOBEII is more stable than that of EM. In terms of computation, for all the datasets, SOBEII converged within 250 epochs, which is rather efficient.

Currently, the formulation of SOBEII only allows modeling of spherical covariances of Gaussian clusters. This satisfies most of the applications in data mining with medium to high dimensionality, as with Wine dataset (due to the central limit theorem). However, it is this limitation that allows EM to out perform SOBEII for the Iris data, which requires modeling with arbitrary covariance matrices. In the future, we will extend SOBEII to allow modeling arbitrary covariance matrices. So far, theoretical analysis of this method has been lacking. In the future, we will investigate how SOBEII can be related to other theoretical frameworks for cluster analysis and its convergence property.

References

[1] J.C. Bezdek. (1981), *Pattern recognition with fuzzy objective function algorithms*, New York, Plenum.

[2] M.Cottrell, S. Ibbou & P. Letrémy, (2004), SOM-based algorithms for qualitative variables, *Neural Networks*, **vol, 17**, p.1149-1167.

[3] A. P. Dempster, N.M. Laird, and D.B. Rubin, (1977), Maximum-likelihood from incomplete data via the em algorithm, *J. Royal Statist. Soc. Ser. B.*, **vol. 39**, p. 1-38.

[4] R.O. Duda, & P.E. Hart, (1973), *Pattern Analysis and Scene Analysis*, John Wiley & Sons.

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

[6] L.O Hall, I.B. Özyurt, and J.C. Bezdek, (1999), Clustering with a genetically optimized approach, *IEEE. Transactions on Evolutionary Computation*, **vol. 3, no. 2,** p103-112.

[7] X. Hu, & L. Xu, (2004), Automatic Cluster Number Determination via BYY Harmony Learning, *Proceedings of International Symposium on Neural Networks*, **part 1**, p 828-833.

[8] T. Kohonen. (1989). *Self-Organization and Associative Memory*, Springer-Verlag, Berlin, Third Edition.

[9] T. Kohonen, T. Kohonen, S. Kaski, K. Lagus, J. Salojärvi, J. Honkela, V. Paatero and A. Saarela, (2000). Self organization of a massive text document collection, *IEEE Transactions on Neural Networks*, **vol 11, no. 3**, p. 574-585.

[10] A.K. Qin & P.N. Suganthan. (2004). Robust growing neural gas algorithm with application in cluster analysis, *Neural Networks*, **vol, 17**, p.1135-1148.

[11] B. Tang and M. Shepherd, (2002), Achieving Self-Organization by Lateral inhibition, *Proc . ICMLC 2002*, **vol. 1**, p.1935-1940.

[12] B. Tang, M.I. Heywood & M. Shepherd, (2003), The self-organization by lateral inhibition model: validation of clustering. *Proc. IJCNN03*, **vol, 1**, p. 781-786.

[13] L. Xu, A. Krzkžak, E. Oja, (1993), Rival penalized competitive learning for clustering analysis, RBF net, and curve detection, *IEEE. Trans. Neural Networks*, **vol. 4**, p. 636-648.

[14] L. Xu, (1995), Bayesian-Kullback Coupled YING-YANG Machines: Unified Learnings and New Results on Vector Quantization, *Proc. NIPS*, p. 977-988.

[15] L.Xu, (2002), BYY harmony learning, structural RPCL, and topological self-organizing on mixture models, *Neural Networks*, **vol. 15,** p.1125-1151.

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

[17] Y. Wang, I. Fujita, H. Ramura, and Y. Murayama. (2002), Contributions of GABAergic inhibition to receptive field structures of monkey inferior temporal neurons, *Cerebral Cortex*, **vol. 12, no. 1**, p. 62-74.

[18] http://www.ics.18.edu/~mlearn/MLSummary.html