Neural Network Analysis of Russian Banks

S. A. Shumsky, A. V. Yarovoy

Lebedev Physics Institute
Leninsky prospect, 53, Moscow, Russia

Abstract

The structure of Russian bank system was studied using two-layer self-organizing neural network. The first, Oja's layer, reduces the initially high dimensionality of banks balances. The second, two-dimensional Kohonen layer, provides the final visualization tool for data analysis. Various coloring of the Kohonen map comprise the "Kohonen atlas" of Russian banks.

1 Introduction.

Russian bank system comprises more than 2000 commercial banks. Each of them is characterized by numerous balance items. For example, rather reduced and superficial description of a bank, so-called first-order balance consists of 32 balance items. More specific second-order balance may have hundreds of items.

Advanced modern methods of nonlinear data processing provide the powerful tool for proper understanding of data regularities and for good visual representation of such multidimensional data. This allows one to extract hidden order from data bases without using any *apriori* information about object in question.

For better understanding human has to operate low-dimensional information in the form of charts and pictures. However, the inevitable result of such dimensionality reduction is the loss of information. The task is — how to reduce dimension with minimal loss of relevant information. Here we perform this task using two different methods based on self-organizing neural networks — the *Oja algorithm* and the *Kohonen map*.

2 Data Description

The data under consideration comprise the first-order balances (32 items each) of 1800 Russian commercial banks. All data belong to 1994 year accounts.

One of important features of Russian bank system is a large "dynamic range" of the bank sizes: the biggest bank is more than 15,000 times larger than the smallest one. However, we sought to characterize the banks regardless of their sizes — to be able to compare various bank strategies, i.e. the shares of money on different accounts. To this end we normalized all balance items on the total amount of assets. This enabled us to compare financial strategies of small and large banks.

After excluding two balance items (total amount of assets and total amount of liabilities that must be equal to the first one) one finally obtains $n = 30$ normalized balance items.

3 Principal Component Analysis by Oja Network

Traditional dimensionality reduction algorithm - *Principal Components Analysis (PCA)* - implies solving the eigenvalue problem for covariance matrix. However, due to implicit linear dependencies among the balance items this problem is an ill-posed one. The neural approach [2] allows one to

avoid solving such an ill-posed problem. To reduce data dimensionality we used a network with $m < 30$ neurons with the following weights updating rule:

$$\Delta \mathbf{w}_k = \alpha\, y_k\, (\mathbf{x} - \sum_{i \le m} y_i \mathbf{w}_i) \tag{1}$$

where $\mathbf{x}$ is the current data point, $\mathbf{w}_k$ is a weight vector of k-th neuron and $y_k = \mathbf{w}_k \mathbf{x}$ is an output of k-th neuron. Such training forces the neurons weights to converge to linear combinations of the first m principal components. Figure 1 shows the fraction of total variance

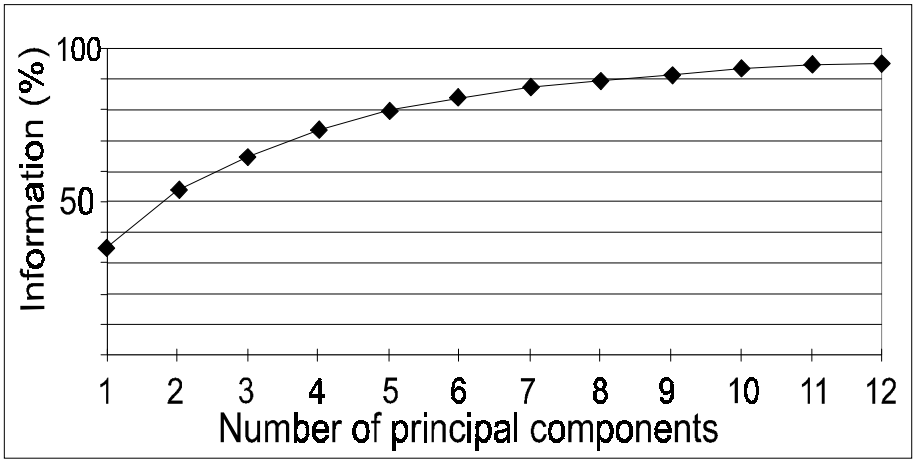

Figure 1: Information versus Number of Principal Components

$$I = \frac{\sum_{i=1}^{m} \sigma_i}{\sum_{i=1}^{n} \hat{\sigma}_i},$$

in the first m principal components roughly corresponding to their information content. Here σ_i is the variance of the i-th output: $\sigma_i = \left\langle (y_i - \langle y_i \rangle)^2 \right\rangle$ and $\hat{\sigma}_i$ is the variance of the i-th input: $\hat{\sigma}_i = \left\langle (x_i - \langle x_i \rangle)^2 \right\rangle$.

One can see that the distribution of information over the principal components is highly nonuniform. The first ten principal components possess 94 per cents of information while the last twenty — only the rest 6 per cents. This is due to a great information redundance of banks balances. Such a redundance allows one to limit consideration to the first several principal components. In the further analysis we used 10 of them.

Though *linear* PCA analysis allowed a threefold reduction of the problem's dimensionality, it is still far from the desired 2-dimensional data representation. The latter was obtained by using *nonlinear* data compression technique - two-dimensional Kohonen map.

4 Kohonen Map of Russian Banks

The Kohonen layer consisted of $20 \times 20 = 400$ neurons on a two-dimensional grid. The inputs of each neuron were 10 outputs of the Oja layer described above (see Figure 2).

In the training phase the weight vectors of the Kohonen neurons were adjusted according to the following rule [1]:

$$\Delta \mathbf{w}_{ik} = \alpha\, (\mathbf{x} - \mathbf{w}_{ik}) \Lambda_{(i-i^*)(k-k^*)} \tag{2}$$

where $\Lambda_{xy} = \exp(-(x^2 + y^2)/2\sigma^2)$, $\mathbf{x}$ is the current data point and (i^*, k^*) — coordinates of the "winning" neuron on Kohonen map. σ and α were gradually decreased during the training until they became sufficiently small.

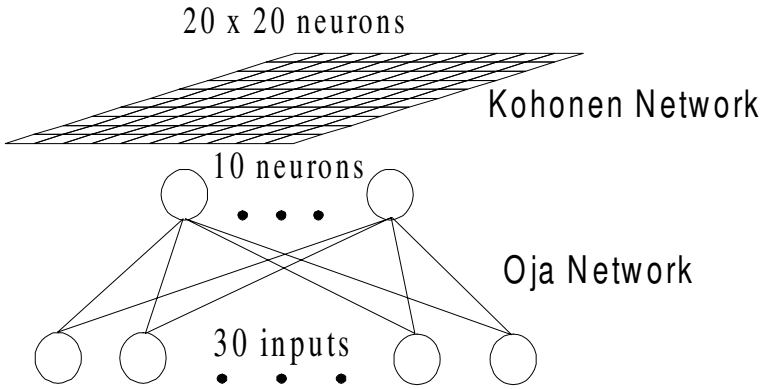

Figure 2: Two-layer self-organizing network for Russian banks analysis.

In the trained network each neuron indicated typically 5 to 6 banks (there were two "empty" neurons, i. e. neurons that were not nearest to at least one data point).

The resulting map itself only provides the "coordinates" for each Russian bank, preserving the local topology of the data. The coloring technique provides further insight in the existing data regularities. E.g. one may mark the cells that contain banks that became bankrupts in the next year. The resulting Figure 3 evidence that the vast majority of bankrupts are concentrated in the right-top corner of the map. This area may thus be considered as a "risk zone".

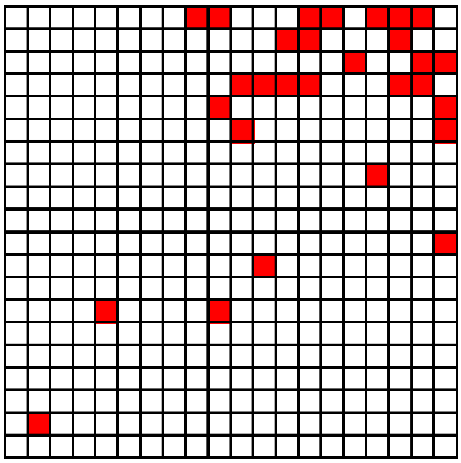

Figure 3: Banks-bankrupts on Kohonen Map

One may then wonder whether the sizes of the banks play some role in choosing their financial strategies. It is worth stressing here that the bank sizes were eliminated from consideration, so small and large banks could well be represented by the same neural cell. The map coloring shown on Figure 4 is caused by the *order* of the bank, i.e. the decimal logarithm of a total amount of assets. The rainbow colors were used to show different sizes of banks. In this palette the bigger the bank the closer the corresponding color to the red end of a spectrum. Two neighboring colors correspond to fivefold difference in size.

According to Figure 4 the banks of different sizes have different financial strategies. For example, very large banks are concentrated at the bottom of the map, while small banks are concentrated in the right bottom corner (group 1) and on the left side of the map (group 2). These two groups

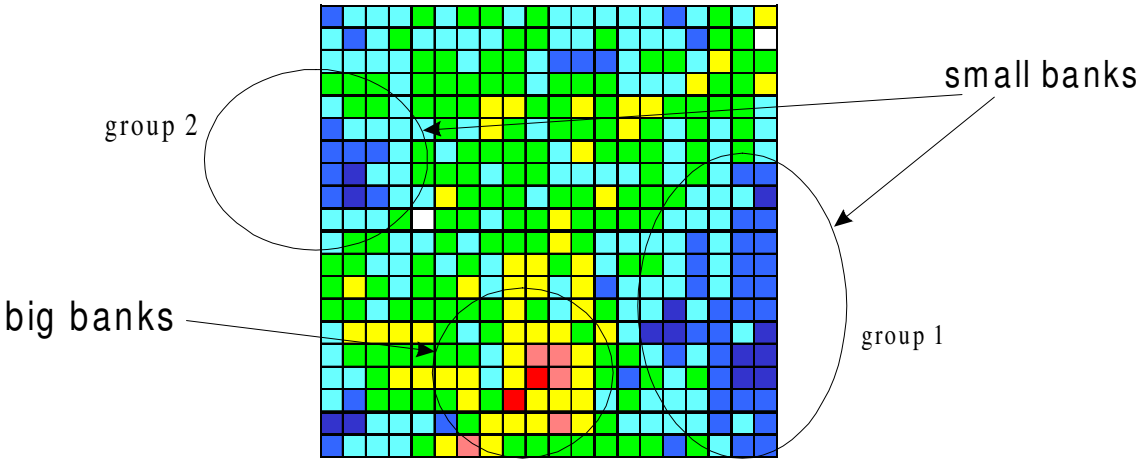

Figure 4: Bank sizes

of small banks appear to adhere very different financial strategies although they have some similar properties (see Figure 5).

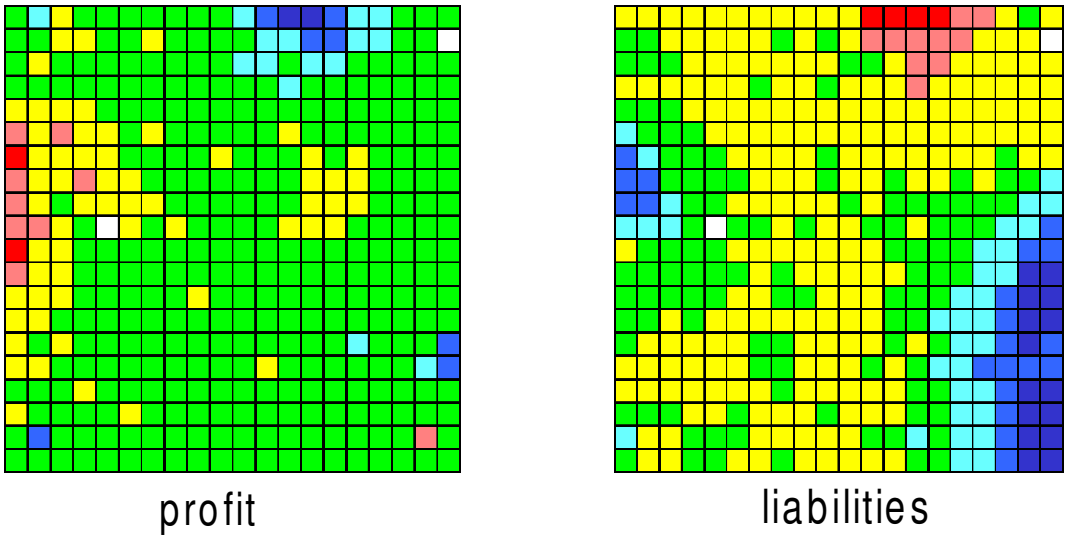

Figure 5: Balance items

Two relevant balance items — "profit" and "liabilities" — are shown on Figure 5. These items help to understand deeper the financial behavior of different groups of Russian banks. The "liabilities" map evidences that both groups of small banks have small amount of liabilities while banks in the risk zone have the biggest amount of liabilities compared to others. These latter banks also demonstrate the smallest profit. Such a dangerous combination explains the large amount of business failures in the risk zone.

Back to two groups of small banks, the very important distinction between them is the item "profit". While group 1 has the profit below average, group 2 has the maximal profit on the map. We should mention that group 1 doesn't contain banks-bankrupts though it has low profit. The analysis of different maps brings one to conjecture that group 1 are "quasi-commercial" banks brought to life by economical changes in modern Russia. Their aim is not getting a commercial profit, but rather to serve as a "cash desk" for their maternal firms and corporations.

5 Conclusion

We performed the analysis of Russian bank system using a two-layer self-organizing neural network. The Oja layer mapped 30-dimensional inputs (balances) to 10-dimensional outputs without significant loss of information. The resulting 10 principal components were fed to the second layer: two-dimensional Kohonen map of Russian banks. Various colorings of such a map produced a "Kohonen atlas" of Russian banks, where different financial groups may easily be found. The above technique provides a powerful tool of high-dimensional financial data visualization and analysis. The results of the studies of the dynamics of Russian banks will be published elsewhere. A relative simplicity of implementation and interpretation extends the applicability of this approach.

Acknowledgments

This work was supported in part by the ISTC project #31-94 and information agency "Prime". We also would like to express our sincere gratitude to Maria Rumyanceva for raw data preparation and her kind assistance during the work.

References

[1] T. Kohonen Self-organized formation of topologically correct feature maps. *Biological Cybernetics,* **43**, 59-69

[2] E. Oja A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology,* **15**, 267-273