Two-level self-organizing maps for analysis of financial statements

Kimmo Kiviluoto, Pentti Bergius

Abstract— **We propose a novel tool based on a hierarchy of two self-organizing maps (SOM's) for analyzing financial statements. The inputs to the first-level SOM are financial indicators derived from a company's annual financial statements; these determine the company's position on the first-level SOM each year. The inputs to the second-level SOM are the coordinates of the company on the first-level SOM during two or more consecutive years. The second-level SOM turns out to give a more accurate description of the state of the company than the first-level SOM; moreover, it is easy to interpret, as each point on the second-level SOM corresponds to a trajectory on the first-level SOM. With our method, several different patterns of corporate behavior can be recognized.**

Keywords— **Self-organizing map, financial statement analysis**

I. Introduction

THE Self-Organizing Map (SOM) [1] has, among its other applications, been used for the analysis of financial statements [2], [3], [4], [5], [6]. It can be used for both visualizing the financial data, and for classifying the companies into healthy and bankruptcy-prone ones.

The studies referenced above are based on financial statements either from a single year or from two consecutive years. However, the practice that has long been preferred by the analysts of Kera Ltd., a Finnish financing company, is to use data from several consecutive years – it has been found that single year data is simply not enough to give a reliable idea of the state of an enterprise. A straigthforward application of this philosophy would be to concatenate financial ratios from several years to a single input vector, which then would be used for SOM training. The problem with this approach is that the map thus obtained is difficult to interpret: there are no simple explications for the different areas of the map.

The solution that we propose in this paper is to proceed in two phases. The annual state of the enterprise is first described with the SOM, so that the state is encoded as a position on the SOM plane. Then, the first-level SOM coordinates for several consecutive years are concatenated, and a second-level SOM is trained with input vectors thus obtained. Now each point on the second-level SOM corresponds to a *trajectory* on the first-level SOM, while the first-level SOM is well-suited for further analysis, as its "coordinate axes" turn out to have very natural interpretations.

Kiviluoto is currently with the Laboratory of Computer and Information Science, Helsinki University of Technology, Espoo, Finland; E-mail: Kimmo.Kiviluoto@hut.fi. Pentti Bergius is with Kera Ltd., Helsinki, Finland; e-mail: Pentti.Bergius@kera.fi

II. Self-Organizing Map as a tool for financial statement analysis

In the present application we are mostly concerned about qualitative analysis of financial statements. The analysis is based on the Self-Organizing Map (SOM), which has a proven record of applications in visualizing high-dimensional data; for a comprehensive treatment of the SOM, see [1]

A. "Semi-supervised" SOM

A straightforward solution for visualizing a set of data is to label each map unit with the data vectors that are mapped to it. This method is practical for small datasets only, however – much smaller than the dataset in the present application. The problem is then how to display the relevant information in such a form that it is easy to see how different attributes, such as the proportion of bankrupt firms, vary in different parts of the map.

The approach we employed here is to train the SOM in a "semi-supervised" manner, used e.g. in [7]. Only the information that can be found from the financial statements is used for determining the shape of the map, other attributes of interest are just carried along with the weight vectors so that they can be later used for visualization. Specifically, denoting the set of financial indicators with vector $\mathbf{x}^{(f)}$ and other attributes with $\mathbf{x}^{(a)}$, the weight vector $\mathbf{m}_j$ associated with each map unit correspondingly consists of two parts: $\mathbf{m}_j = [\mathbf{m}_j^{(f)T} \ \mathbf{m}_j^{(a)T}]^T$. The best-matching unit is then found with the rule

$$c = \underset{j}{\operatorname{argmin}} ||\mathbf{x}^{(f)} - \mathbf{m}_j^{(f)}|| \tag{1}$$

but the weight vectors are updated using the rule

$$\mathbf{m}_j := \mathbf{m}_j + \alpha(t)h(j,c)(\mathbf{x} - \mathbf{m}_j), \qquad \forall j \tag{2}$$

where $\mathbf{x} = [\mathbf{x}^{(f)T} \ \mathbf{x}^{(a)T}]^T$; $\alpha(t)$ is the learning rate, and $h(j,c)$ is the neighborhood function which here has the form of a Gaussian – we experimented also with other neighborhood functions, such as the "bubble" function [1], but Gaussian neighborhood seemed to yield a smoother mapping which is easier to inspect visually. Then, the relative values of map unit components are easily visualized as grey-level pictures for each component plane.

The semi-supervised SOM turned out to perform considerably better than the alternative visualization method, in which the attributes of each map unit are simply averages of the attributes of data vectors mapped to that particular map unit. The reason is that the smoothing by the neighborhood function is able to filter out some of the noise;

this was demonstrated also in [5], where companies were classified into healthy and bankrupt-prone on the basis of their financial statements.

B. Trajectory maps

The long-term behavior of a company is visualized using two SOMs in a hierarchy. The first-level SOM is trained with yearly financial statements, so that for a given year, a company can be positioned on the first-level SOM based on its financial statement for that year. The second-level SOM is then trained with the company's coordinates on the first-level SOM during two or three consecutive years, as illustrated in figure 1. This way, each unit on the second-level SOM corresponds to a *trajectory* on the first-level SOM, capturing one typical pattern of change in the financial statements from year to year.

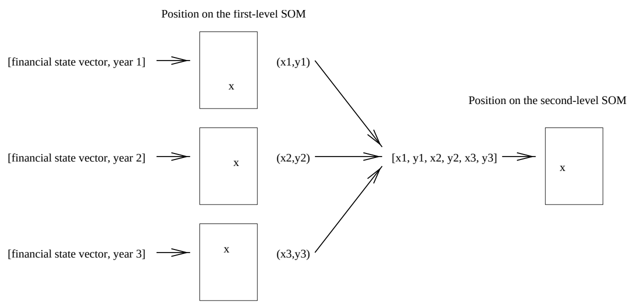

Fig. 1. The second-level SOM is trained with vectors consisting of an enterprise's positions on the first-level SOM during two or three consecutive years. (Figure from [6])

Let us note that this method depends crucially on how well the data can be described using only the first-level SOM coordinates. If the intrinsic dimension of the first-level SOM training data is higher than the dimension of the SOM, the map tries to adapt to the data by folding itself – a phenomenon analyzed in detail in [8]. This folding gives rise to discontinuities in the mapping from the input space onto the map, which can eventually make the first-level SOM coordinates entirely useless for describing the original dynamics in the input space.

Therefore, the intrinsic dimension of the input space should first be somehow assessed. If the sample size permits, one could try to directly estimate the intrinsic dimension using e.g. the algorithm of Pineda and Sommerer [9] or some other method described in the same volume. An alternative is to visually inspect the shape of the trained SOM looking for possible folds, which can be done, for instance, using the Sammon mapping [10] or the "Curvilinear Component Analysis" by Demartines and Herault [11]. In this study, visual checking with Sammon mapping was used.

C. Data preprocessing

Before the financial indicators derived from financial statements were used for training the SOM, they were preprocessed. Our choice for the pre-processing technique was histogram equalization performed separately for each indicator; this method seemed to suit SOM slightly better than other candidates we experimented with, such as variance normalization. In effect, histogram equalization transforms the original highly kurtotic componentwise distributions to nearly uniform distributions.

III. Material

The material used in the present study consists of small and medium-sized Finnish enterprises. The sample was selected from a collection of partial histories of Finnish SME's on the basis of the line of business and size. It was also required that the history and state of the enterprise was known well enough: if there was no data available for a longer period than two years before the bankruptcy, or if the last known financial statements of a supposedly non-failed enterprise were very poor, the company was rejected from the sample.

In the final sample, there were 11 072 financial statements. These were given by 2 579 companies, of which 756 eventually failed, so there were 2 606 financial statements that were given at most five years before failure.

For the financial indicators that were used to train the first-level SOM we chose three commonly used ratios that measure the profitability and solidity of an enterprise.

IV. Results

The first-level SOM is shown in figure 2. From the financial indicator planes it can be seen that the map coordinates correspond roughly to the solidity and the profitability of the company: solidity increases from top to bottom, profitability from left to right. The bankruptcy indicator planes show how the companies drift upwards and to the left as they approach bankruptcy. Earlier, increased failure risk is mostly associated with low solidity, but later on also with decreased profitability.

On figure 3, a few examples of company trajectories on the first-level SOM are shown. The trajectories generally tend to rotate clockwise: a decrease in profitability – a leftward movement – normally results in a decrease in solidity as well, which produces an upward movement. Exceptions to this rule indicate abnormalities, such as changes in the capital structure of the company.

A revealing way to look at the three-year trajectory map is displayed in figure 4. Here, the first-level SOM trajectories that correspond to the selected units of the trajectory map are plotted on top of those units. The trajectories smoothly change throughout the map; the change would be even smoother, if space would permit plotting all the units.

A closer analysis of the second-level maps reveals that they do capture information that escapes the first-level SOM. For instance, on the first-level SOM, the failing companies often jump out temporarily from the high-risk region. On the second-level map, however, there appear to be certain "absorbing states" or areas that failing companies generally do not leave. A similar statement holds true also for a group of very well performing companies. Because of this property, the trajectory maps seem to be a promising tool for rating of enterprises.

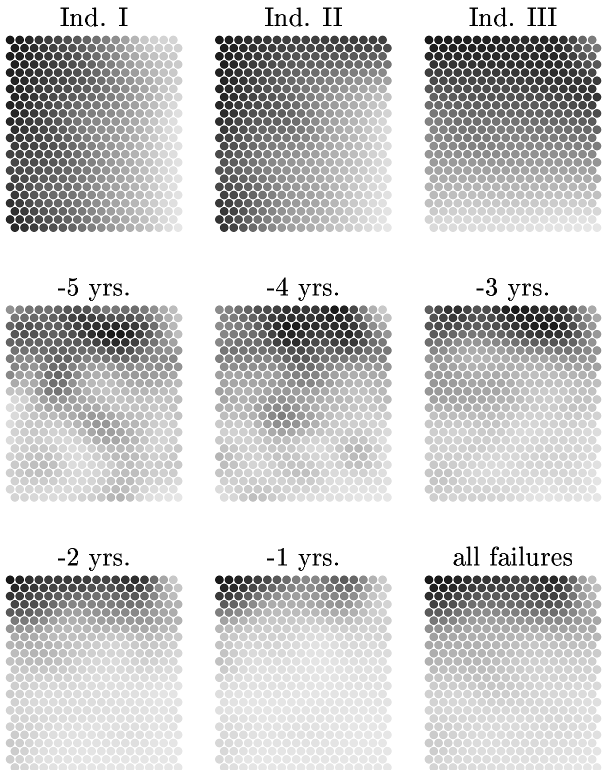

Fig. 2. The first-level SOM. Light color corresponds to good relative values of financial indicators, or low proportion of financial statements given by failed companies. (Figure from [6])

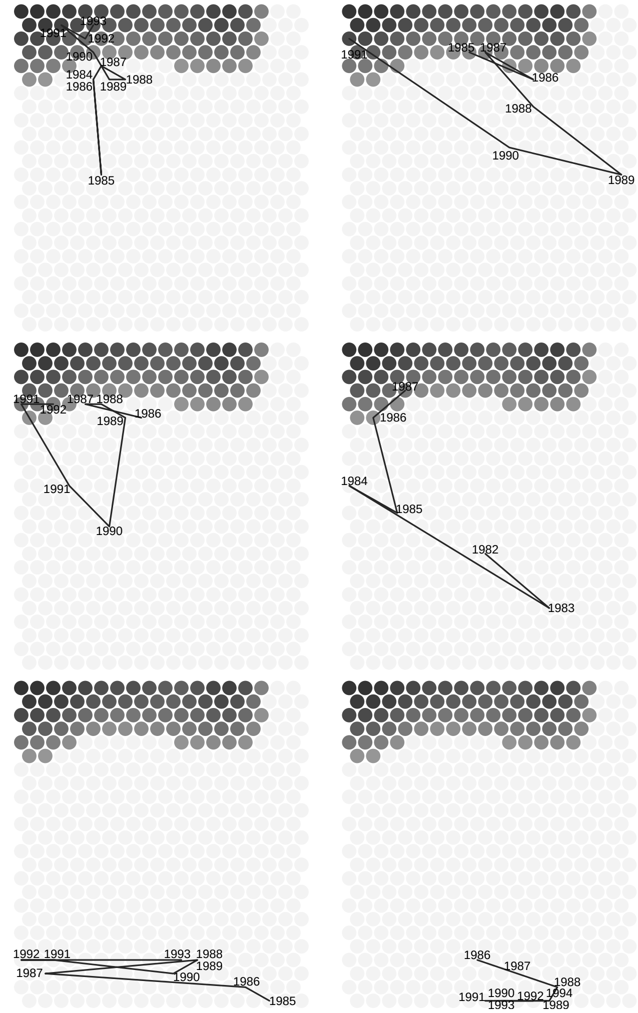

Fig. 3. Trajectories of six companies on the first-level SOM; the four upmost trajectories were drawn by companies that eventually failed. The year the enterprise was mapped to each trajectory point is plotted next to the trajectory; the area with high failure risk is marked with a (thresholded) darker shade. (Figure from [6])

V. Discussion

The Self-Organizing Map offers valuable new insights to the analysis of financial statements. With the two-level SOM, it is possible to recognize different patterns of corporate behaviour and find attributes associated with those patterns. This makes our method look like a promising tool for a more general analysis of financial statements. In particular, applying the two-level SOM for corporate self-benchmarking and corporate rating seems feasible. Also increased risk of corporate failures can be detected with our method.

References

[1] T. Kohonen, *Self-Organizing Maps*. Springer Series in Information Sciences 30, Berlin Heidelberg New York: Springer, 1995.

[2] B. Martín-del-Brío and C. Serrano-Cinca, "Self-organizing neural networks for the analysis and representation of data: Some financial cases," *Neural Computing & Applications*, vol. 1, pp. 193–206, 1993.

[3] B. Back, G. Oosterom, K. Sere, and M. van Wezel, "A comparative study of neural networks in bankruptcy prediction," in *Multiple Paradigms for Artificial Intelligence (SteP94)*, Finnish Artificial Intelligence Society, 1994.

[4] S. A. Shumsky and A. Yarovoy, "Neural network analysis of Russian banks," in *Proceedings of the workshop on self-organizing maps (WSOM'97)*, (Espoo, Finland), Neural Networks Research Centre, Helsinki University of Technology, June 1997.

[5] K. Kiviluoto and P. Bergius, "Analyzing financial statements with the self-organizing map," in *Proceedings of the workshop on self-organizing maps (WSOM'97)*, (Espoo, Finland), pp. 362–367, Neural Networks Research Centre, Helsinki University of Technology, June 1997.

[6] P. Bergius and K. Kiviluoto, ch. "Maps for analysing failures of small and medium-sized enterprises" in *Visual Explorations in Finance using Self-Organising Maps* (eds. G. Deboeck and T. Kohonen). Springer London, 1998. In press. (Book and chapter titles provisional.).

[7] J. Heikkonen, P. Koikkalainen, and E. Oja, "Self-organizing maps for collision-free navigation," in *World Congress on Neural Networks*, vol. III of *International Neural Network Society Annual Meeting*, pp. 141–144, Neural Network Society, July 1993.

[8] H. Ritter and K. Schulten, "Convergence properties of Kohonen's topology conserving maps: Fluctuations, stability, and dimension selection," *Biological Cybernetics*, vol. 60, pp. 59–71, Nov. 1988.

[9] F. J. Pineda and J. C. Sommerer, ch. "Estimating generalized dimensions and choosing time delays: a fast algorithm" in *Time series prediction: forecasting the future and understanding the past* (eds. A. S. Weigend and N. A. Gershenfeld), vol. XV of *Santa Fe Institute studies in the sciences of complexity*, pp. 367–385. Addison-Wesley, 1993.

[10] J. W. S. Jr., "A nonlinear mapping for data structure analysis," *IEEE Transactions on Computers*, vol. 5, pp. 401–9, May 1969.

[11] P. Demartines and J. Hérault, "Curvilinear component analysis: a self-organizing neural network for nonlinear mapping of data sets," *IEEE Transactions on Neural Networks*, vol. 8, pp. 148–154, Jan. 1997.

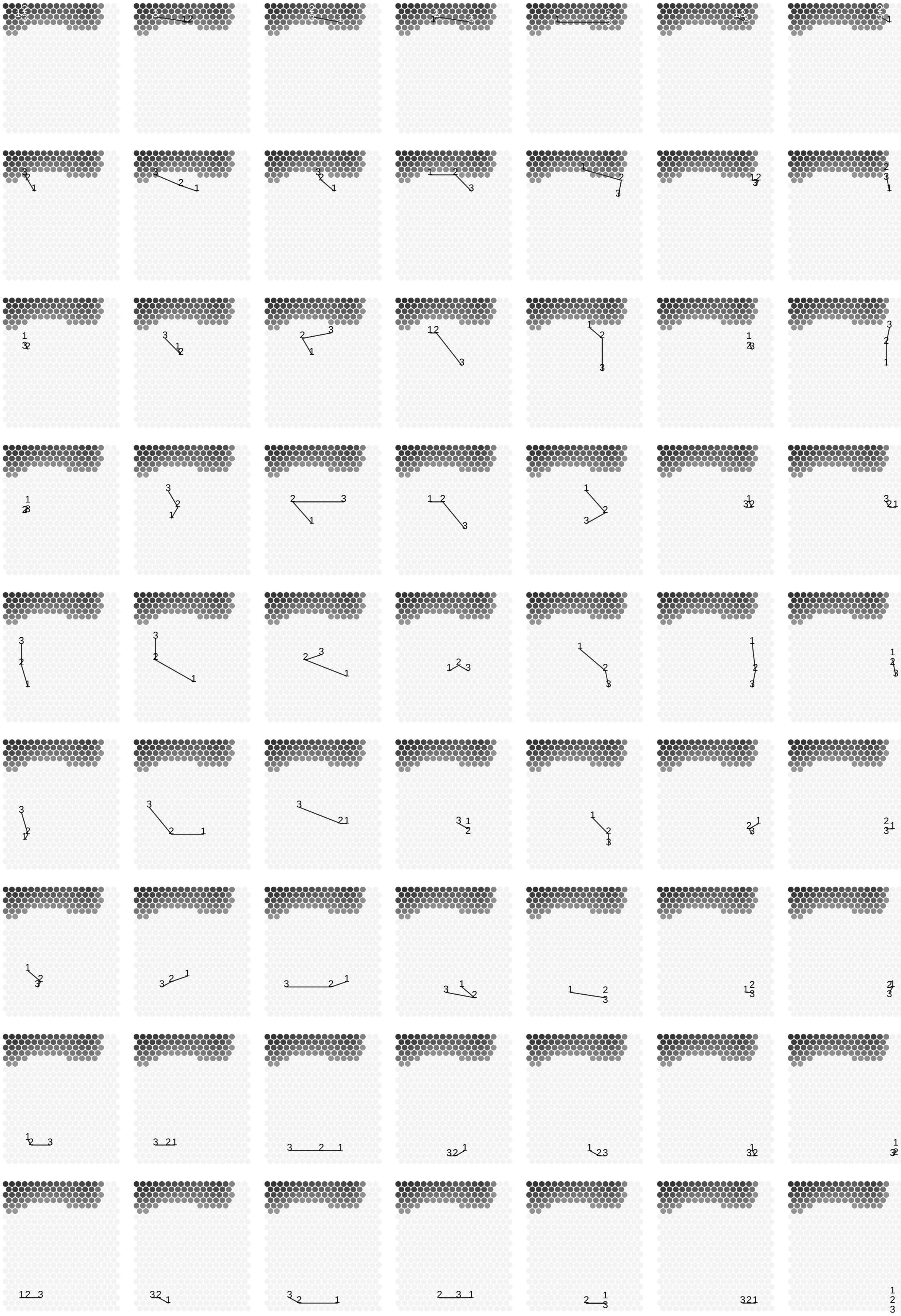

Fig. 4. The three-year trajectory map "opened": On top of map units are plotted the corresponding first-level SOM trajectories; approximately every third unit shown. (Figure from [6])