Kaski, S. and Lagus, K. (1996) Comparing Self-Organizing Maps. In C. von der Malsburg, W. von Seelen, J. C. Vorbrüggen, and B. Sendhoff (Eds.) *Proceedings of ICANN96, International Conference on Artificial Neural Networks*, Lecture Notes in Computer Science vol. 1112, pp. 809-814. Springer, Berlin.

Comparing Self-Organizing Maps

Samuel Kaski and Krista Lagus

Helsinki University of Technology
Neural Networks Research Centre
Rakentajanaukio 2C, FIN-02150 Espoo, Finland

Abstract. In exploratory analysis of high-dimensional data the self-organizing map can be used to illustrate relations between the data items. We have developed two measures for comparing how different maps represent these relations. The other combines an index of discontinuities in the mapping from the input data set to the map grid with an index of the accuracy with which the map represents the data set. This measure can be used for determining the goodness of single maps. The other measure has been used to directly compare how similarly two maps represent relations between data items. Such a measure of the dissimilarity of maps is useful, e.g., for analyzing the sensitivity of maps to variations in their inputs or in the learning process. Also the similarity of two data sets can be compared indirectly by comparing the maps that represent them.

1 Introduction

The self-organizing map (SOM) [4, 5] algorithm forms a kind of a nonlinear regression of an ordered set of reference vectors m_i, $i = 1, \ldots, N$, into the data space $\Re^n$. Each reference vector belongs to a map unit on a regular map lattice. In exploratory data analysis (data mining) with the SOM the aim is to extract and *illustrate* the essential structures within a statistical data set by a map that, as a result of an unsupervised learning process, follows the distribution of the data in the input space. Each data sample is mapped to the unit containing the most similar reference vector, whereby the relations of the data samples become reflected in geometrical relations (order) of the samples on the map. The density of the data points in different regions of the input space (reflected in the distances between the reference vectors of neighbor units) can be visualized with gray levels on the map display [6, 9].

2 Measures of Goodness of Maps

Measures are needed for choosing good maps from a sample set of maps resulting from a stochastic learning process, or for determining good learning parameters for the maps.

2.1 Previously Proposed Measures

The accuracy of a map in representing its input can be measured with the average *quantization error*, i.e., the distance from each data item to the closest reference vector. If also the distance from the reference vectors of the *neighbors* (units that lie within a specified radius on the map grid) of the winner is incorporated [5], the measure becomes sensitive to the local orderliness of the map. Although these two measures are necessary in guaranteeing that the map represents the data set well, they cannot be used to compare maps with different stiffnesses since they favor maps with specific neighborhood radii.

Several orderliness measures have been proposed that compare the relative positions of the reference vectors in the input space with the positions of the corresponding units on the map lattice (e.g., [1]). As has been pointed out by Villmann et al. [10], however, these measures cannot distinguish between folding of the map along nonlinearities in the data manifold and folding within a data manifold. The former is a highly desirable property whereas the latter causes discontinuities in the mapping from the input space to the map grid, which may be undesirable in some applications.

A more sensitive measure [10] computes the adjacency of the "receptive fields", or cells in the Voronoi tessellation, of the different map units *within the data manifold.* In a perfectly ordered map only units that are neighbors on the map lattice may have adjacent receptive fields. A possible problem with this measure is that noise or nonrepresentative inputs may easily cause some receptive fields to be erroneously judged as adjacent within the manifold.

Kiviluoto [3] has used a more gradual measure of the adjacency of the receptive fields: The proportion of samples for which the nearest and the second nearest units reside in non-neighboring locations on the map. Even this measure does not, however, consider the *extent* of the discontinuities in the mapping from the input space to the map grid.

Kraaijveld et al. [6] have compared different mapping methods by computing the accuracies with which a given data set can be classified in the mapped spaces. Although their goodness measure is not sufficiently general for our purposes since it requires classified input samples, the way they computed distances between data points has been found useful also in our studies.

2.2 A Novel Measure

We formed a measure that combines an index of the continuity of the mapping from the data set to the map grid with a measure of the accuracy of the map in representing the set (the quantization error). For each data item x we compute the distance $d(x)$ *from* x *to the second nearest reference vector* $m_{c'(x)}$ passing first from x to the best matching reference vector $m_{c(x)}$, and thereafter *along the shortest path* to $m_{c'(x)}$ through a series of reference vectors. In the series each reference vector must belong to a unit that is an immediate neighbor of the previous unit. If there is a discontinuity in the mapping near x, such a distance along the map from unit $c(x)$ to $c'(x)$ is in general large, whereas if the units are neighbors the distance is smaller.

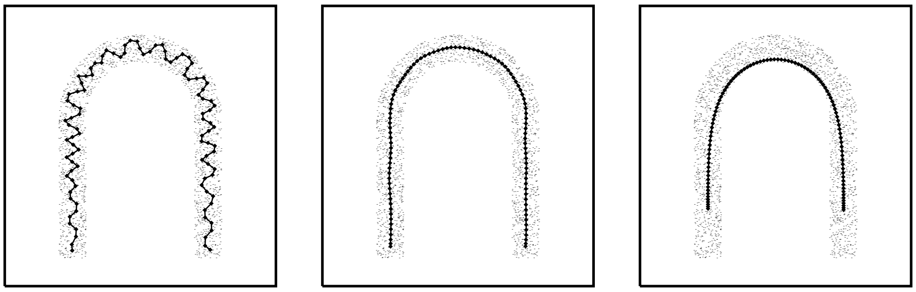
The distance $d(x)$ can be expressed more formally as follows: Denote by $I_i(k)$ the index of the kth unit on a path along the map grid from unit $I_i(0) = c(x)$ to $I_i(K_{c'(x),i}) = c'(x)$. In order for the function I_i to represent a path along the map grid the units $I_i(k)$ and $I_i(k+1)$ must be neighbors for $k = 0, \ldots, K_{c'(x),i} - 1$. Using these notations the distance $d(x)$ is

$$d(x) = ||x - m_{c(x)}|| + \min_i \sum_{k=0}^{K_{c'(x),i}-1} ||m_{I_i(k)} - m_{I_i(k+1)}|| \ . \tag{1}$$

The goodness C of the map is defined as the average (denoted by E) of the distance over all input samples (low values denote good maps),

$$C = E[d(x)] \ . \tag{2}$$

In simulations with a simple data set (Fig. 1) C measured a satisfactory combination of the continuity of the mapping and the quantization error, a result not obtainable with the previously proposed methods.



$C = 0.052$ $C = 0.043$ $C = 0.059$

Fig. 1. The goodness measure C of SOMs with varying stiffnesses produced by varying the final neighborhood width in the learning process. The input (small dots) came from a two-dimensional, horseshoe-shaped distribution. The reference vectors of the 100-unit, one-dimensional SOMs are shown in the input space as large black dots, with lines connecting reference vectors belonging to neighbor units. The best (lowest) value of C is yielded by the SOM in the middle that covers all of the horseshoe without folding unnecessarily.

3 A Novel Measure of Dissimilarity of Maps

For a given data set there may exist several different representations that are all useful for different purposes. Therefore it may not always be sensible to compare the goodnesses of the maps as was done in Sec. 2.2. It might in any case be useful to know *how* different the maps are from each other. A measure of the dissimilarity of maps could be used, e.g., for detecting outlier maps or for analyzing the sensitivity of the maps for variations in the inputs or in the learning process.

We define the dissimilarity of two maps, L and M, as the average (normalized) difference in how they represent the distance between two data items. The *representational distance* $d_L(x,y)$ between the pair (x,y) of data samples, represented by map L, is defined as follows. The distance is computed along the shortest path which passes through the best matching reference vectors $m_{c(x)}$ and $m_{c(y)}$, and through a series of reference vectors. In the series the units corresponding to each successive pair of reference vectors must be immediate neighbors. Using the notation introduced in Sec. 2.2, denote by $I_i(k)$ the index of the kth unit on a path from $I_i(0) = c(x)$ to $I_i(K_{c(y),i}) = c(y)$. The distance between samples x and y on map L is then

$$d_L(x,y) = \|x - m_{c(x)}\| + \min_i \sum_{k=0}^{K_{c(y),i}-1} \|m_{I_i(k)} - m_{I_i(k+1)}\| + \|y - m_{c(y)}\| \;, \quad (3)$$

and the dissimilarity of maps L and M is defined to be

$$D(L,M) = E\left[\frac{|d_L(x,y) - d_M(x,y)|}{d_L(x,y) + d_M(x,y)}\right] . \quad (4)$$

Here the expectation E is estimated over all pairs of data samples (x,y) in a representative set. To reduce the computational complexity of the measure the reference vectors of one or all of the maps can be used as the representative set.

It can be shown that D is a dissimilarity measure in the mathematical sense. To demonstrate that D does indeed measure the dissimilarity of maps we have applied it in a case study to compare maps that had progressively more different input data sets (Fig. 2).

4 A Demonstration of the Use of the Dissimilarity Measure

Assume a scenario where SOMs are used by several parties to explore their data sets and to present summaries of the data. The parties could be individual people, institutions, or software agents, and the data sets might consist of information about any specific topic area, e.g., encoded documents or economical statistics (cf. [2, 5]). The parties might make the SOMs accessible through, for example, the Internet as advertisements or reports of their work, although they might not want to open their data sets for public use, e.g., due to confidentiality or the size of the data.

The SOMs are representations of the knowledge, or “expertise”, inherent in the data sets of the parties. It might therefore be of interest for the parties to assess the similarity of their SOMs. We have demonstrated the use of the measure D (4) in comparing maps describing different phonemes (Fig. 3). Maps taught with similar data sets (e.g., /m/ and /n/) were found to be more similar than maps taught with dissimilar sets (e.g., /m/ and /s/).

The significance of the measured dissimilarity between two maps could be assessed by computing the probability that the maps represent the same data set, for example using a nonparameteric statistical test. The baseline distribution of the dissimilarities, under the hypothesis that the maps have been taught with the

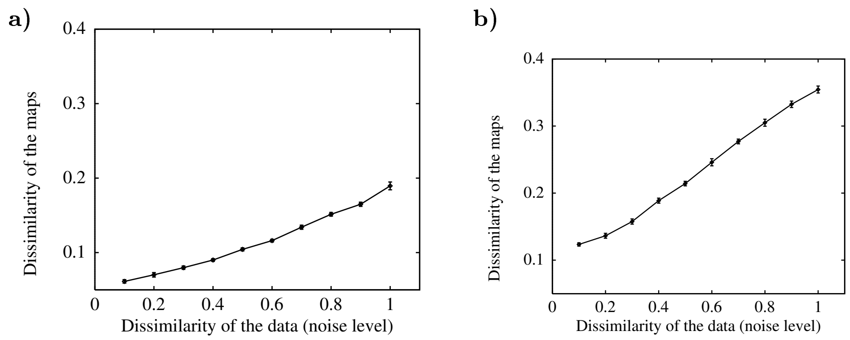


Fig. 2. Demonstration of a sensitivity analysis using the dissimilarity measure. Varying amounts of noise were added to a data set that consisted of 39 indicators for each country in a set of 78 countries, describing different aspects of their welfare [2]. The dissimilarity D between the SOMs taught with noisy data and a SOM taught with the original data set was computed when (a) the maps were of equal size (13 by 9 units) and had equal learning parameters (the final width of the neighborhood was two), and (b) when the map taught with the noisy data was different in size (16 by 7 units) and had different learning parameters (final neighborhood width was one instead of two). In both cases the dissimilarity D of the maps increased when the dissimilarity of their inputs increased. The bars in the figure denote the standard errors of the means of ten distances computed between maps that had different random input sequences while learning. The noise level is the standard deviation of the i.i.d. Gaussian noise. The variance of each data dimension was normalized to unity.

same data set, can be formed by teaching a set of maps with different (stochastic) input sequences. Also different stochastically chosen learning parameters and initial states can be used if the learning procedures of the maps are unknown.

5 Discussion

We have proposed for the comparison of SOMs two measures that are suitable especially for data mining applications. In data mining the map lattice must for illustratory purposes be regular and of a low dimension, whereby neither a perfectly topography preserving mapping [7] nor matching of the dimensions of the map and the input space [8] would be useful in general.

The proposed measure of the goodness of a map can be used to choose maps that do not fold unnecessarily in the input space while representing the input data distribution.

The measure of the dissimilarity of two maps can be used to compare directly how the maps illustrate relations between data items.

In the measures, the representational distances between data points are computed in the input space along paths following the "elastic surface" formed by the SOM. Such distances reflect the perceptual distance of data items on a map display, on which distances between neighboring reference vectors have for data mining purposes been illustrated with gray levels.

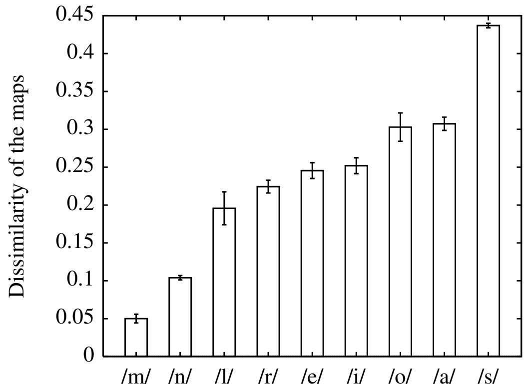


Fig. 3. Demonstration of the use of the dissimilarity measure D for comparing SOMs representing different data sets. The sets consisted of 20-dimensional short-time cepstra collected around the middle parts of phonemes of one male speaker (over 900 samples in each class). For each data set, 10 maps of the size of 6 by 4 units were taught using different random input sequences. The average of the distances between those maps and a common reference map are shown in the figure, together with the standard deviations. The reference map was chosen (based on the goodness measure C) from a batch of maps representing the set /m/.

References

1. Bauer, H.-U., Pawelzik, K. R.: Quantifying the neighborhood preservation of self-organizing feature maps. IEEE Tr. Neural Networks **3** (1992) 570–579
2. Kaski, S., Kohonen, T.: Exploratory data analysis by the self-organizing map: Structures of welfare and poverty in the world. *In* Neural Networks in the Capital Markets, World Scientific (to appear)
3. Kiviluoto, K.: Topology preservation in self-organizing maps. *In* Proc. ICNN96, IEEE Int. Conf. on Neural Networks (to appear)
4. Kohonen, T.: Self-organized formation of topologically correct feature maps. Biol. Cybern. **43** (1982) 59–69
5. Kohonen, T.: Self-Organizing Maps. Springer, Berlin (1995)
6. Kraaijveld, M. A., Mao, J., Jain, A. K.: A non-linear projection method based on Kohonen's topology preserving maps. *In* Proc. 11ICPR, 11th Int. Conf. on Pattern Recognition. IEEE Comput. Soc. Press., Los Alamitos, CA (1992) 41–45
7. Martinetz, T., Schulten, K.: Topology representing networks. Neural Networks **7** (1994) 507–522
8. Speckmann, H., Raddatz, G., Rosenstiel., W.: Considerations of geometrical and fractal dimension of SOM to get better learning results. *In* M. Marinaro and P. G. Morasso, eds, Proc. ICANN94, Int. Conf. on Artificial Neural Networks. Springer, London (1994) 342–345
9. Ultsch, A., Siemon, H. P.: Kohonen's self organizing feature maps for exploratory data analysis. *In* Proc. INNC90, Int. Neural Network Conf. Kluwer, Dordrecht (1990) 305–308
10. Villmann, T., Der, R., Martinetz, T.: A new quantitative measure of topology preservation in Kohonen's feature maps. *In* Proc. ICNN'94, IEEE Int. Conf. on Neural Networks. IEEE Service Center, Piscataway, NJ (1994) 645–648