ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 205-210

Measuring Topology Preservation in Maps of Real-World Data

M. Herrmann[1], H.-U. Bauer[1], Th. Villmann[2]

[1] MPI für Strömungsforschung, Bunsenstr. 10, 37073 Göttingen, Germany
[2] Universität Leipzig, Klinik für Psychosomatische Medizin
Karl-Tauchnitz-Str. 25, 04107 Leipzig, Germany

Abstract. Topography of neural maps is an advantageous property, e.g. in the presence of noise in a transmission channel or for data visualization. Yet, this property is difficult to define and to quantify. Reviewing some recently proposed measures to quantify topography, we give results for maps trained on synthetic data as well as on four real-world data sets. The measures are found to do not a perfect, but an adequate job, e.g. in selecting a topographically optimal output space dimension.

1. Introduction

Even though many people seem to share a common intuitive understanding of what is meant by topology preservation in neural maps, it has turned out to be difficult to pinpoint this notion in an unequivocal and mathematically rigorous fashion. In a colloquial manner, topology preservation (or topography) of a map means the mapping of similar data points to close locations in the map layer. Thus, topography of a map is related to the notion of similarity, which may be defined on various levels leading to different methods for defining and measuring topography. Distances [5] and distance rankings [2, 1] as well as geometrical [9] or topological relations [8] have been used for quantifying topography. Each approach is motivated and illustrated by convincing, albeit handcrafted examples. Real-world-data on the other hand could pose unforeseen problems due to high dimensionality and sparseness, spatial heterogeneity, multi-fractal structure, scaling ambiguity, as well as the low quality of the data including redundancy, noise, or missing values.

The present paper aims at relating different topography measures, and at comparing their power of analysis for maps generated for different, illustrative and challenging, data sets. At any level of generality a consequence of topography is the continuity of a map. Therefore, the task of determining the effective data dimension seemed to us particularly suited for a fair comparison.

2. Measuring topology preservation

A neural map Ω assigns outputs $\mathbf{r} \in \mathcal{V}$ to inputs $\mathbf{v} \in \mathcal{A}$. The resp. dimensions are $D^{\mathcal{V}}$ and $D^{\mathcal{A}}$. In our simulations all maps have been produced by Koho-

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 205-210

nen's self-organizing maps (SOMs) [7]. We should remark, however, that the measures given below[1] characterize the topography of maps irrespective of the algorithm that generated the map.

The topographic product P [1] relates for each neuron the sequence of input space neighbors to the sequence of output space neighbors. The sign of P approximately indicates the relation between input and output space topology. $P < 0$ corresponds to a too low-dimensional input space and $P > 0$ to a too high-dimensional input space. $P \approx 0$ indicates an approximate match.

Spearman's ρ has been used in [2] to express the fact that "the *relative positions* of all neighbors of every point" [2] are to be preserved. Formally, this involves a normalized mean squared difference between distance ranks in input and output space. ρ takes values in $[-1, 1]$ with $\rho = 1$ indicating exact *metric topology preservation.*

The Zrehen-measure Z quantifies "the correctness of the following statement: A pair of neighbor cells $\mathbf{r}$ and $\mathbf{r}'$ is locally organized if the straight line joining their weight vectors $\mathbf{w_r}$ and $\mathbf{w_{r'}}$ contains points which are closer either to $\mathbf{w_r}$ or to $\mathbf{w_{r'}}$ than they are to any other" [9]. All ($Z = 0$)-maps are considered to be topology preserving. This holds in particular for dimension-reducing maps. If $D^{\mathcal{A}}$ exceeds the effective data dimension, then $Z > 0$. Using this measure the optimal output dimension is selected as the largest $D^{\mathcal{A}}$ with Z being not significantly different from zero.

In the **topographic function** Φ approach [8] a graph of neighborhood relations between the reference vectors is constructed by the method of masked Voronoi polygons. The function value $\Phi(k)$, $k > 0$, counts how many pairs of nearest neighbors in the output space have a corresponding pair in the input space with neighborhood order larger than k. Backward-mapping distortions are recorded analogously in the negative-k terms. If $\Phi(k) = 0$ for all k the map is considered to be perfectly topology preserving. The values of the negative (positive) k components of the topographic function indicate whether the output dimension is too high (small). Thus, the sign of $\bar{\Phi} = \Phi(+1) - \Phi(-1)$ expresses what terms are predominant. The shape of $\Phi(k)$ allows a detailed discussion of the magnitude of distortions occurring in a map.

Fractal dimension analysis is another approach to the determination of the effective data dimension D^{eff}. To check the consistency of our topography results, we also determined the Grassberger-Procaccia-dimension D^{GP} [6] for our data sets. For comparison with the maps, we evaluated D^{GP} at that length scale which corresponds to the size of the Voronoi regions induced by the maps.

3. Illustrative examples

We have calculated the topography measures for maps[2] from a two-dimensional square-shaped input space to output spaces which form a line, a square or a

[1] For brevity we have to refer to the cited papers for details.

[2] All maps were trained by Kohonen's algorithm: 10^5 learning steps, neighborhood width decreased from a third of the output space diameter to 0.1, learning step size decreased from 0.9 to 0.01, three map realizations per output dimension $D^{\mathcal{A}}$ and node number N.

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 205-210

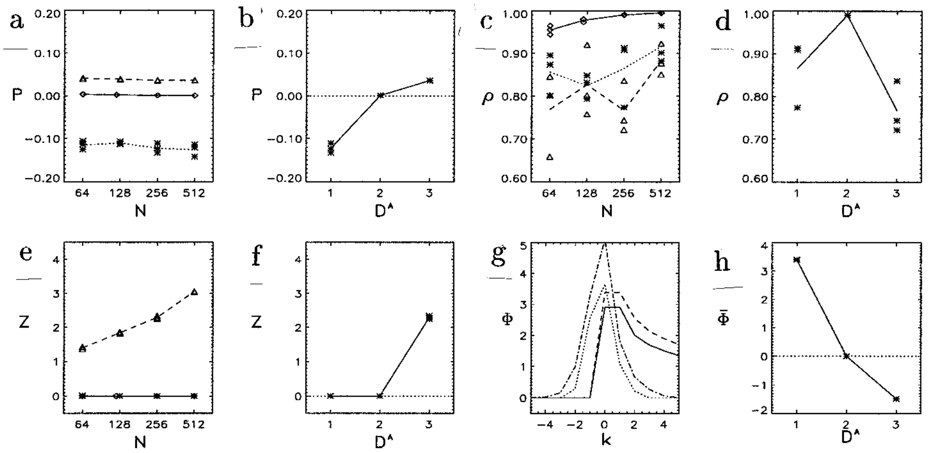

Figure 1: **(a)** *Values of the topographic product* P *as a function of node number for SOMs mapping the unit square to lines with* $N = 64, 128, 256, 512$ *units (asterisks, averaged values yield dotted lines), square grids with* $N = 8\times8, 11\times 11, 16\times16, 22\times23$ *units (diamonds, solid lines), and cubes with* $N = 4\times4\times 4, 5\times5\times5, 7\times6\times6, 8\times8\times8$ *units (triangles, dashed line).* **(b)** P *as a function of output space dimension* $D^{\mathcal{A}}$ *at* $N \approx 256$ *nodes. The zero of* P *at* $D^{\mathcal{A}} = 2$ *clearly selects* $D^{\mathcal{A}} = 2$ *as optimal.* **(c)** *Same as* (a)*, but for Spearman's* ρ*. The variability of the values indicates a considerable fluctuation of this measure for non-optimal maps.* **(d)** *The maximum of the averaged values nevertheless uniquely determines* $D^{\mathcal{A}} = 2$*.* **(e)** *Same as above, but for the Zrehen-measure. Maps with* $D^{\mathcal{A}} = 1,\ 2$ *yield* $Z = 0$*. The strong increase of* Z *at* $D^{\mathcal{A}} = 3$ *in* **(f)** *indicates that the* $D^{\mathcal{A}} = 2$*-maps are topographically optimal.* **(g)** *Topographic functions* $\Phi(k)$ *for the same maps (solid:* $D^{\mathcal{A}} = 1$, $N = 64$*, dashed:* $D^{\mathcal{A}} = 1$, $N = 256$*, dotted:* $D^{\mathcal{A}} = 3$, $N = 4\times4\times4$*, dash-dotted:* $D^{\mathcal{A}} = 3$, $N = 7\times6\times6$*). For maps with* $D^{\mathcal{A}} = 2$ *we have* $\Phi(k) = 0$ *indicating optimality (not depicted).* **(h)** $\bar{\Phi} = \Phi(+1) - \Phi(-1)$ *as a function of* $D^{\mathcal{A}}$, $N \approx 256$.

cube. In a second example a curved two-dimensional surface embedded in three dimensions was combined with seven additional noisy dimensions to form ten-dimensional inputs.

Among these maps, all of the topography measures perform quite well, see Figs. 1, 2. The ($D^{\mathcal{A}} = 2$)-maps are identified as the most topography preserving.

In the second example in addition to dimension selection also a scale selection task was addressed. The correct aspect ratio (28×9) between the different directions within a given output space dimensionality was picked out by all maps. The topography values of the other ($D^{\mathcal{A}} = 2$)-maps were still substantially better than those for maps with other output dimensions. For moderate noise $\nu = 0.1$ the measures based on the graph distance failed in both tasks, whereas the other ones did not show any significant deterioration. Finally, for $\nu = 0.2$ all measures indicated that an output dimension $D^{\mathcal{A}} = 2$ failed to be topographic, since the total noise variance exceeded the variance of the two-dimensional signal.

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 205-210

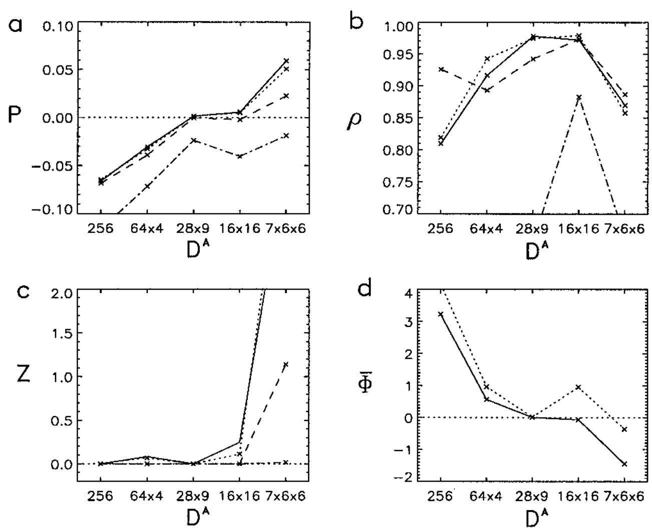

Figure 2: *Results of the topography measures for the synthetic nonlinear noisy data set. Solid lines: no noise ($\nu = 0$), dotted lines: $\nu = 0.05$, dashed lines: $\nu = 0.1$, dash-dotted lines: $\nu = 0.2$. The selection of the correct output space scale (28×9) can be seen, as well as the breakdown of topography for noise levels above $\nu = 0.1$.*

4. Examples involving real-world data

We checked for possible consequences of these effects by generating SOM-maps for experimental data from a chaotic dynamical system (time delay coordinates of a driven pendulum with embedding dimension $D^{\mathcal{V}} = 5$, $n = 28000$ data points), speech data (German numerals, $D^{\mathcal{V}} = 19$, $n = 2013$), and two types of image data (Lena image, $D^{\mathcal{V}} = 16$ combined from patches of 4×4 graylevel pixels, $n = 16384$; and Landsat image data, $D^{\mathcal{V}} = 6$, $n = 65536$). In each case we trained maps with varying output dimension $D^{\mathcal{A}}$ and tried to spot the most topography preserving among these. The total number of neurons was $N \approx 256$ in each map, other parameters as in described in Sect. 3.

Table 1 lists results obtained at these data sets. For the pendulum data the $D^{\mathcal{A}} = 3$-maps are selected by all measures to be the most topography preserving. This is in good agreement with the relevant fractal dimension to be $D^{\text{GP}} \approx 3.1$. On the speech data the different measures pick out maps with dimension $D^{\mathcal{A}} = 2-3$, but an clear-cut decision cannot be drawn. However, the performance of speech recognition systems which make use of the topography of the map will certainly decrease for ($D^{\mathcal{A}} = 1$)- or ($D^{\mathcal{A}} \geq 4$)-maps. On the Lena image ρ and Z both yield $D^{\mathcal{A}} = 2$, P yields $D^{\mathcal{A}} = 2 - 3$ and $\bar{\Phi}$ yields $D^{\mathcal{A}} = 3 - 4$. Apart from $\bar{\Phi}$, the different measures agree quite well with the value of a Grassberger-Procaccia-analysis ($D^{\text{GP}} \approx 1.9$). Maps of the Landsat data are optimally topographical at $D^{\mathcal{A}} = 2$, as compared to $D^{\text{GP}} \approx 1.7$ in a Grassberger-Procaccia-analysis. In this example, the topographic function failed to yield a sensible value for the optimal output space dimension $D^{\mathcal{A}}$.

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 205-210

data set	measure	D^A=1	D^A=2	D^A=3	D^A=4	D^A=5
Pendulum	P	-0.2709	-0.0444	-0.0033	0.0117	0.0368
	ρ	0.7586	0.8733	0.8879	0.7658	0.5836
	Z	0.0065	0.1465	0.3223	1.0516	2.4519
	$\bar{\Phi}$	7.2917	3.8177	0.6005	-1.3203	-2.5981
DPI	P	-0.2285	-0.0554	-0.0035	0.0195	0.1770
	ρ	0.6639	0.8568	0.6071	0.6540	0.3113
	Z	0.0118	0.2194	0.6405	1.3537	2.6296
	$\bar{\Phi}$	2.2891	-0.0469	-2.0238	-3.3438	-3.8848
Lena	P	-0.086	-0.022	0.029	0.078	0.134
	ρ	0.584	0.761	0.667	0.530	0.368
	Z	0.068	0.627	2.452	5.730	9.369
	$\bar{\Phi}$	7.401	4.135	1.508	-0.789	-2.373
Landsat	P	-0.117	-0.018	0.024	0.071	
	ρ	0.697	0.813	0.733	0.556	
	Z	0.002	0.080	0.859	3.360	
	$\bar{\Phi}$	13.17	9.08	5.42	2.05	

Table 1: Values of the topography measures for maps of different output dimensionality D^A, each value averaged over three maps, for different data sets.

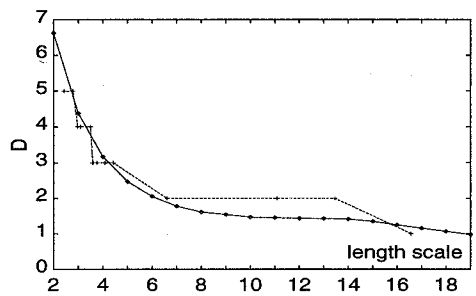

Figure 3: *Dimensional estimates D^{GP} (solid line) and D^{CCA} (dashed line) as a function of underlying length scales for the Landsat data set (cf. Sect. 4.).*

5. SOMs, PCA, and MDS

SOMs have been compared (cf. e.g. [4, 2]) to other data representation algorithms such as principal component analysis (PCA) and multi-dimensional scaling (MDS). Since topology preserving mapping does not aim at exactly preserving inter-data distances nor a detailed clustering structure of the data, comparisons (e.g. [4] on distortion error, metric topology preservation, and clustering quality) appear unfavorable for SOM, although the outcome is often a close call. If one focuses on global data features the capabilities specific to SOMs are revealed more clearly. For dimensional analysis via PCA we required a large portion of the variance of the data to be covered by projections onto a small number D^{PCA} of principal components. For the Landsat data ($D^{\mathrm{PCA}} = 2$) and the pendulum data ($D^{\mathrm{PCA}} = 3$) higher components had obviously small eigenvalues. The Lena data and the synthetic data set are dominated by one eigenvector which would suggest $D^{\mathrm{PCA}} = 1$. The relatively

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 205-210

small set of DPI data did not allow any conclusion, since the first three components accounted only for 75 % of the variance and higher components decayed very slowly only.

Multi-dimensional scaling minimizes directly the mean squared difference between inter-data distances in the input and output space. The results obtained are equivalent to PCA when Euclidean metrices are used. Convergence problems occur when a large number of data is mapped directly, such that a vector quantization (cf. [3]) or clustering (cf. [4]) of the data is necessary. Among the various weighting schemes a similarity to SOMs is obtained when small distances in the output space are given a largers weights, cf. [3]. Varying the output dimension the data dimension could be resolved for a range of length scales which are determined by the number of quantization units (cf. Fig. 3).

6. Conclusion

The different topography measures not only coincide among themselves but they also designate a map with a value of $D^{\mathcal{A}}$ which corresponds to the smallest integer greater than D^{GP}. For the justification of the neighborhood graph representation of the input space topology, in [8] the case of a strongly curved data manifold was referred to. The real-world data studied here suggest that noisyness and sparseness of the data are the more relevant problems to be considered when measuring topography. Provided an appropriate evaluation of the maps is performed, SOMs have superior capabilities in revealing global data features such as the effective data dimension.

References

[1] H.-U. Bauer, K. Pawelzik, Quantifying the Neighbourhood Preservation of Self-Organizing Feature Maps. *IEEE TNN* **3**(4), 570-579 (1992).

[2] J. C. Bezdek, N. R. Pal, An Index of Topology Preservation for Feature Extraction, *Patt. Recogn.* **28**, 381-391 (1995).

[3] P. Demartines, J. Hérault, CCA: "Curvilinear Component Analysis". In: *Quinzieme Coloque Gretsi*, Juan-les-Pins (1995).

[4] A. Flexer, Limitations of Self-Organizing Maps for Vector Quantization and Multidimensional Scaling. To appear in: *Neur. Inf. Proc. Syst.* (1997).

[5] G. J. Goodhill, S. Finch, T. J. Sejnowsky, A Unifying Measure for Neighborhood Preservation in Topographic Mappings. In: *Proc. 2nd Joint Symp. Neur. Comp.* Inst. f. Neur. Comp., La Jolla, CA, **5**, 191-202 (1995).

[6] P. Grassberger, I. Procaccia, Measuring the Strangeness of Strange Attractors. *Physica D* **9**, 189-208 (1983).

[7] T. Kohonen, *Self-Organizing Maps.* Springer, Berlin (1995).

[8] Th. Villmann, R. Der, M. Herrmann, Th. Martinetz, General Definition of Topology Preservation in Self-Organizing Feature Maps and the Topographic Function for its Measurement. To appear in *IEEE TNN* (1997).

[9] S. Zrehen, Analyzing Kohonen Maps With Geometry. In: *Proc. ICANN'93*, Eds. St. Gielen and B. Kappen, Springer, London (1993).