World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:2, No:6, 2008

Estimating an Optimal Neighborhood Size in the Spherical Self-Organizing Feature Map

Alexandros Leontitsis, Archana P. Sangole

***Abstract*—**This article presents a short discussion on optimum neighborhood size selection in a spherical self-organizing feature map (SOFM). A majority of the literature on the SOFMs have addressed the issue of selecting optimal learning parameters in the case of Cartesian topology SOFMs. However, the use of a Spherical SOFM suggested that the learning aspects of Cartesian topology SOFM are not directly translated. This article presents an approach on how to estimate the neighborhood size of a spherical SOFM based on the data. It adopts the L-curve criterion, previously suggested for choosing the regularization parameter on problems of linear equations where their right-hand-side is contaminated with noise. Simulation results are presented on two artificial 4D data sets of the coupled Hénon-Ikeda map.

***Keywords*—**Parameter estimation, self-organizing feature maps, spherical topology.

I. Introduction

THE Self-Organizing Feature Maps (SOFMs) [1] is a class of neural networks capable of recognizing the main features of the data they are trained on. There is extensive literature on its biological and mathematical concepts and even more on its implementation in a variety of areas including medicine, finance, chaos and data mining in general ([2]-[6]). It is an unsupervised learning algorithm that clusters data while it is simultaneously attempting to preserve the topology of the input space within the structure of its predefined lattice. Selected examples on the various implementations of the SOFM and the different ways of visualizing the data using SOFMs can be found in [2], [3], and [5].

Adaptive learning is realized by exploiting redundant and complementary information embedded in the data. Organization of the input is brought about by lateral interaction among the nodes (cluster units) in the SOFM lattice. Higher the degree of lateral interaction, greater is the degree of organized learning. Lateral interaction among the cluster units in the SOFM lattice is primarily determined by the topology of the lattice. Topology, in this context, is defined as the geometric connectivity between the cluster units or nodes in the grid that represents the configuration of the SOFM space. Therefore, a topological discontinuity in the lattice will restrict the lateral interaction between the cluster units and therefore restrict learning.

The Spherical SOFMs (S-SOFMs), introduced [7] as a natural extension of the SOFMs, have an inherit capability of visualizing high dimensional data ([8]-[11]). This is attributed to the spherical topology of the SOFM lattice which not only has an overall symmetry and continuity in its structure, but also provides a 3D framework for visualizing the data.

Key parameters of learning in the SOFM include the neighborhood size, the learning rate, and convergence of the algorithm. While much has been published on estimating the neighborhood size in Cartesian topology SOFMs, there is no literature on the S-SOFMs regarding this parameter. Extensive implementation of the spherical SOFM suggested that the lessons learned in the case of Cartesian topology SOFMs did not directly translate to S-SOFMs. This article discusses how to estimate an optimal neighborhood size in an S-SOFM. It is inspired by the findings from fundamental research literature on the regularization parameter on system of linear equations. Its implementation is illustrated using examples of simulated chaotic data.

II. The Spherical Self Organized Feature Map

An S-SOFM (also known as "glyph") is essentially a non-linear mapping from the data space to the surface of a sphere. The data space is usually of Cartesian form, but can also be of any other form. If $\boldsymbol{P}$ represents the m-dimensional spatially continuous input space that comprises of a set of activation patterns, defined by metric relationship $\mathbf{p}_i \in \mathbf{P}$, and $\mathbf{w}$ denotes the spatially discrete S-SOFM space (Fig. 1), then, in the mathematical sense the S-SOFM non linear transformation may be expressed as

$$\mathbf{\Phi} : \mathbf{P} \rightarrow \mathbf{w} \tag{1}$$

where $\mathbf{\Phi}$ is the S-SOFM non-linear mapping between the m-

Manuscript received April 20, 2005. Sincere thanks are due to Dr. Kenneth J. Ottenbacher, Director of Rehabilitation Sciences, at the University of Texas Medical Branch, Galveston-Texas, for his support through the National Institute of Disability and Rehabilitation Research (NIDRR) grant no. H133P990001.

A. Leontitsis is with the Department of Education, University of Ioannina, 45110 – Dourouti, Ioannina Greece, (phone: +30-26510-95722; fax: +30-26510-95816; e-mail: me00743@ cc.uoi.gr), and with the Center for Research and Applications of Non Linear Systems, University of Patras, 26500 – Rio, Patras, Greece.

A. P. Sangole, is with the Division of Rehabilitation Sciences, The University of Texas Medical Branch, Galveston, TX 77555-1137, USA (e-mail: apsangol@utmb.edu).

dimensional input data space and the m-dimensional weight vectors of the SOFM space. The SOFM space is represented by a network of inter-connected nodes, where each node has a weight vector associated with it which is of the same dimension as the input space. During learning, every input vector gets assigned to a cluster unit whose weight vector is closest to it in the Euclidean sense.

$$\mathbf{w}_j(\text{new}) = \mathbf{w}_j(\text{old}) + a \cdot h(r,s) \cdot [\mathbf{p}_i - \mathbf{w}_j(\text{old})] \quad (2)$$

where 'a' is the learning parameter and h(r,s) is the neighborhood function. This function is of the form $h(r,s) = \exp\left(-r^2/(sR)\right)$, where 's' is the neighborhood size parameter, and 'R' is the size of the neighborhood that is able to cover a hemisphere and it is considered constant. The hemisphere is the largest possible neighborhood that can be considered.

The cluster units in the S-SOFM space therefore represent a reduced set of "internally ordered" prototypical representation vectors of the input space. The fundamental assumption of the S-SOFM algorithm is that clusters existing in the input space will also exist in the low-dimensional mapping space. Therefore the distribution of the input vectors in the S-SOFM space will reflect some physical aspects of the data. Associations between the numeric vectors are reflected in the relative topological positions of the cluster units on the grid, to which they have been assigned.

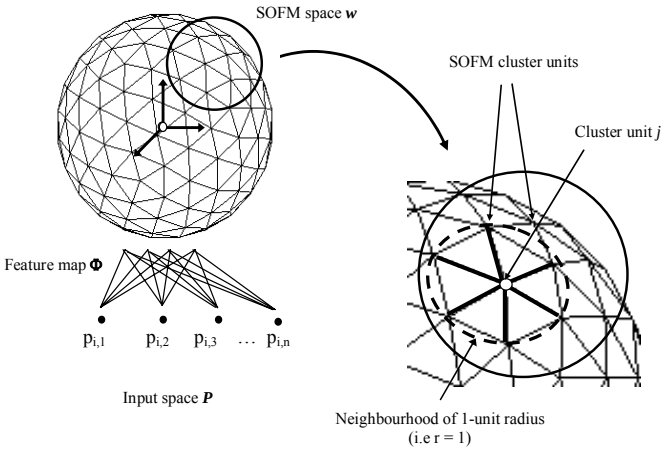

Fig. 1 The spherical self-organizing feature map [8]

Learning in the S-SOFM takes place due to lateral interaction between neighboring nodes. Therefore the lattice of the S-SOFM plays a significant role in the learning process. A discontinuity in the lattice will result in a restricted neighborhood and consequently restrict learning. In the more common 1- and 2- dimensional SOFM some researchers have investigated different ways to find the optimal learning parameters [12].

III. The Neighborhood Parameter by the L-Curve Criterion

The neighborhood size is essentially a regularization parameter. One of the most promising methods for estimating the regularization parameter is the L-curve criterion [13]. This criterion is widely used in linear algebra problems for handling solutions of linear systems with more equations than unknowns, and has found applications in image de-blurring. The algorithm of this criterion can be summarized in the following steps:

1) Estimate the smoothness and the accuracy for a wide range of the regularization parameter.
2) Plot smoothness (horizontal axis) against accuracy (vertical axis). The resulting plot should be L-shaped (Fig. 2).
3) The corner of the L-shape corresponds to the optimum regularization parameter. It is a value that results both high smoothness and high accuracy.

The original L-curve plot [13], requires accuracy and smoothness to be plotted on the log-log scale. If this rule is relaxed using a linear scale on both axes, as is the case in the example data set presented in this article, the result is a sharp curve.

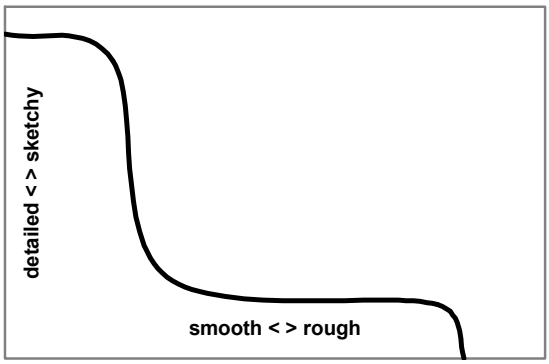

Fig. 2 The ideal L-curve shape
Note that in practice the curve may not be so sharp

In the L-curve criterion, smoothness is defined as the sum of distances between each vertex of the glyph ($\mathbf{w}_j$) and the mean of its nearest neighbors ($\mathbf{W}_j$). The lesser this sum, smoother is the glyph. Likewise, accuracy is defined as the sum of the distances between each pattern ($\mathbf{p}_i$) and its closest feature ($\mathbf{w}_j$).

IV. Simulation Results

Consider two data sets formed by the time series of the Hénon map [14] (**x** and **y**) and the Ikeda map [15] (**u** and **v**) of lengths N_1=1024 and N_2=4096

$$
\begin{aligned}
x_{i+1} &= 1 - 1.4x_i^2 + y_i \\
y_{i+1} &= 0.3x_i \\
u_{i+1} &= 1 + 0.9\left(u_i \cos(t_i) - v_i \sin(t_i)\right) \\
v_{i+1} &= 0.9\left(u_i \sin(t_i) + v_i \cos(t_i)\right) \\
t_i &= 0.4 - \frac{0.6}{1 + u_i^2 + v_i^2}
\end{aligned}
\qquad (3)
$$

All the time series are standardized in order to have mean 0 and standard deviation 1. The first data set forms a 1024-by-4 pattern space $\mathbf{P_1}=[\mathbf{x_1}\ \mathbf{y_1}\ \mathbf{u_1}\ \mathbf{v_1}]$ (where $\mathbf{x_1}$, $\mathbf{y_1}$, $\mathbf{u_1}$, and $\mathbf{v_1}$ are column vectors of length 1024) and the second data set forms a 4096-by-4 pattern space $\mathbf{P_2}=[\mathbf{x_2}\ \mathbf{y_2}\ \mathbf{u_2}\ \mathbf{v_2}]$ (where $\mathbf{x_2}$, $\mathbf{y_2}$, $\mathbf{u_2}$, and $\mathbf{v_2}$ are column vectors of length 4096).

These 4 dimensional objects are very complex and cannot be visually perceived. They cannot be projected in a lower dimensional space because they have non-linear dependencies within them, which will be lost due to linear projection. Therefore non-linear data projection techniques to create meaningful low-dimensional representations of the data are much needed.

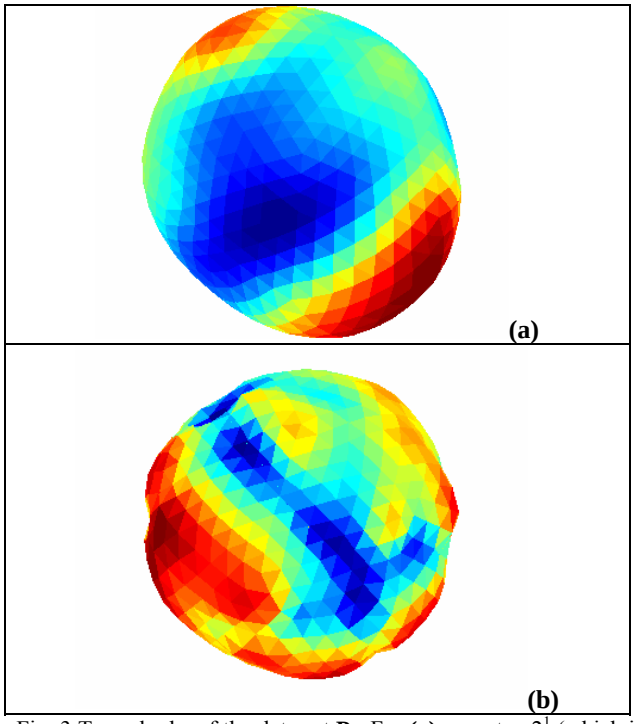

Fig. 3 Two glyphs of the data set $\mathbf{P_1}$. For **(a)** we set $s=2^1$ (which is very smooth but inaccurate) and for **(b)** we set $s=2^{-3}$ (which is very rough, and at the same time it is indicative of being over-fitted on the data)

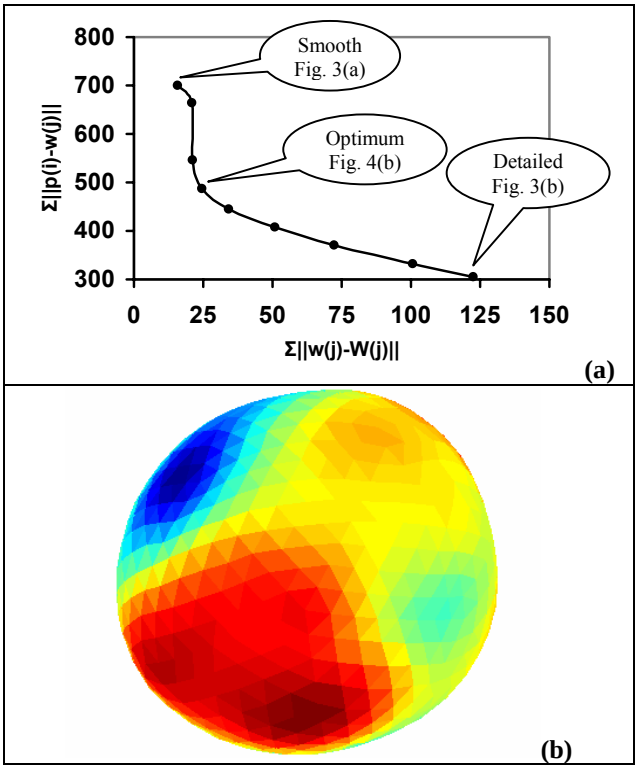

Fig. 4 The L-curve for the data set $\mathbf{P_1}$ and the corresponding glyph for the optimal value of s, which is $2^{-0.5}$.

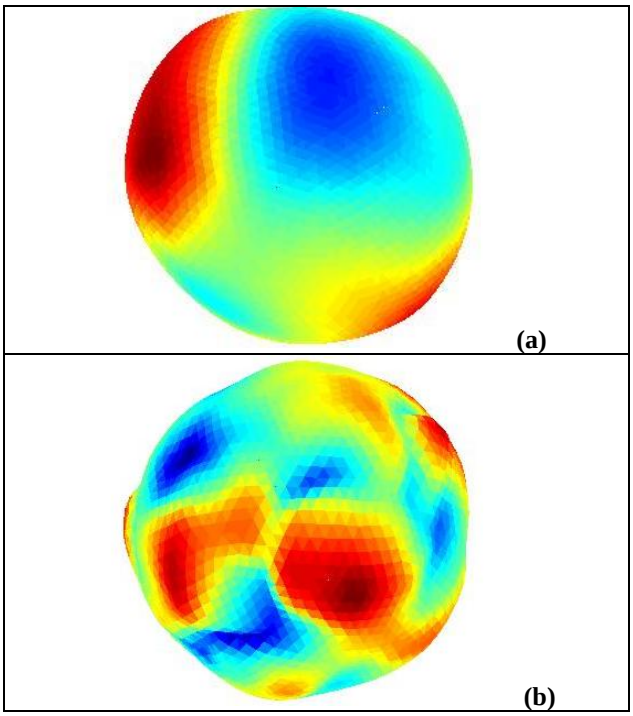

Fig. 5 Two glyphs made by the data set $\mathbf{P_2}$. For **(a)** we set $s=2^1$ (which is very smooth but inaccurate) and for **(b)** we set $s=2^{-3}$. (which is very rough, and at the same time gives the impression of being overfitted on the data

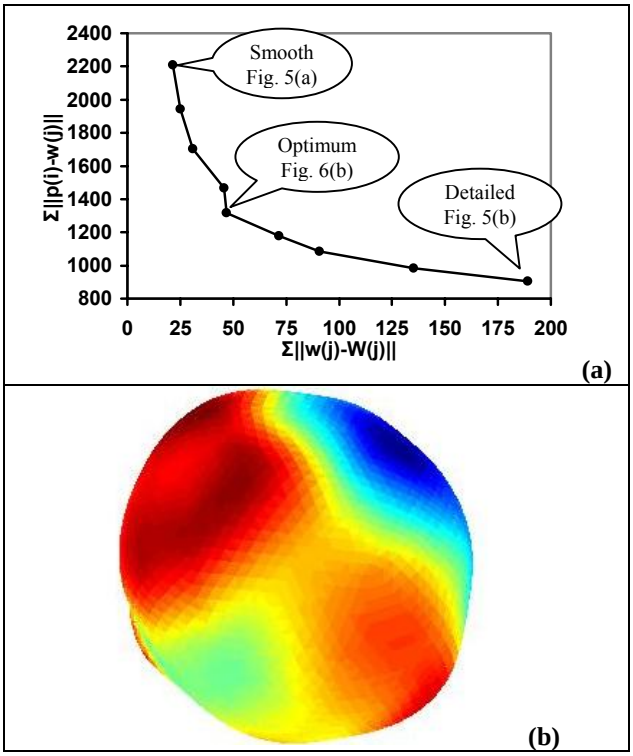

Fig. 6 The L-curve for the data set P_2 and the corresponding glyph for the optimal value of s, which is $2^{-0.5}$

The main disadvantage in the implementation of the S-SOFM so far is the absence of a criterion to choose an optimal neighborhood size parameter 's'. A large neighborhood results in smoother glyphs (which is desirable), but ignores the small-scale features. On the other hand, a small neighborhood takes into account all the features, no matter how large they are, but results in rough glyphs (which is undesirable). Therefore a cut-off point should be considered as the best tradeoff between smoothness and accuracy. Figs. 3 and 5 show the glyphs of the pattern space for neighborhoods that belong to the 2 extremes (i.e. $s=2^1$, and $s=2^{-3}$), while Figs. 4 and 6 shows the L-curves and the glyphs obtained by the optimum neighborhood parameter choice.

As observed in Figs. 4 and 6, the neighborhood size parameter 's' is neither dependent on the length of the data set, nor on the resolution of the glyph. The optimum parameter size ıs the same on the short and the long data sets ($\mathbf{P_1}$ and $\mathbf{P_2}$ respectively). Moreover, the glyphs of Figs. 3 and 4 have 642 nodes, while the glyphs of Figs. 5 and 6 have 2562 nodes. We observe that in both cases of glyph resolution the optimum parameter size is the same. This result is very desirable since it indicates that the proposed method defines a neighborhood selection parameter which is dependent on the nature of the data features and not on other parameters irrelevant to them (like the length of the data set or the resolution of the glyph).

V. Conclusions

A method for choosing the neighborhood size parameter regarding the Spherical Self Organizing Feature Maps was presented. This method provides a means to visualize high dimensional data as both smooth and detailed objects. It is based on the L-curve criterion which has been successfully applied on regularization problems in the past.

Future research in enhancing the potential of the Spherical SOFM as a data visualization tool requires an automatic estimation of the neighborhood parameter. This will save computational time in the proposed trial-and-error approach that examines different values of 's' prior to estimating an optimal value.

Another aspect of interest is the estimation of a reasonable number of training cycles (epochs). In the examples that were discussed, the data were trained for at most 60 cycles. Intuitively, this may be over-spending for the computational time. Unfortunately, inferences in this regard cannot be made based on existing methods for estimating the SOFM parameters. Although the performance of these methods have been successfully demonstrated in 1- and 2- dimensional SOFMs, additional research and testing is required to relatively comment in the context of the spherical SOFM.

Acknowledgment

The authors thank the reviewers for their helpful comments and suggestions.

References

[1] T. Kohonen, "Self-organized formation of topologically correct feature maps", *Biological Cybernetics*, vol. 43, pp. 59–69, 1982
[2] G. Deboeck, T. Kohonen, *Visual Explorations in Finance with Self-organizing Maps*. London: Springer-Verlag, 1998
[3] M. Gross, F. Seibert, "Visualization of multi-dimensional image data sets using a neural network", *The Visual Computer*, *International Journal of Computer Graphics*, vol. 10, pp. 145–159, 1993
[4] M. Gross, *Visual Computing: The Integration of Computer Graphics, Visual Perception and Imaging*. Berlin: Springer-Verlag, 1994
[5] J. Vesanto, "SOM-based data visualization methods", *Journal of Intelligent Data Analysis*, vol. 3, pp. 111–126, 1999
[6] J. Vesanto, E. Alhoniemi, "Clustering of the self-organizing map", *IEEE Transactions on Neural Networks*, vol. 11, pp. 586–600, 2000
[7] H. Ritter, "Self-organizing maps on non-Euclidean spaces", in *Kohonen Maps*, E. Oja and S. Kaski, Amsterdam: Elsevier Science B. V., 1999, pp. 97–109
[8] A. Sangole, G. K. Knopf, "Representing high-dimensional data sets as close surfaces", *Journal of Information Visualization*, vol. 1, pp. 111–119, 2002
[9] A. P. Sangole, "Data-driven Modeling using Spherical Self-organizing Feature Maps", Ph.D. dissertation, Dept. of Mech. and Mat. Eng., The University of Western Ontario, London, ON, Canada, 2003
[10] A. Sangole, G. K. Knopf, "Geometric representations for high-dimensional data using a spherical SOFM", *International Journal of Smart Engineering System Design*, vol. 5, pp. 11–20, 2003
[11] A. Sangole, G. K. Knopf, "Visualization of random ordered numeric data sets using self-organized feature maps", *Computers and Graphics*, vol. 27, pp. 963–976, 2003
[12] K. Haese, "Self-organizing feature maps with self-adjusting learning parameters", *IEEE Transactions on Neural Networks*, vol. 9, pp. 1270–1278, 1998
[13] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*. SIAM, 1997, pp. 186–193

[14] M. Hénon, "A two-dimensional map with strange attractor", *Communications in Mathematical Physics*, vol. 50, pp. 69–77, 1976
[15] K. Ikeda, "Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system", *Optics Communications*, vol. 30, p. 257, 1979

Alexandros Leontitsis was born in Ioannina, Greece, 1977. His first degree is in Business Administration, University of Macedonia, Thessaloniki, Greece, 2001, MSc. in Statistics, University of Kent at Canterbury, Canterbury, U.K., 2002, and Ph.D. in Chaos Theory and Computational Intelligence, Department of Education, University of Ioannina, Greece, 2004. He is specialized in the analysis of noisy chaotic time series, with most of his applications on financial data.

He is currently working on the Department of Education, University of Ioannina, 45110 – Dourouti, Ioannina, Greece, as a Post-Doc researcher. He is also a research fellow at the Center for Research and Applications of Nonlinear Systems, University of Patras, 26500 – Rio, Patras, Greece. He has published a textbook, co-authored by Costas Siriopoulos, entitled *Chaos – Time series analysis and prediction*, Thessaloniki: Anikoula, 2000, (in Greek) and papers to Journals related to Nonlinear Science, Applied Mathematics and Finance. He has served as a referee to papers regarding the analysis of chaotic time series. His research interests include computational aspects of noisy chaotic time series, scientific data exploration and visualization, improvement of computational intelligence techniques (especially the particle swarm optimization), numerical analysis, and simulation.

Dr. Leontitsis was a member of the Multinational Finance Society and was awarder by an EPSRC scholarship.

Archana P. Sangole is currently a NIDRR (National Institute of Disability and Rehabilitation Research) post-doctoral fellow in the Division of Rehabilitation Sciences at University of Texas Medical Branch, Galveston-Texas, USA. She is also a non-tenured Assistant Professor in the division.

She completed her PhD in mechanical engineering at the University of Western Ontario, London-Ontario, Canada in 2003. Her research interests in the past have included scientific data visualization, 3D freeform surface reconstruction, interactive design systems, and data exchange issues in computer-aided-design (CAD). The work was presented at various international conferences and has resulted in several refereed publications.

Driven by a personal desire, Dr. Sangole moved to Galveston-Texas in 2003 to pursue a research career in rehabilitation. Now, her research activities mainly concentrate on motor control issues in neuromuscular rehabilitation. A few specific interests include non-linearity of postural control in brain injury, kinematics of the neurological hand, upper extremity kinematics after brain injury and, visualization of clinical data.