International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:2, No:9, 2008

# Topology Preservation in SOM

E. Arsuaga Uriarte, and F. Díaz Martín

***Abstract*****—**The SOM has several beneficial features which make it a useful method for data mining. One of the most important features is the ability to preserve the topology in the projection. There are several measures that can be used to quantify the goodness of the map in order to obtain the optimal projection, including the average quantization error and many topological errors. Many researches have studied how the topology preservation should be measured. One option consists of using the topographic error which considers the ratio of data vectors for which the first and second best BMUs are not adjacent. In this work we present a study of the behaviour of the topographic error in different kinds of maps. We have found that this error devaluates the rectangular maps and we have studied the reasons why this happens. Finally, we suggest a new topological error to improve the deficiency of the topographic error.

***Keywords*****—**Map lattice, Self-Organizing Map, topographic error, topology preservation.

## I. Introduction

THE Self-Organizing Map (SOM) is a neural network algorithm that is based on unsupervised learning. It has properties of both vector quantization and vector projection algorithms. The SOM has proven to be a valuable tool in data mining with applications in full-text and financial data analysis [1]-[3]. It has also applied successfully in various engineering applications in pattern recognition, image analysis, process monitoring and fault diagnosis [4], [5].

The SOM provides a non-linear, ordered, smooth mapping of high-dimensional input data manifolds onto the elements of a regular, low-dimensional array. The main characteristic of the projection provided by the algorithm is the preservation of neighbourhood relations; as far as possible, nearby data vectors in the input space are mapped onto neighbouring locations in the output space [6], [7], [8]. This feature makes the Self-Organizing Map very useful in data analysis and data visualization where a common goal is to represent data from a high-dimensional space in a low-dimensional space so as to preserve the internal structure of the data in the input space [9], [10], [11].

Preserving neighbourhood's relations in the mapping makes possible to see more clearly in the output space the structure hidden in the high-dimensional data, such as clusters [12], [13].

In order to guarantee the correct analysis of the input data we have to be sure the mapping has been correctly chosen. For this purpose, there are different measures to quantify the goodness of a map. The accuracy of the maps in preserving the topology, or neighbourhood relations, of the input space has been also measured in various ways.

In this work, the behaviour of a widely used topological error called topographic error is studied. We examine its behaviour in detail in many maps and detect a tendency to undervalue the rectangular self-organizing maps. Furthermore, we suggest an improved version to quantify the neighbourhood preservation.

We start in Section 2 with a brief review of different measures to quantify a Self-Organizing Map. In section 3 we present a case studied and we analize the behaviour of the topographic error. We also suggest an improvement of this error. Finally, in Section 4 we present conclusions and further studies.

## II. Quantifying the Goodness of SOM

The Self-Organizing Map is defined in the training phase. During the training, we have to make assumptions about several parameters of the map, such as learning parameters, map topology and map size. These features influence in the final map, thus it is very important to choose these parameters carefully in order to reach the appropriate map [6], [14], [15]. Once we have tested different choices, we can use some measures to evaluate the quality of the map and select the optimal one to represent our data.

Several measures have been used to evaluate the quality of a Self-Organizing Map. A widely used measurement is the quantization error. This error measures the average distance between each data vector and its best matching unit (BMU). The quantization error is calculated as shown in formula (1), where N is the number of data-vectors and $m_{\vec{x}_i}$ is the best matching prototype of the corresponding $\vec{x}_i$ data-vector:

$$qe = \frac{1}{N}\sum \left\| \vec{x}_i - m_{\vec{x}_i} \right\| \quad (1)$$

This error evaluates the fitting of the neural map to the data. Thus, the optimal map is expected to yield the smallest average quantization error. The smaller the quantization error, the smaller the average of the distance from the vector data to the prototypes, and that means, that the data vectors are closer to its prototypes.

But what happens with the topological preservation? It is

Manuscript received August 31, 2006.
E. Arsuaga Uriarte is with the Faculty of Engineering, University of Deusto, Bilbao (e-mail: earsuaga@eside.deusto.es).
F. Díaz Martín is with the Faculty of Engineering, University of Deusto, Bilbao.

quite another question whether the quantization error on its own describes the topological order of the map: several indicators of topological errors have been used in addition to quantization error to control the conservation of topology [16]. Topology preservation has, however, turned out to be quite difficult to define sensibly for a discrete grid. There seem to exist two different approaches for measuring the degree of topology preservation [9].

In the first approach the relations between the reference vectors and the relations between the corresponding units on the map lattice are compared such as the topographic product does [17], [18].

An alternative approach for measuring topology preservation is to use input samples to determine how continuous the mapping from the input space to the map grid is. One of the most extended indices for this purpose is the topographic error [19]. It is also one of the errors proposed by Kohonen himself [6]. This error measures the proportion of all data vectors for which first and second best-matching units (BMU) are not adjacent vectors. So the lower the topographic error is, the better the Self-Organizing Map preserves the topology.

$$te = \frac{1}{N}\sum_{i=1}^{N} u(\vec{x}_i) \qquad (2)$$

The topographic error is calculated as shown above where the function $u(\vec{x}_i)$ is 1 if $\vec{x}_i$ data vector's first and second BMUs are adjacent and, 0 otherwise.

## III. Case Study

As we have said above, in the learning phase of a Self-Organizing Map we have to decide about learning parameters, map size and map topology. Then we can quantify the goodness of the maps, so that we can get the optimal one. The average quantization error and the topographic error are measures used for this purpose. In order to study the behaviour of the topographic error, we have chosen networks with different characteristics, according to their dimensionality and topology.

The lattice type of the output space can be defined to be rectangular or hexagonal. And the size has been varied getting 16 different neural networks. First of all, we have normalized each attribute scale such that its variance taken over all the items is unity. We have performed the learning process in the same way for every map. We have used the SomToolbox [20] package made available by the Helsinky University of Technology for all researchers in the net. All maps have been trained using the batch-training algorithm (with *som_make)*. They have been linearly initialized in the subspace spanned by the two eigenvectors with greatest eigenvalues computed from the training data. The maps were trained in two phases: a rough training with large initial neighbourhood width and a fine-tuning phase with small initial neighbourhood width. The neighbourhood function was Gaussian. The neural networks have been created after normalizing the variables to avoid any difference in the variables

To illustrate the performance of the proposed research, we have used a simple and synthetic data set. The variables values are between [0,10]. As we work with a simple two-dimensional-example, we can evaluate the quality of a map even without the help of any appropriateness measure. The topology preserving quality of the map can be assessed by visual inspection by performing a Sammon mapping onto a 2D space.

In the next table we show the different maps with its corresponding average quantization error and topographic error. The average quantization error decreases as the map's size increases. This is obvious because as the number of units increases there are more neurons to represent the data, therefore each data vector will be closer to its best matching unit. Besides this there is no any significative difference with respect the topology of the map; rectangular and hexagonal maps fit the input data in a similar way.

TABLE I
Average Quantization Error and Topographic Error of the SOMs

| Map Size | Quantization Error | | Topographic Error | |
|---|---|---|---|---|
| | Rectangular | Hexagonal | Rectangular | Hexagonal |
| 3x4 | 1.2016 | 1.2761 | 0 | 0.0169 |
| 4x3 | 0.8805 | 1.2521 | 0.0169 | 0.0169 |
| 4x4 | 0.8852 | 0.9433 | 0 | 0 |
| 5x5 | 0.5999 | 0.6897 | 0.0678 | 0.0169 |
| 7x5 | 0.4848 | 0.5296 | 0.1356 | 0.0678 |
| 7x6 | 0.4362 | 0.4686 | 0.0847 | 0.0339 |
| 8x5 | 0.4596 | 0.4937 | 0.1525 | 0.0508 |
| 8x7 | 0.3731 | 0.3939 | 0.1017 | 0.0339 |

On the other hand we have the topographic error. In general, as the dimension increases the topographic error also increases. This is due to the growing complication to obtain the units in order while the number of prototypes increases. In addition, we have to remark that there is some difference in the topographic error of rectangular and hexagonal maps. As we show above, the topographic error considers, for each input vector, the distance of the best matching unit and second best matching unit on the map: If the units are not neighbours, then the topology is not preserved. However it seems that the hexagonal maps preserve the topology respect to this error better than the rectangular ones. But, is this true? Do the rectangular maps preserve the neighborhood relations worse that the hexagonal ones?

As the topographic error gives the proportion of all data vectors for which first and second best BMUs are not adjacent units, we can calculate precisely the number of data vectors that doesn't meet condition. Furthermore, we can even know exactly which are the data vectors for which the two first BMUs are not adjacent. In order to study this case and to make the task easier, we have implemented a function (called *som_disordered*) in matlab that returns the pair of neurons that are not adjacent but, happen to be the first and second BMUs

International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:2, No:9, 2008

TABLE II
TOPOGRAPHIC ERROR OF THE SOMS

| Map Size | Rectangular Maps | | |
|---|---|---|---|
| | Number of disordered data-vectors | Disordered pair of neurons | Relation |
| 3x4 | 0 | -- | -- |
| 4x3 | 1 | 12 - 7 | Diagonal |
| 4x4 | 0 | -- | -- |
| 5x5 | 4 | 2 - 8 | Diagonal |
| | | 19 - 25 | Diagonal |
| | | 3 - 9 | Diagonal |
| | | 3 -9 | Diagonal |
| 7x5 | 8 | 2 - 10 | Diagonal |
| | | 29 - 23 | Diagonal |
| | | 21 - 27 | Diagonal |
| | | 16 - 22 | Diagonal |
| | | 16 - 22 | Diagonal |
| | | 16 - 22 | Diagonal |
| | | 6 - 12 | Diagonal |
| | | 13 - 7 | Diagonal |
| 7x6 | 5 | 2 - 10 | Diagonal |
| | | 36 - 30 | Diagonal |
| | | 23 - 29 | Diagonal |
| | | 23 - 29 | Diagonal |
| | | 23 - 29 | Diagonal |
| 8x5 | 9 | 2 - 11 | Diagonal |
| | | 34 - 27 | Diagonal |
| | | 26 - 13 | Diagonal |
| | | 26 - 13 | Diagonal |
| | | 26 - 13 | Diagonal |
| | | 26 - 13 | Diagonal |
| | | 7 - 13 | No diagonal |
| | | 6 - 13 | Diagonal |
| | | 7 - 13 | No diagonal |
| 8x7 | 6 | 10 - 11 | Diagonal |
| | | 49 - 42 | Diagonal |
| | | 34 - 41 | Diagonal |
| | | 34 - 41 | Diagonal |
| | | 34 - 41 | Diagonal |
| | | 14 - 17 | Diagonal |

Pair of neurons that make increase the topographic error. The last column contains the relation between pairs of neurons.

of some data vector.

In Table II, we present the pair of neurons which make increase the topographic error in our example. We have studied which are exactly these pair of neurons and also the relations between the pair of units in the rectangular maps.

As we can see in Table II, the majority of pairs that make increase the topographic error are diagonal. The topographic error in rectangular maps increases due to nearby diagonal units. But those maps actually are not disordered.

Although the topographic error does not consider the diagonal neuron neighbours, it does not really mean that the map does not conserve the local relations. It could be said that even if diagonal units are not neighbours, they are "special no neighbours". It is quite different to be diagonal units or units which are far away from each other. But the topographic error does not make any difference between different kinds of adjacent units. This could be one of the reasons why this error devaluates the rectangular maps. It seems that the error does not behave in the same way for rectangular and hexagonal maps.

Furthermore we would like to remark the number of adjacent units a neuron has in each lattice. A "central unit" in a hexagonal map is next to other six units, whereas in a rectangular lattice a "central unit" has only four neighbours. This means it is much easier, at least with respect to the error we are working with, for a hexagonal map to be organized rather than for a rectangular one.

Consequently, we would like to suggest a new measure for rectangular maps to improve the deficiency of the topographic error. The new error, called Alfa Error, takes into account different kind of no neighbours, the diagonal relations to be precise.

The Alfa Error is based on assigning weights to different kind of no neighbours. For instance, in each case we will have to decide the weight the diagonal neighbours should take. Let's call K the weight we want to give to the diagonal relation. This way, when we want the diagonal neighbours to be considered as "half-neighbours" K would be equal to two; if we want "third-neighbours" K should be equal to three and so on. Besides this, if we want the diagonal neighbour to be considered as a neighbour then K should be equal to 0. To summarize, the Alfa Error gives the opportunity to decide the importance the diagonal neighbours have in a rectangular map.

The Alfa Error we propose to quantify the conservation of topology is calculated as follows:

$$\alpha Error = \frac{1}{N}\sum_{i=1}^{N}\alpha(\vec{x}_i) \qquad (3)$$

Depending the relation of diagonal neurons, the function $\alpha(\vec{x}_i)$ is defined in a different way. If we want K to be equal to zero, $\alpha(\vec{x}_i)$ is going to be 1, if $\vec{x}_i$ data-vector's first and second best matching unit are nor adjacent neither diagonal, and 0 otherwise.

$$\alpha(\vec{x}_i) = \begin{cases} 1 & \text{, if first and second BMUs are not adjacent neither diagonals} \\ 0 & \text{, otherwise} \end{cases}$$

If K is equal to any other value ( $K \neq 0$ ) the function is going to be defined as follows:

TABLE III
THE ALFA ERROR FOR RECTANGULAR MAPS.

| Map Size | Rectangular Maps | | | Hexagonal Maps |
|---|---|---|---|---|
| | Topographic Errorr | Alfa Error | | |
| | | K=0 | K=2 | |
| 3x4 | 0 | 0 | 0 | 0.0169 |
| 4x3 | 0.0169 | 0 | 0.0085 | 0.0169 |
| 4x4 | 0 | 0 | 0 | 0 |
| 5x5 | 0.0678 | 0 | 0.0339 | 0.0169 |
| 7x5 | 0.1356 | 0 | 0.0678 | 0.0678 |
| 7x6 | 0.0847 | 0 | 0.0423 | 0.0339 |
| 8x5 | 0.1525 | 0.339 | 0.0932 | 0.0508 |
| 8x7 | 0.1017 | 0 | 0.0508 | 0.0339 |

$$\alpha(\vec{x}_i) = \begin{cases} 1 & \text{, if first and second BMUs are not adjacent neither diagonals} \\ \frac{1}{K} & \text{, if first and second BMUs are diagonals} \\ 0 & \text{, otherwise} \end{cases} \quad (4)$$

In Table III, we show the results obtained using the new topological error for rectangular maps.

We have calculated the error considering K=0, which means that diagonal neighbours are treated as neighbours. As most of the pairs that made increase the topographic error were diagonal units, it was obvious that the new error was going to decrease.

When comparing the different lattice maps, we observed that according to the new measurement with the rectangular networks, better results are obtained. Remember that when considering the diagonal neighbour units, we make a "centric neuron" to have 8 neighbours. So now, the rectangular maps have more adjacents than the hexagonal ones.

Furthermore, a way between the two measures would be considering K equal to 2. This means the diagonals would be counted as half of one no-vicinity. We can observe that in these cases the error is more neutral, because we obtain similar values for hexagonal and rectangular maps. This seems to have more sense because in general, it doesn't have to exist any significative difference between different lattice maps.

## IV. Conclusion

The topographic error seems to have a tendency to depreciate rectangular maps. This can be due to the smaller number of neighbours a "centric neuron" has in a rectangular map; to be more precise, a "centric neuron" has 6 neighbours in a hexagonal map whereas they only have 4 in rectangular ones.

In addition, in this work we have seen that in many cases nearby diagonal neurons are the reason why the topographic error increases in rectangular maps. This happens because diagonal units represent nearby data although they are not neighbours. But this doesn't mean the map is disordered.

Consequently we suggest a new error called Alfa Error, where a special care is given to the diagonal neurons. We suggest considering the diagonal units as neighbours as the researcher decides. A reasonable value could be K=2 where the diaonal units are considered half-neighbours.

In further studies, we will concentrate to extend the Alfa Error's idea considering more kind of different neighbours. We also are going to test this errors in a bigger database examples.

## References

[1] J. Vesanto. "Using SOM in Data Mining," Thesis for the degree of Licentiate of Science in Technology. Espoo, Finland, 2000.
[2] G. Deboeck, T. Kohonen. "Visual Explorations in Finance with Self-Organizing Maps," Springer Finance. Springer-Verlag, London, 1998.
[3] K. Lagus. "Text Mining with the WEBSOM," Acta Polytechnica Scandinaviva, Mathematics and Computing Series No.110, Espoo. Finnish Academies of Technology, 2000.
[4] Dolnicar, S. The Use of Neural Networks in Marketing: Market Segmentation with Self-Organizing Feature Maps. Proceeding of the Workshop on Self-Organizing Maps (WSOM'97), 38-43 (1997).
[5] Kohonen, T., Oja, E., Simula, O., Visa, A., Kangas, J. Engineering application of the self-organizing map. Proceedings of the IEEE, Vol.84, No.10, 1358-1384 (1996)
[6] J Kohonen, T.: Self-Organizing Maps. Vol. 30. of Springer Series in Information Sciences, 3rd ed., Springer-Verlag, Berlin Heidelberg (2001)
[7] T. Kohonen, "Self-organized formation of topologically correct feature maps," Biological Cybernetics Vol. 43, 59-69 (1982).
[8] H. Ritter, T. Martinetz, K. Schultzen. "Neural Computation and Self-Organizing Maps: An Introduction," Reading, MA: Addison-Weley (1992)
[9] S. Kaski, "Data Exploration Using Self-Organizing Maps." PhD thesis, Helsinky University of Technology, Acta Polytechnica Scandinavica: Mathematics, Computing and Management in Engineering, 82. (1997).
[10] B. D. Ripley, "Pattern Recognition and Neural Networks." Cambridge University Press, Cambridge, Great Britain (1996).
[11] .J. Mao, A. K. Jain, "Artificial neural networks for feature extraction and multivariate data projection." IEEE Transaction on Neural Networks, Vol.6, 296-317 (1995)
[12] J. Vesanto, E. Alhoniemi, "Clustering of the Self-Organizing Map." IEEE Transactions on Neural Networks (2000).
[13] K. Obermayer, H. Ritter, K. Schultzen, "Development and spatial structure of cortical feature maps: A model study". Advances in Neural Information Proceedings Systems, Vol. 3, 11-17, San Mateo, CA (1991).
[14] M. Herrmann, "Self-Organizing Feature Map with Self-Organizing Neighborhood Widths." IEEE International Conference on Neural Networks (1995).
[15] A. Flexer, "Limitations of self-organizing maps for vector quantization and multidimensional scaling." Technical Report oefai-tr-96-23. The Australian Research Institute for Artificial Intelligence. (1997).
[16] G. J. Goodhill, T. J. Sejnowski. "Quantifying neighbourhood preservation in topographic mappings." Proceedings of the 3rd Joint Sympostium on Neural Computation. La Jolla. CA Vol. 6, 61-82 (1996).
[17] H.U. Bauer, K. R. Pawelzik, "Quantifying the Neighborhood Preservation of Self-Organizing Feature Maps." IEEE Transactions on Neural Networks, Vol. 3, No. 4, 570-579 (1992).
[18] J. C. Bezdek, N.R. Pal, "An index of topological preservation for feature extraction." Pattern Recognition, Vol. 28, No. 3, 381-391 (1995).
[19] K. Kiviluoto, "Topology Preservation in Self-Organizing Maps." Proceedings of International Conference on Neural Networks (ICNN), 294-299 (1996).
[20] SomToolbox, Helsinky University of Technology Available: www.cis.hut.fi/projects/somtoolbox