ESANN'2006 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2006, d-side publi., ISBN 2-930307-06-4.

Weighted Differential Topographic Function: A Refinement of Topographic Function

Lili Zhang[1] and Erzsébet Merényi [2] *

1- Rice University, Rice Quantum Institute and Electrical & Computer Engineering MS-366
2- Rice University, Electrical & Computer Engineering MS-380
both at 6100 Main Street, Houston, TX 77005 - USA

Abstract. Topology preservation of Self-Organizing Maps (SOMs) is an advantageous property for correct clustering. Among several existing measures of topology violation, this paper studies the Topographic Function (TF) [1]. We find that this measuring method, demonstrated for low-dimensional data in [1], has a reliable foundation in its distance metric for the interpretation of the neighborhood relationship in the input space, for high-dimensional data. Based on the TF, we present a Differential Topographic Function (DTF) to reveal the topology violation more clearly and informatively. In addition, a Weighted Differential Topographic Function (WDTF) has been developed. For real world data, the DTF and WDTF unravel more details than the original TF, and help us estimate the topology preservation quality more accurately.

1 Introduction

Kohonen's Self-Organizing Map (SOM) algorithm has successfully been applied to many areas of science. One important property of the map $M_A = (\psi_{A \to M}, \psi_{M \to A})$ formed by an SOM is topology preservation, i.e., both the mapping $\psi_{M \to A}$ from M to A as well as the inverse mapping $\psi_{A \to M}$ from A to M are neighborhood preserving. M represents the input data manifold, and A represents the output SOM lattice. Topology preservation is needed for correct detection of the input data structure. However, this property may be lost if a dimensional mismatch occurs or the learning process is not properly parameterized. Therefore, we need a diagnostic tool to help monitor the topology violation during SOM learning, and remedy it in a timely manner.

The differences between various topology violation measures consist of two factors. One is the distance metric adopted to describe the neighborhood relationship; the other is the form of the cost function, i.e., the penalty for violation. The former is an embedded fundamental concept, while the latter is a matter of clear representation. As for the former factor, the most easily used metric is the Euclidean norm $\|\cdot\|_E$. One example is the Topographic Product (TP) [2] proposed by Bauer and Pawelzik, which is an evaluation based on the agreement of the near-neighbor rankings in both the input and the output spaces. This measure is fast to compute, but it fails for nonlinear data manifolds. Villmann et al. proposed a more advanced measure, Topographic Function (TF) [1], which is based on the induced Delaunay triangulation D_M [3]. Compared with the TP, the TF is a better

* This work was partially supported by grant NNG05GA94G from the Applied Information Systems Research Program, NASA, Science Mission Directorate. Figures are in color, request color copy by email: llzhang@rice.edu, erzsebet@rice.edu

representation, in that it is an informative function showing the severity of violation for different folding lengths, rather than a single summary value.

Since the TF has only been demonstrated for low-dimensional data, we aim to find out if it is valid and helpful when dealing with high-dimensional data. We also propose a clearer representation, the Differential Topographic Function, derived from the TF, and another useful representation, the Weighted Differential Topographic Function, which provides advantageous views of topology violation by adding the connection strength (see in Section 2.3) as a weighting factor.

2 Measuring Approaches

2.1 Topographic Function (TF)

The definition of the TF is restated as the foundation, from which DTF and WDTF are developed. For a commonly used 2-d square SOM lattice, [1] defined for each processing element (PE, also called lattice node or neuron elsewhere) i

$$\begin{cases} f_i(k) \overset{def}{=} \#\left\{ j \mid \|i-j\|_{\max} > k \wedge d_{D_M}(i,j)=1 \right\} & k=1,2,\ldots,\max_{i,j\in A}\|i-j\|_{\max} \\ f_i(-k) \overset{def}{=} \#\left\{ j \mid \|i-j\|_{E} = 1 \wedge d_{D_M}(i,j)>k \right\} & k=1,2,\ldots,\max_{i,j\in A} d_{D_M}(i,j) \end{cases}.$$

The maximum-norm is defined as $\|\cdot\|_{\max} \overset{def}{=} \max_{j=1}^{d_A}\left|(\cdot)_j\right|$, where d_A denotes the dimensionality of the SOM lattice, and d_{D_M} indicates the graph distance in the induced Delaunay triangulation D_M [3]. (The definitions of the induced Delaunay triangulation and the corresponding graph distance metric are explained in [3].) The TF is then expressed as

$$\Phi(k) \overset{def}{=} \begin{cases} \frac{1}{N}\sum_{i\in A} f_i(k) & k>0 \\ \Phi(1)+\Phi(-1) & k=0 \\ \frac{1}{N}\sum_{i\in A} f_i(k) & k<0 \end{cases}\cdot$$

Positive k and negative k represent the foldings in the input and output spaces respectively. The largest k, which holds a nonvanishing value of the TF indicates the largest range of topology violation. In order to compare the TF between two different output lattice structures, k can be normalized to [-1, 1].

In the original definition of the TF, no distinction was made between active and dead PEs. Nevertheless, dead PEs are inegligible sources of topology violation. For example, in case of dimensional mismatch or highly structured data, dead PEs usually take the role of gaps to make a better topology match between the input and output spaces. Since we use 2-d SOMs, dimensional mismatch is not a rarity, so the question arises whether the violation induced by the dead PEs should be counted in the TF. We think it is reasonable to exclude them because of the following two points. 1. Dead PEs do not represent data; 2. Dead PEs that died during learning record their weights right at the moment of death. These weights might remain unchanged afterwards, hence can be a source of topology violation.

2.2 **Differential Topographic Function (DTF)**

Essentially, the TF is an integral function. To show the range of topology violation more clearly, a supplementary differential form can be calculated. We name it as the Differential Topographic Function. From the similar formulae

$$\begin{cases} g_i(k) \overset{def}{=} \#\left\{ j \mid \|i-j\|_{\max} = k \wedge d_{D_M}(i,j) = 1 \right\} & k = 2,..., \max\limits_{i,j \in A} \|i-j\|_{\max} \\ g_i(-k) \overset{def}{=} \#\left\{ j \mid \|i-j\|_E = 1 \wedge d_{D_M}(i,j) = k \right\} & k = 2,..., \max\limits_{i,j \in A} d_{D_M}(i,j) \end{cases},$$

the DTF can be obtained directly by the first difference of the TF

$$DTF(k) \overset{def}{=} \tfrac{1}{N} \sum_{i \in A} g_i(k) = \begin{cases} \Phi(k-1) - \Phi(k) & k \geq 2 \\ \Phi(k+1) - \Phi(k) & k \leq -2 \end{cases}.$$

It shows us the severity of the folds for each folding length k.

2.3 **Weighted Differential Topographic Function (WDTF)**

Due to outliers and noise in high-dimensional real world data and the limitation of our commonly used 2-d SOM, it is not surprising to have more violating connections than in low-dimensional data. However, if looked into closely, many long-range violating connections are very weak. Let us define the connection strength, denoted by $CONN(i,j)$, between prototypes i and j as the number of data points for which either i or j is the best matching unit and the other is the second best matching unit [4]. $CONN(i,j)$ obviously form an $N \times N$ matrix $CONN$, where N is the number of all SOM PEs or prototypes. Both TF and DTF count the number of violating connections regardless of their strengths. In order to show the relative importance of each connection, we incorporate the connection strength as a weighting factor into the DTF. For each PE i, define

$$h_i(k) \overset{def}{=} \sum_{\substack{\|i-j\|_{\max} = k, \\ d_{D_M}(i,j)=1}} CONN(i,j), \; k = 2,..., \max_{i,j \in A} \|i-j\|_{\max}.$$

Then the WDTF will be $WDTF(k) = \frac{1}{2P} \sum_{i \in A} h_i(k)$, where P is the total number of data points. WDTF is only defined for positive k, since the connectivity strength matrix $CONN$ is defined in the input space and used to describe the severity of violation in the SOM lattice. Fortunately, the SOM algorithm tends to eliminate the topology violation in the input space, i.e., TF often vanishes for negative k.

3 Experiments and Results

3.1 **8-class synthetic data**

This is a synthetic 6-element spectral image, where each of 128x128 pixels is represented by a 6-element vector. The image contains 8 spectral classes, created by adding ~10% Gaussian noise to class prototypes. For complete description, see [5].

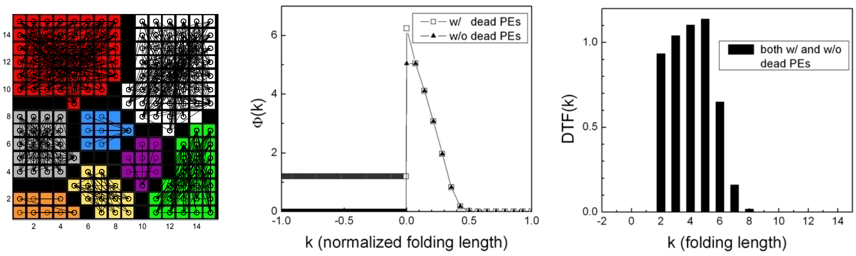

Fig. 1: SOM of the 8-class synthetic data with known class labels (colors) superimposed. The clustering learned by the SOM is perfect. **Left:** The 567 violating connections are visualized in the cluster map and all remain within clusters. **Middle:** The TF. **Right:** The DTFs are the same with and without counting the dead PEs.

All violating connections shown in Fig. 1 (left) are drawn on the SOM grid, according to the neighborhood relationship described by the induced Delaunay triangulation. The gaps formed by the empty PEs delineate the clusters learned by the SOM. None of the violating connections extend outside the clusters. Since the classes were assigned by construction, they represent the ground truth, and it is clear that the cluster structure formed by the SOM is a perfect match. In this sense, the TF is constructed on a reliable distance metric. The TF or the DTF in Fig. 1 show the same: all violating connections are local. No connection has a size greater than 8, the 'diameter' of the largest cluster. One might conclude that, for the purpose of cluster capture, this level of topology violation is inconsequential.

3.2 Lunar Crater Volcanic Field (LCVF) data

This is a noisy real world spectral image comprising ~250,000 194-d data points (image pixels). We have a credible manual clustering, shown in Fig. 2, to serve as ground truth [6]. The SOM was produced after 300K learning steps.

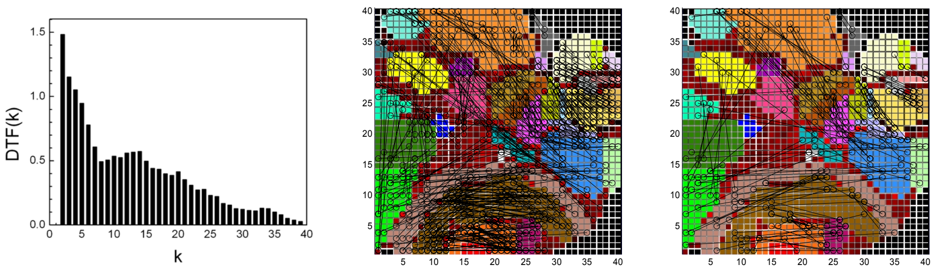

Fig. 2: 40×40 SOM of the LCVF data. **Left:** The DTF. **Middle:** Visualized in the cluster map are all 521 violating connections with strengths above 15. **Right:** Visualized are the 165 inter-cluster violating connections with strengths above 15.

As seen in Fig. 2, violating connections exist for every folding length. A threshold for the connection strength is necessary in Fig. 2 (middle) to enable us to see connecting lines rather than black clouds. The connections in the bottom annular area interestingly profile the shapes of the four adjacent clusters. In the upper right area, almost all connections follow the same direction, showing the close relationship along the line. The reason is that the spectra of these clusters along this direction form a series of slightly varying signatures. Due to the limitation of SOM dimension, their intertwined relationship is represented by topology violations. Since within-cluster violations are not important for cluster capture, Fig. 2 (right) gives the view of all inter-cluster violations above strength 15 ($\approx$ mean + one standard deviation). On the whole, the visualization of the violating connections reproduces what the credible clustering result tells us, though there are some long range connections which could be brought about by boundary effects.

A longer continued run to 8M steps was also done for this data. To compare its topology preservation quality with that of 300K steps, we plot the DTFs and some connection properties in Fig. 3. The two DTFs differ slightly, and there are even more long range violating connections (size 36-39) as we have more learning steps. Based solely on the DTF, one might say: 8M is worse than 300K. Although not shown here, the TF, an integral form of the DTF, indicates the same. However, the details of the connection strength for each folding length provide a different view.

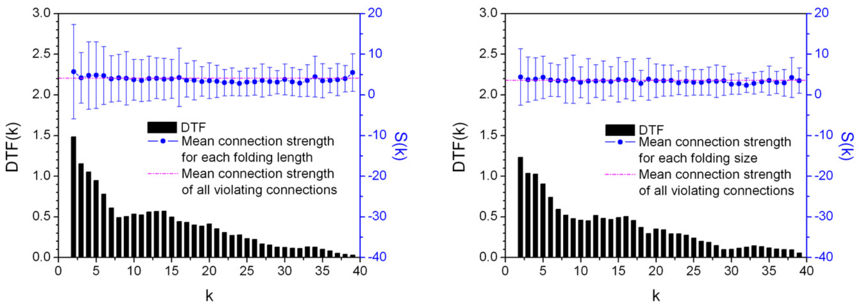

Fig. 3: Comparison of the DTF, the means and the standard deviations of the strengths of violating connections for each folding length, after 300K (**Left**) and 8M (**Right**) learning steps. The red lines indicate the total average strength of all violating connections, and the vertical bars indicate the standard deviations.

Fig. 3 shows the plots of the DTF and the properties of the connection strengths. The average strength of violating connections for folding length k is calculated as

$$S(k) = \frac{\sum_{i \in A} h_i(k)}{\sum_{i \in A} g_i(k)} = \frac{2P \times WDTF(k)}{N \times DTF(k)}.$$

As we see, the average strengths tend to get closer in a longer learning run. After 8M learning steps, most of the standard deviations shrink. Many high maximum peaks of the connection strengths are significantly lowered too (not shown here).

Fig. 4 (left) is a plot of two normalized DTFs, denoted by NDTF, and calculated as

$$NDTF(k) = \frac{N \times DTF(k)}{2C},$$

where C is the total number of connections. It shows the distribution (percentage) of the number of connections across folding lengths. Likewise, the meaning of the WDTF in Fig. 4 (right) is the percentage of connection strength for each k. The WDTF shows more clearly that the 8M learning steps quench many high peaks, especially the local ones, compared with the 300K case. It tells us that even though the number of the violating connections does not change much (Fig. 4 left), many violating connections become weaker (Fig. 4 right). By this graph, we would conclude that the 8M steps result is better. If calculated, the ratio of the violating connections to the total connections also decreases from 73.6% to 70.9%, and the percentage of the number of the data points involved in the non-violating connections increases from 80.3% to 83.7%. These data support our conclusion.

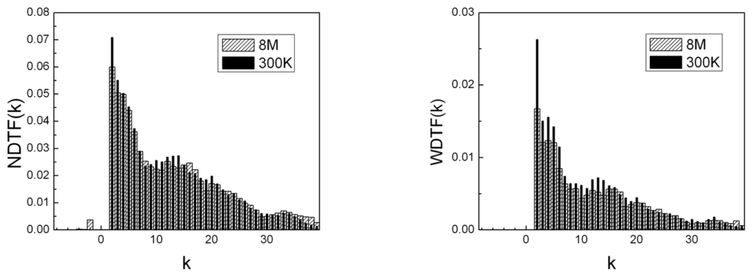

Fig. 4: Comparisons between the results after 300K and 8M learning steps, superimposed. The NDTFs (**Left**) and the WDTFs (**Right**) represent the percentage of connections and the percentage of connection strength for each folding length, respectively.

4 Conclusion

From our experiments, we conclude that the distance metric embedded in the definition of the TF is reasonable and reliable. Therefore, with the help of the DTF, the TF provides a clear overview of the severity and range of topology violation. However, we should be careful before stamping an SOM "good" or "bad" in topology preservation simply based on the TF result. The statistics of the connection strengths, as seen in the LCVF example, show more detailed change, not discovered by the TF. WDTF, a novel measure, can provide more accurate interpretation of the severity of violation, by adding the connection strength as a weighting factor.

The authors thank Dr. Thomas Villmann for valuable discussions.

References

[1] T. Villmann, R. Der and M. Herrmann, Topology preservation in self-organizing feature maps: exact definition and measurement, *IEEE Transactions on Neural Networks*, 8:256-266, 1997.

[2] H.-U. Bauer and K. Pawelzik, Quantifying the neighborhood preservation of self-organizing feature maps, *IEEE Transactions on Neural Networks*, 3:570-579, 1992.

[3] T. Martinetz and K. Schulten, Topology representing networks, *Neural Networks*, 7:507-522, 1994.

[4] K. Tasdemir and E. Merényi, Considering topology in the clustering of self-organizing maps, *Proc. 5th Workshop on Self-Organizing Maps, Paris, 5th -8th September*, pp.439-446, 2005.

[5] E. Merényi and A. Jain, Forbidden Magnification? II., *Proc. 12th European Symposium on Artificial Neural Networks, Bruges, Belgium, 28th-30th April*, pp.57–62, 2004.

[6] T. Villmann and E. Merényi, Extensions and Modifications of the Kohonen-SOM and Applications in Remote Sensing Image Analysis. In *"Self-Organizing Maps: Recent Advances and Applications"* (U.Seiffert and L.C. Jain Eds.), Springer-Verlag, pp.121-145, 2001.