MULTIDIMENSIONAL SCALING AND KOHONEN'S SELF-ORGANIZING MAPS

Włodzisław Duch and Antoine Naud
Department of Computer Methods, Nicholas Copernicus University,
ul. Grudziądzka 5, 87–100 Toruń, Poland.
e-mail: duch @ phys.uni.torun.pl, naud @ phys.uni.torun.pl
WWW: http://www.phys.uni.torun.pl

Two methods providing representation of high-dimensional (input) data in a lower-dimensional (target) space are compared. Although multidimensional scaling (MDS) and Kohonen's self-organizing maps (SOM) are dedicated to very different applications both methods are based on an iterative process that tends to approximate the topography of high-dimensional data and both can be used to model self-organization and unsupervised learning. In general it is impossible to find a lower-dimensional representation that preserves exactly the topography of high-dimensional data. An error function is defined to measure the quality of representations and is minimized in an iterative process. The minimal error measures the unavoidable distortion of the original topography represented in the target space.

I. INTRODUCTION

Human experts can analyze data in at most three dimensions, therefore evaluation of high-dimensional data is possible only if its dimensionality is reduced. Such reduction is also useful in pattern recognition [1] where "the curse of dimensionality" plagues many computational procedures. Statistical method known as the multidimensional scaling (MDS) technique has foundations in the work of Torgerson [2] and in the Coombs theory of data [3]. Computer programs and applications of MDS have been developed, among others, by Kruskal [4] at the Bell Laboratories, by Lingoes, Roskam and Borg [5] in Ann Arbor, and by Shepard [6] in Palo Alto. MDS was used by experts in mathematical psychology who wanted to obtain a lower-dimensional representation of psychological data. These data are related to perception (perceived nearness of objects, preferences, feature intensity or affinity) and are obtained by subjective evaluation of similarities or dissimilarities between different items, characterizing a small part of human psychological spaces. MDS techniques were developed to provide a two or three-dimensional image of the observed data, reducing their complexity and allowing their analysis by a human expert. Unfortunately MDS methods are almost unknown outside the mathematical psychology field. A very similar visualization method, called the nonlinear mapping, has been developed by Sammon [7].

The self-organizing map (SOM) algorithm introduced in 1981 by Kohonen [8] is usually presented as a particular type of artificial neural network. The network is first trained on the high-dimensional samples, in such a way that the weight vectors of the array of neurons in the output layer tend to approximate the probability density function of the high-dimensional data. This iterative learning process is unsupervised, or self-organizing, since there is no intervention of a teacher giving information about clusters that exist in the data space. Kohonen was working in pattern recognition (especially in speech recognition) when he developed the SOM algorithm, and his intentions were the following [9]: "The SOM has not been meant for statistical pattern recognition; it is a clustering, visualization, and abstraction method. Anybody wishing to implement decision and classification processes should use LVQ (Learning Vector Quantization) instead of SOM."

Despite this warning SOM is used quite frequently as a network for classification (cf. the book [10] or the bibliography on SOM stored in the ftp archive `cochlea.hut.fi` in the `/pub/ref/` catalog, file `references.bib.Z`). The most interesting aspect of SOM is its ability to visualize the high-dimensional data, although the method does not provide any measures of the quality of this visualization. After a brief presentation of the SOM algorithm we will describe the MDS algorithm and compare these two methods in a few cases. In summary we will comment on the differences observed between the maps and the two-dimensional representations obtained and the prospects to develop reliable visualization and classification methods using a combination of MDS and SOM ideas.

II. SELF-ORGANIZING MAPS

The SOM algorithm allows to perform in an unsupervised manner a visualization of high-dimensional input data, usually in a target space of one, two or three-dimensions. We will assume here a two-dimensional array of nodes. SOM seems to preserve the topography of the input data. This means that if some of the high-dimensional points are grouped in clusters, then their representations in the map are also grouped in clusters and the relative distances between clusters are to some degree preserved. The SOM network takes as input a set of labeled sample vectors and gives as output an array of network nodes with the input vector labels attached to these nodes.

Let N be the dimension of the n sample vectors $X(t) \in \Re^N, t = 1, 2, ..., n$, where each sample vector $X(t)$ is identified by a label. The two-dimensional output layer contains $i = 1, ..., xdim \times ydim$ nodes W_i, each serving as a codebook vector of dimension N. The training of the weight (codebook) vectors of the map's nodes is realized by the following algorithm:

For a given number of iterations do:

1. Pick up randomly one sample vector $X(t)$
2. Find the nearest weight vector W_c: $\|X - W_c\| = \min_j \{\|X - W_j\|\}$
3. Update the weights W_i according to the rule:

$$W_i(t+1) = W_i(t) + h_{ci}(t) \cdot [X(t) - W_i(t)]$$

where $h_{ci}(t)$ is the neighborhood function that can be of type:

- "bubble": $h_{ci}(t) = \alpha(t)$ if $\|W_c - W_i\| \leq r(t)$ and $h_{ci}(t) = 0$ if $\|W_c - W_i\| > r(t)$;
- "gaussian": $h_{ci}(t) = \alpha(t) \cdot \exp\left(\frac{-\|W_c - W_i\|}{2\sigma^2(t)}\right)$

Only the neurons within the neighborhood $h_{ci}(t)$ are moved near to $X(t)$. The learning rate $\alpha(t) \in [0, 1]$ decreases monotonically with time, $\sigma(t)$ and $r(t)$ are neighborhood radiuses decreasing also monotonically. Although one-dimensional Kohonen maps have been analyzed in some details little is known about the self-organization process in two or three dimensions [9]. The main problem is the lack of quantitative measure to determine what exactly "the good map" is.

III. MULTIDIMENSIONAL SCALING

MDS techniques emerged from the need to visualize in a two- or three-dimensional space high dimensional objects described by some measure of their similarities or dissimilarities. The problem is to find the coordinates of points representing the multivariate items in the two or three-dimensional space in such a manner that the low-dimensional interpoint distances correspond to the dissimilarities of the original objects. MDS takes as input a symmetric matrix of the similarities or dissimilarities between objects, whereas SOM needs absolute coordinates of these objects in the high-dimensional space. Note that for the MDS input space does not even need to be a metric space. If a given observation concerns n objects there are $n(n-1)/2$ distances between these objects. SOM algorithm in the same case needs $n \times N$ input values, where N is the dimension of the input vectors. If the number of objects $n > 2N + 1$ Kohonen map uses more information than MDS.

Let n be the number of observed objects in the high-dimensional input space $X_1, X_2, ..., X_n$, and let δ_{ij} be the observed similarities between objects X_i, equivalent to distances $\delta_{ij} = \|X_i - X_j\|$ in metric spaces. Let Y_i be the low dimensional target space point representing the input object X_i and let d_{ij} be the distance between Y_i and Y_j. We have to place the points $\{Y_i, i = 1, ..., n\}$ in the target space in such a way that the distances d_{ij} are as close as possible to the original distances δ_{ij}. A sum-of-squared error function can be used as a criterion to decide whether a given configuration of image points is better than another. There are two commonly used criterion:

- Kruskal's `stress`: $S = \sqrt{\frac{\sum_{i>j}(\delta_{ij} - d_{ij})^2}{\sum_{i>j}\delta_{ij}^2}}$

- Lingoes' `alienation coefficient`: $K = \sqrt{\frac{1-\sum_{i>j}(\delta_{ij} \cdot d_{ij})^2}{\sum d_{ij}^2}}$

The best configuration is found iteratively:

0. Define a starting configuration for the points Y_i randomly or by a principal components analysis,

While the criterion function significantly decreases, do:

1. Compute the distances d_{ij}.
2. Compute the value of the criterion functions S and K.
3. Find a new configuration of the points Y_i by a gradient-descent procedure such as Kruskal's linear regression or Guttman's rank-image permutation.

Looking for quantitative measures of the preservation of topography between the high-dimensional input and low-dimensional target spaces Duch [11] has introduced the stress-like measure $D_1 = S$ and its quadratic version:

$$D_2(Y;\delta) = \sum_{i>j}^{n} \left(\delta_{ij}^2 - \sum_{l=1}^{k} \left(y_i^{(l)} - y_j^{(l)} \right)^2 \right)^2$$

where $y_i^{(l)}$ are components of Y_i objects in the k-dimensional target space and the reduction in the number of the degrees of freedom going from N dimensions to k dimensions is taken into account by setting all components of $Y_0 = 0$ and $k-1$ components of Y_1 to zero, $y_1^{(l)} = 0, l = 1..k-1$. For this measure we may obtain the best representation by solving a set of non-linear equations [11] instead of minimization:

$$\sum_{j\neq i}^{n} \left(y_i^{(m)} - y_j^{(m)} \right)^3 + \sum_{j\neq i}^{n} \left(y_i^{(m)} - y_j^{(m)} \right) \sum_{l\neq m}^{k} \left(y_i^{(l)} - y_j^{(l)} \right)^2 - \sum_{j\neq i}^{n} \delta_{ij}^2 \left(y_i^{(m)} - y_j^{(m)} \right) = 0$$

Unfortunately it is as hard to solve this system of nonlinear equations as it is to minimize the stress function.

IV. COMPUTATIONAL RESULTS

Minimization in MDS is usually done via gradient procedure. Since we are looking for a global minimum we have used simulated annealing method for minimization. We have applied SOM and MDS algorithms to a number of cases in which the quality of maps could be assessed easily. Due to the lack of space we will present only two cases: the data related to semantic maps about animals and a series of hypercubes in 3-5 dimensions, with cube corners represented in two-dimensional target space. Configurations of points obtained from SOM and MDS are compared in figures below.

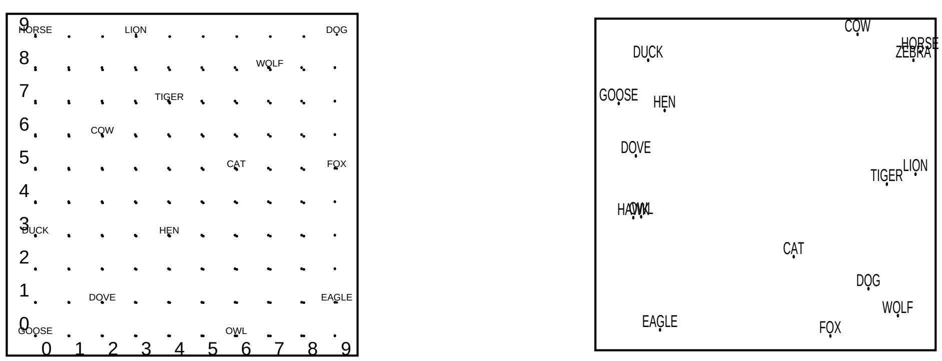

FIG. 1. The two-dimensional representations of the 13-dimensional semantic data obtained by SOM (left) with a 10 x 10 neurons map, a training of 10000 cycles, with final stress of 0.25, and the MDS (right) with final stress of 0.20 after 10 iterations.

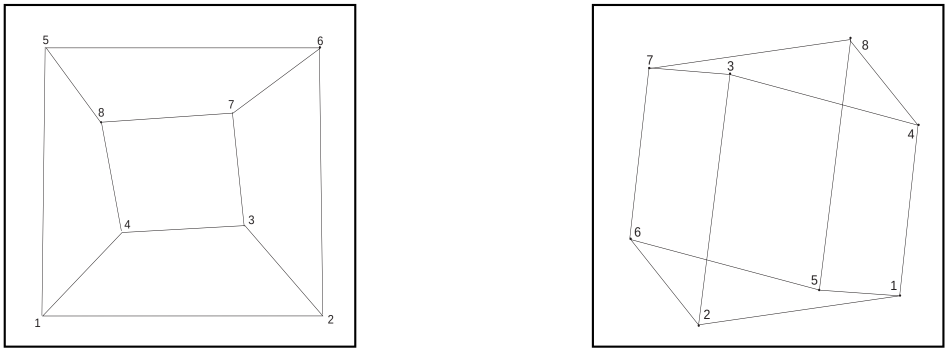

FIG. 2. The two-dimensional representations of the 8 points of the 3D cube obtained by SOM (left) with a 20 x 20 neuron map, a training of 10000 cycles, a final quantization error of 0.001 and stress 0.321, MDS (right) has the final stress value 0.246 after 22 iterations.

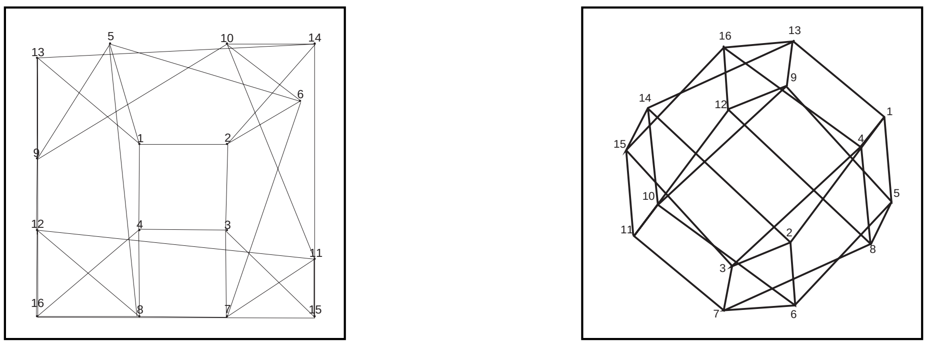

FIG. 3. The two-dimensional representations of the 16 points of the 4D hypercube obtained by SOM (left) with a 20 x 20 neurons map, a training of 10000 cycles, quantization error of 0.001, stress value 0.327, MDS (right) has the final stress of 0.312 after 18 iterations.

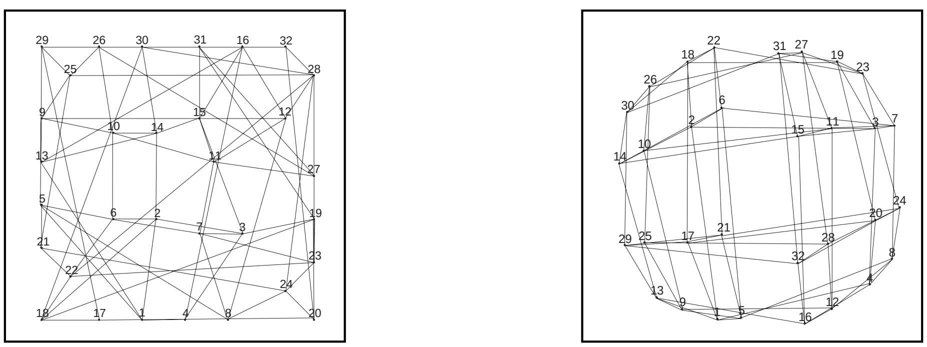

FIG. 4. The two-dimensional representations of the 32 points of the 5D hypercube obtained by SOM (left) with a 20 x 20 neurons map, a training of 10000 cycles, the stress value of 0.353 and by MDS (right) with a final stress of 0.333 after 18 iterations.

In figures 2-4 all corners of the hypercube that are adjacent to each other are connected by lines. All these lines should be short but SOM tries to use all neurons and places many codebook vectors at the boarders, missing the best configuration even in the 3D case. Perhaps with infinitely slow learning (i.e. decreasing the neighborhood function $h_{ci}(t)$ to zero infinitely slowly) it would be possible to avoid "freezing" wrong configurations - 10000 iterations were not sufficient to obtain global minimum of the stress function.

V. SUMMARY AND CONCLUSIONS

Theoretical considerations as well as computational experience shows that MDS is a better approach for visualization of multidimensional data than SOM. Multidimensional scaling provides a well defined measure of the quality of maps, a measure that may also be used to compare the quality of different Kohonen maps. Computational demands of the two methods in the learning phase are similar.

It is possible to modify the original Kohonen SOM algorithm to take into account minimization of stress measure. MDS may also be useful in initialization of large SOM maps used for classification since minimization of stress adding one additional object is more computationally demanding than calculation of output from the Kohonen layer. Introducing the coordinate mesh in the MDS target space and the codebook vectors in each node of this mesh one obtains a combination of the LVQ and MDS methods useful not only in visualization and classification but also in approximation problems (W. Duch and A. Naud, work in progress)

Acknowledgments Support by the Polish Committee for Scientific Research, grant 8T11F 00308, is gratefully acknowledged.

---

[1] R.O. Duda, P.E. Hart, Pattern classification and scene analysis. (Menlo Park, Cal.: J. Wiley 1973)

[2] W.S. Torgerson, Multidimensional scaling. I. Theory and method. Psychometrika, 17 (1952) 401-419

[3] C.H. Coombs, A theory of data. (New York: J. Wiley 1964)

[4] J.B. Kruskal, Nonmetric multidimensional scaling: a numerical method. Psychometrika 29 (1964) 115-129.

[5] J.C. Lingoes, E.E. Roskam, I. Borg (Eds.), Geometric representation of relational data. (Ann Arbor: Mathesis Press 1979).

[6] R.N. Shepard, The analysis of proximities: multidimensional scaling with an unknown distance function. Psychometrika 27 (1962) 125-140.

[7] J.W. Sammon Jr., A nonlinear mapping for data structure analysis. IEEE Transactions on Computers C-18 (1969) 401-409.

[8] T. Kohonen, The self-organizing map. Proceedings of the IEEE 78 (1990) 1464-1480.

[9] T. Kohonen, Self-organizing maps. (Heidelberg Berlin, Springer-Verlag 1995).

[10] Zupan Jure, Gasteiger Johann, Neural networks for chemists: an introduction (VCH, Weinheim, Germany 1993)

[11] W. Duch, Quantitative measures for the self-organized topographical mapping. Open Systems and Information Dynamics 2 (1995) 295-302