IJCSE International Journal of Computer Sciences & Engineering Open Access
Research Paper Volume-3, Issue-7 E-ISSN: 2347-2693

Self-organizing Map with Modified Self Organizing Map Clustering

Kamalpreet Kaur Jassar[1]* and Dr. Kanwalvir Singh Dhindsa[2]

[1*,2]*Dept. of CSE, BBSBEC, Fatehgarh Sahib, PUNJAB, INDIA*

www.ijcseonline.org

Received: Jun/03/2015 Revised: Jul/10/2015 Accepted: July/24/2015 Published: July/30/ 2015

Abstract- Clustering is a very well-known technique of data mining which is mostly used method of analyzing and describing the data. It is one of the techniques to deal with the large geographical datasets. Clustering is the mostly used method of data mining. KohonenSOM is a classical method for clustering. In this paper, a new approach is proposed by combining neural network and clustering algorithms. We propose a modified Self Organizing Map algorithm which initially starts with null network and grows with the original data space as initial weight vector, updating neighbourhood rules and learning rate dynamically in order to overcome the fixed architecture and random weight vector assignment of simple SOM. In this paper, existing SOM and modified SOM have been compared by using different parameters.

Keywords-Clustering Algorithms, Learning rate, Weight vector, SOM, Modified SOM

I. INTRODUCTION

A self-organizing map (SOM) is a kind of artificial neural network that is trained using unsupervised learning to produce a low dimensional typically two dimensional as output. It is discretized representation of the input space of the training samples, called a map. Self-organizing maps are different than other artificial neural networks in the sense that they use a neighbourhood function to preserve the topological properties of the input space. The main set back of this technique, however, is that the number of output nodes is predefined and only the adjacent nodes are taken as neighbourhood [8]. SOM is a clustering method because it organizes the data in clusters (cells of map) such as the instances in the same cell are similar, and the instances in different cells are different. In this point of view, SOM gives comparable results to state-of-the art clustering algorithm such as K-Means [11]. SOM is also considered as data visualization technique because it allows us to visualize data in a low dimensional representation space (basically in 2D).

The Kohonen SOM algorithm is a very powerful tool for data analysis [21]. It was originally designed to model organized connections between some biological neural networks. It was also immediately considered as a very good algorithm to realize vectorial quantization, and at the same time pertinent classification, with nice properties for visualization [20]. Self-Organizing Maps (SOMs) have been used in GIScience both for clustering georeferenced data and for the specialization of various non-geographic datasets. The original SOM proposed by Kohonen does not take into account the particular role that geographic location has in most problems involving the clustering of geo-referenced data. In the original SOM algorithm, all variables are treated equally. When clustering geo-referenced data, spatial location is particularly important, since objects that are geographically far away should not be clustered together, even if they are similar in all other aspects. Although the term "Self-Organizing Map" could be applied to a number of different approaches, we shall use it as a synonym of Kohonen's Self Organizing Map, or SOM for short, also known as Kohonen Neural Networks. The basic idea of a SOM is to map the data patterns onto n-dimensional grid of neurons or units. That grid forms what is known as the output space, as opposed to the input space where the data patterns are. This mapping tries to preserve topological relations, i.e., patterns that are close in the input space will be mapped to units that are close in the output space, and vice-versa. So as to allow an easy visualization, the output space is usually 1 or 2 D.

II. RELATED WORK

Aneetha and Bose [3] proposed a modified Self Organizing Map algorithm which initially starts with null network and grows with the original data space as initial weight vector, updating neighbourhood rules and learning rate dynamically in order to overcome the fixed architecture and random weight vector assignment of simple SOM. New nodes are created using distance threshold parameter and their neighbourhood is identified using connection strength and its learning rule and the weight vector up-dation is carried out for neighbourhood nodes. The k-means

clustering algorithm is employed for grouping similar nodes of Modified SOM into k clusters using similar measures.
A new approach is proposed by Berglund and Sitte [8]. The parameter less self-organizing map (PLSOM) is a new neural network algorithm based on the self-organizing map (SOM). It eliminates the need for a learning rate and annealing schemes for learning rate and neighborhood size. We discuss the relative performance of the PLSOM and the SOM and demonstrate some tasks in which the SOM fails but the PLSOM performs satisfactory.

Different data clustering algorithms has been studied and compared by Abbas [2]. These are compared according to the factors like size of dataset, type of dataset, number of clusters and tool used. The algorithms considered for investigation are k-means algorithm, self organizing map algorithm, hierarchical clustering algorithm and expectation maximization algorithm. Conclusions extracted from comparative study of these algorithms belong to the performance, quality and accuracy of algorithms.

Hosseini[13] suggests the similarities between the mechanisms used in the TASOM (Time Adaptive Self-Organizing Map) neural network and AIS (Artificial Immune Systems) are analyzed. To demonstrate the similarities, AIS mechanisms are incorporated into the TASOM network such as the weight updating is replaced by a mutation mechanism. Learning rate and neighborhood sizes are also replaced by the clonal selection process used in AIS. This new network is called TAISOM. Experimental results with TAISOM are implemented for uniform and Gaussian distributions for one and two-dimensional lattices of neurons. These experiments show that TAISOM learns its environment as expected so that neurons fill the environments quite well and the neurons also preserve the topological ordering.

III. ALGORITHMS

A. Existing SOM Algorithm

In Kohonen SOM (Kohonen, 1982) discussed spatially continuous input space in which our input vectors live. The aim is to map from this to a low dimensional spatially discrete output space, the topology of which is formed by arranging a set of neurons in a grid. SOM provides such a nonlinear transformation called a feature map [11].
The learning algorithm of SOM is detailed below in the following steps:
Step 1 Initialize the map's node's weight vector.
Step 2 Grab an input vector.
Step 3 Traverse each node in the map.
Step 4 Use Euclidean distance formula to find similarity between the input vector and the map's node's weight vector.
Step 5 Track the node that produces the smallest distance (this node will be called the Best Matching Unit or BMU).
Step 6 Update the nodes in the neighborhood of BMU by pulling them closer to the input vector.

$$Wv(t+1) = Wv(t) + \theta(v,t)\,\alpha(t)\,(D(t)) - Wv(t)) \quad (1)$$

Where $\alpha(t)$ is a monotonically decreasing learning coefficient and D(t) is the input vector. The neighborhood function $\theta(v,t)$ depends on the lattice distance between the BMU and neuron v.

B. Modified SOM Algorithm

In existing Self-Organising Maps (SOM), there is a problem of dependence on the learning rate, the size of the neighbourhood function and the decrease of these parameters as training progresses. In improved approach, a simple modification is done in existing SOM that completely eliminates the learning rate, the decrease of the learning rate and the decrease of the neighbourhood size [36]. A new learning rule is introduced. This has been done by making the learning rate and neighbourhood size dependant ona variable calculated from the internal state of the SOM, rather than on externally applied variables

The learning algorithm of improved SOM is detailed below in the following steps:
Step 1 Initialize the map node's weight vectors.
Step 2 Traverse each input vector in the input data set
Step 3 Use the Euclidean distance formula to find the similarity between the input vector and the map's node's weight vector.
Step 4 Track the node that produces the smallest distance (this node is the best matching unit, BMU).
Step 5 Update the nodes in the neighborhood of the BMU.

The new learning rule is introduced in this algorithm is as follows:

$$Wv(t+1) = Wv(t) + \theta(v,t)\,\alpha(t)\,(D(t)) - Wv(t)) + \theta(v,t)\,(t/T) \quad (2)$$

Where $\alpha(t)$ is a monotonically decreasing learning coefficient and D(t) is the input vector. The neighborhood function $\theta(v,t)$ depends on the lattice distance between the BMU and neuron v. Current time is represented by t and T is the total time.

IV. COMPARISON

Table 1. Comparative results of existing and modified SOM

Parameters	Existing SOM	Modified SOM
Computational Time	The existing approach consumes much time	The improved approach consumes half the time taken by existing approach
Complexity	The existing algorithm is more complex as it depends on learning rate	The improved algorithm is simpler to implement as it does not depends on learning rate
Cost	Existing algorithm is more expensive	Improved SOM is less expensive than existing SOM
Error Rate	It has higher error rate as compared with improved approach	It has lower error rate as compared with existing approach
Efficiency	The existing algorithm is less efficient to find clusters	The improved algorithm is more efficient than the existing approach

- Computation time (in milliseconds) – It is the length of time required to perform a computational process. The time taken by the modified SOM is very less as compare to the time taken by existing on the same data set.
- Complexity can be measured by two factors:
  - Time complexity – Time complexity of an algorithm signifies the total time required by the program to run to completion. For modified SOM algorithm it is low as compared to existing, which is beneficial for better performance.
  - Space complexity – Space complexity of an algorithm is total space taken by algorithm with respect to input size. For modified algorithm it is high as compared to existing which is beneficial for better performance.
- Cost - The cost incurred for the two algorithms depends on the complexity factor and the no. of iterations in the dataset used. As the complexity of algorithm increase, the cost factor also increases. The improved SOM is observed to be less costly than existing SOM.
- Error rate (in %) - Modified SOM comparatively gives less error rate (55%) than existing(57%). Value of Error rate lies between 0 and 1.
- Efficiency (in %) - Results produced by modified SOM are more efficient than existing.

V. CONCLUSIONS

The results of the implementation of modified SOM algorithm has led to some important conclusions. This approach relies on the idea that the learning rate and neighbourhood size should not vary according to the iteration number, but rather vary according to how well the map represents the topology of the input space. It also markedly decreases the number of iterations required to get a stable and ordered map. Improved SOM completely eliminates the selection of the learning rate, the annealing rate and annealing scheme of the learning rate and the neighbourhood size, which have been an inconvenience in applying SOMs. It learns continuously from its environment, and only a one-time initialization is needed to work in its possibly changing environment. The improved SOM also reduces the training time and preserves generality. This is achieved without inducing a significant computation time or memory overhead. It also covers a greater area of the input space, leaving a smaller gap along the edges.

VI. FUTURE SCOPE

The improved SOM can be applied to many familiar problems. The future scope of this is listed below:

- In processing a stereo sound signal, improved SOM can be used to determine the direction of the sound source and orienting the microphones toward the sound source. Thus improved SOM can deal with cases where the number of input dimensions is far higher than the number of output dimensions.
- The SOM related methods are finding wide application in more and more fields, to make the methods more efficient, robust and consistent is a key challenge, especially for large-scale, real-world applications.
- Improved SOM can handle very high-dimensional and clustered data. Thus it can be used in image segmentation. For general pattern recognition, it may have more potential than implied by current practice, which often limits the SOM to a 2-D map and empirically chosen model parameters.
- Probabilistic extensions of the SOM may provide useful tools in deciphering and interpreting the information content and relationships conveyed among stimuli and responses.
- The algorithm may be further extended in order to deal with complex biological signals and networks, for example in handling spikes and more importantly multiple, perhaps in homogeneous and population spike trains.

VII. REFERENCES

[1] Agrawal, R., Mehta, M., Shafer, J., Srikant, R., Arning, A. and Bollinger, T, "The Quest Data Mining System", Proceedings of 1996 International

Conference on Data Mining and Knowledge Discovery (KDD'96), Portland, Oregon, **1996**, pp. **244-249**.

[2] Abbas, O.A., "Comparison between Data Clustering Algorithms", TheInternational Arab Journal of Information Technology, Vol.5, No.3, **2008**, pp. **320-325.**

[3] Aneetha, A.S. and Bose, S., "The combined approach for anomaly detection using neural networks and clustering techniques", Computer Science & Engineering: An International Journal, Vol.2, No.4, **2012**, pp. **37-46**.

[4] Bacao, F., Lobo, V. and Painho, M., "Self-organizing Maps as Substitutes for K-Means Clustering", International Conference on Computational Science, Springer-Verlag Berlin Heidelberg, Vol.3516, **2005**, pp. **476 – 483**.

[5] Bacao, F., Lobo, V. and Painho, M., "Clustering census data: comparing the performance of self-organizing maps and k-means algorithms", Proceedings of KDNet(European Knowledge Discovery Network of Excellence), Knowledge-Based Services for the Public Sector, Workshop 2: Mining Official Data, Petersberg Congress Hotel, Bonn, Germany, Vol.2, **2004.**

[6] Birdi, M., Gangwar, R.C. and Singh, G., "A Data Mining Clustering Approach for Traffic Accident Analysis of National Highway-1", International Journal of Advanced Research in Computer Science and Software Engineering, Vol.4, Issue-10, **2014**, pp. **44- 47.**

[7] Bhatia, S.K. and Dixit, V.S., "A Propound Method for the Improvement of Cluster Quality", International Journal of Computer Science Issues, Vol.9, Issue-4, No.2, **2012**, pp. **216-222**.

[8] Berglung, E. and Sitte, J., "The Parameter-less Self-Organizing Map Algorithm", IEEE Transactions on Neural Network, Vol.17, No.2, **2006**, pp. **305-316.**

[9] Chen, Y., Qin. B., Liu, T., Liu, Y. and Li, S., "The Comparison of SOM and K- means for Text Clustering", Computer and Information Science, Vol.3, No.2, **2010**, pp. **268-274**.

[10] Dhingra, S., Gilhotra, R. and Ravishanker, R., "Comparative Analysis of Kohonen-SOM and K-Means data mining algorithms based on Academic Activities", International Journal of Computers & Technology, Vol.6, No.1, **2013**, pp. **237-241.**

[11] Ganda, R. and Chahar, V., "A Comparative Study on Feature Selection Using Data Mining Tools",International Journal of Advanced Research in Computer Science and Software Engineering, Vol.3, Issue-9, **2013**, pp. **26-33**.

[12] Halkidi, M., Batistakis, Y. and Vazirgiannis, M., "On Clustering Validation Techniques", Journal of Intelligent Information Systems, Vol.17, No.2, **2001**, pp. **107–145.**

[13]Hosseini, H.S., "The Time Adaptive Self-Organizing Map is a Neural Network Based on Artificial Immune System", Proceedings of IEEE World Congress on Computational Intelligence, International Joint Conference on Neural Networks,Sheraton Vancouver Wall Centre Hotel, Vancouver, Canada, Vol.6, No.3, **2006**, pp.**1007-114**.

[14] Hemalatha, M. and Saranya, N.N., "A Recent Survey on Knowledge Discovery in Spatial Data Mining", International Journal of Computer Science Issues, Vol.8, Issue-3,**2011**, pp. **473-479**.

[15] Johal, H.S., Singh, B., Singh, H., Nagpal, A. and Virdi, H.S.,"UsingKohonen-SOM & K-Means Clustering Techniques to Analyze QoS Parameters of RSVP", Proceedings of the World Congress on Engineering and Computer Science, Vol.1, **2012**, pp. **431-436.**

[16] Kohonen, T., "The self-organizing maps", Proceedings of the IEEE, Vol.78, No.9,**1990**, pp. **1464-1480**.

[17]Kohonen, T., "Self-Organized Formation of Topologically Correct Feature Maps", Biological Cybernetics, Springer, Espoo, Finland, Vol.43, **1982**, pp. **59-69.**

[18] Kohonen, T., "Essentials of the self-organizing maps", International Conference on Neural Networks, Vol.37,**2013**, pp. **52-65.**

[19] Kaur, J. and Singh, G., "Review of Error Rate and Computation Time of Clustering Algorithms on Social Networking Sites", International Journal of Computer Application, Vol.113, No.8, **2015**, pp. **32-35.**

[20] Mingoti, S.A. and Lima, J.O., "Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms", European Journal of Operational Research, Vol.174, **2006**, pp. **1742–1759**.

[21] Murugavel, P. and Punithavalli, M., "Improved Hybrid Clustering and Distance-based Technique for Outlier Removal", International Journal on Computer Science and Engineering, Vol.3, No.1,**2011**, pp. **333-339.**

[22] Mishra, M. andBehera, H.S., "Kohonen Self Organizing Map with Modified K-means Clustering for High Dimensional Data Set", International Journal of Applied Information Systems, Vol.2, No.3, **2012**, pp. **34-39.**

[23]Ravikumar, S. andShanmugam, A., "Comparison of SOM Algorithm and K-Means Clustering Algorithm in Image Segmentation", International Journal of Computer Applications, Vol.46, No.22, **2012**, pp. **21-25.**

[24] Raghuwanshi, S.S. and Arya, P.N., "Comparison of K-means and Modified K-means for Large Data-set", International Journal of Computing, Communications and Networking, Vol.1, No.3, **2012**, pp. **106-110**.

[25] Sumathi, N., Geetha, R. and Bama, S.S., "Spatial data mining-techniques trends and its applications", Journal of Computer Applications, Vol.1, No.4,**2008**, pp. **28-30.**

[26] Sundararajan, S. and Karthikeyan, S., "A Study On Spatial Data Clustering Algorithms In Data Mining", International Journal Of Engineering And Computer Science, Vol.1, Issue-1, **2012**, pp. **37-41.**

[27] Subitha, N. and Padmapriya, A., "Clustering Algorithm for Spatial Data Mining:An Overview", International Journal of Computer Applications, Vol.68, No.10,**2013**, pp. **28-33.**

[28] Sharma, K. and Dhiman, R., "Implementation and Evaluation of K-Means, Kohonen-SOM, and HAC Data Mining Algorithms base on Clustering", International Journal of Computer Science Engineering& Information Technology Research, Vol.3, Issue-1, **2013**, pp. **165-174.**

[29] Sharma, H. and Kaler, N.K., "A Synthesized Approach for Comparison and Enhancement of Clustering Algorithms in Data Mining for Improving Feature Quality", International Journal of Soft Computing and Engineering, Vol.4, Issue-2, **2014**, pp. **114-117.**

[30] Toor, A.K. and Singh, A., "Analysis of Clustering Algorithms Based on Number of Clusters, Error Rate, Computation Time and Map Topology on Large Data Set", International Journal of Emerging Trends & Technology in Computer Science, Vol.2, Issue-6, **2013**, pp. **94- 98.**

[31] Torma, M., "Comparison between three different clustering algorithms", The Photogrammetric Journal of Finland, Vol.13, No.2, **1993**, pp. **85-95**.

[32] Uriarte, E. A. and Martin, F.D., "Topology Preservation in SOM", International Scholarly and Scientific Research & Innovation, Vol.2, No.9, **2008**, pp. **829-832.**

[33] Yin, H., "The Self-Organizing Maps: Background, Theories, Extensions and Applications", Computational Intelligence: A Compendium, Springer Berlin Heidelberg, Vol.115, **2008**, pp. **715-762.**

