ISSN 1392 – 124X INFORMATION TECHNOLOGY AND CONTROL, 2007, Vol.36, No.1A

PREDICTION OF CHANGES OF BANKRUPTCY CLASSES WITH NEURO-DISCRIMINATE MODEL BASED ON THE SELF-ORGANIZING MAPS

Egidijus Merkevičius, Gintautas Garšva

Vilnius university, Kaunas faculty of humanities, Department of Informatics
Muitinės St. 8, LT- 44280 Kaunas, Lithuania

Abstract. It is suggested in this paper generating SOM as one that could be applied for forecasting of bankruptcy classes for other than trained companies. The various aspects of the proposed Neuro-discriminate model based on the multiple discriminate analysis, supervised learning neural network and self-organizing maps are analyzed. There is compared how accuracy of prediction changes executing algorithm with different discriminate models of bankruptcy: Springate, Zmijewski and Shumway.

1. Introduction

Bankruptcy is described as inability or impairment of ability of an individual or organization to pay their creditors, in other words – default. One of the most important tasks of the creditors is to manage the credit risk-making forecast of changes in financial standing.

During the past 15 years investigations in area of SOM applications to financial analysis have been done. Doebeck described and analyzed most cases in [0]. Martin-del-Prio and Serrano-Cinca generated SOM's of Spanish banks and subdivided those banks into two large groups, the configuration of banks allowed establishing root causes of the banking crisis [0].

Based on Kiviluoto's study [0], through visual exploration one can see the distribution of important indicators (i.e. bankruptcy) on the map.

The following authors have estimated only historical and current financial data of the companies and afterwards they have interpreted it for forecasting bankrupt of those companies. In this paper, we suggest generating SOM as one that could be applied for forecasting of bankruptcy classes for other than trained companies.

The methodology of the model is described in the first section. Second part presents short description of proposed Neuro-discriminate model. The experiments with the actual financial data and analysis of results are described in the third section and the last section of the paper makes some conclusions.

2. Methodology

Artificial neural networks (ANN) are divided into supervised and unsupervised learning [0],[0]. The Self-organizing map (SOM) is an unsupervised learning artificial neural network that is generated without defining output values [0]. The outcome of this process is a two-dimensional cluster map that can visually demonstrate the financial units, which are scattered according to similar characteristics.

Multiple discriminant analysis is a statistical technique used to classify an observation into one of several a priori groupings dependent upon the observation's individual characteristics. It is used primarily to classify and/or make predictions in problems, where the dependent variable appears in qualitative form [0],[0].

In this paper we analyze the various aspects of the Neuro-discriminate model (further – Model) presented in [0]. Methods used in the Model are original with no major adjustments, so they are not presented. More detailed description of multiple discriminate analysis, supervised learning neural network and self-organizing maps are presented in [0],[0],[0],[0]. Figure 1 shows the concept of the Model.

A description of the Model concept is as follows [0, 0]:

1) Changes of companies financial standing are determined as changes of the indexes of bankruptcy model during two periods straight (0 – negative changes, 1 – positive changes);
2) The components of discriminate bankruptcy model are used for training of unsupervised neural network and generating SOM. Testing of

accuracy of the SOM is executed via calculation of corresponding nodes between training and testing data.

3) The accuracy of forecasting is improved via changing of weights. Feed-forward neural network (further - FF ANN) is used in the Model as a tool for changing of weights. The main principle is taken from the core of ANN theory – training an ANN the weights in the ANN are adjusted while the ANN gives the same outputs as in the training data. In other words, the goal of the ANN is to get optimal set of weights via changing them. In that way the inputs of FF ANN would be data of testing and the outputs would be the original generated labels of SOM nodes. Initial weights in the ANN are set as weights of original bankruptcy model as described above.

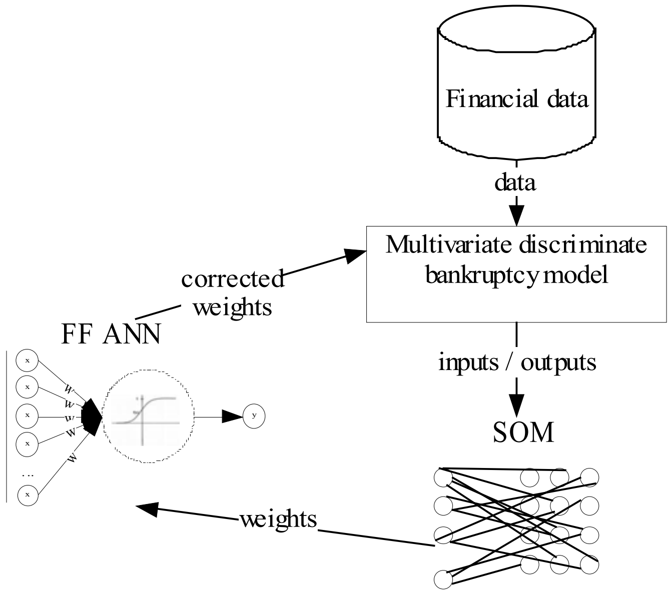


Figure 1. The concept of the model

3. Experiments and analysis

The testing of proposed Model has been executed using real financial dataset: companies from NASDAQ list loaded from EDGAR PRO Online database [0].

The main task of the experiments is to compare how accuracy of prediction changes executing algorithm with different discriminate models of bankruptcy: Springate [0] (further – SPRINGATE), Zmijewski [0] (further – ZMIJEWSKI) and Shumway [0] (further – SHUMWAY). In other words, it is expected, that proposed Model can be used with various financial ratios as inputs independent of they weights. The basic characteristics of the dataset and the experiments are as follows:

- Number of companies: 9364.
- Dataset consists of annual financial statements of 7 periods consecutively. The dataset was rebuilt in the way that the string of record was a pair of financial statement straight. After that count of records seeks 56184 records.
- Risk classes of bankruptcy (or outputs) are determined as follows: if the index of selected bankruptcy model – SPRINGATE, ZMIJEWSKI or SHUMWAY – of the second period is less than index of the first period then the risk class is determined as 0, otherwise – 1.
- Two-level filtering of data was executed: first, count of records was reduced to 46353 records after elimination of missing data, second, count of records was reduced to 36167 records after elimination records, whose outputs (0 – negative changes, 1 – positive changes) don't correspond between all three bankruptcy models.
- Records are divided to the two subsets – the one for the training (TRAINDATA), the second – for the testing (TESTDATA). On the each iteration, the separation of data into the training and testing data has been executed randomly with the ratio 70:30.
- According to bankruptcy models, there are variables (inputs) for each of 2 periods for the training of SOM as follows.

The main principle of the measurement of prediction accuracy is as follows: the labeled nodes of trained SOM are labeled with the outputs of testing data and the corresponding nodes of training and testing data are calculated.

Remarks and conclusions of experiment with ZMIJEWSKI:

- Accuracy of prediction increased from 73.12% to 93.86% after 22 iterations of the cycle. After further iterations of the cycle, differences between changed and original variables do not increase and remains at the same level.

Table 1. Variables and ratios

Variables	Ratios of SPRINGATE	Ratios of ZMIJEWSKI	Ratios of SHUMWAY
p0w1 and *p1w1*	Working capital/Total assets (1.03)	-4.336	-7.811
p0w2 and *p1w2*	EBIT/Total assets (3.07)	Net income/Total assets (-4.513)	Net income/Total assets (-6.307)
p0w3 and *p1w3*	EBT/Short-term liabilities (0.66)	Total liabilities/Total assets (5.679)	Total liabilities/Total assets (4.068)
p0w4 and *p1w4*	Total sales/Total assets (0.4)	Short-term assets/ Short-term liabilities (0.004)	Short-term assets/ Short-term liabilities (-0.158)

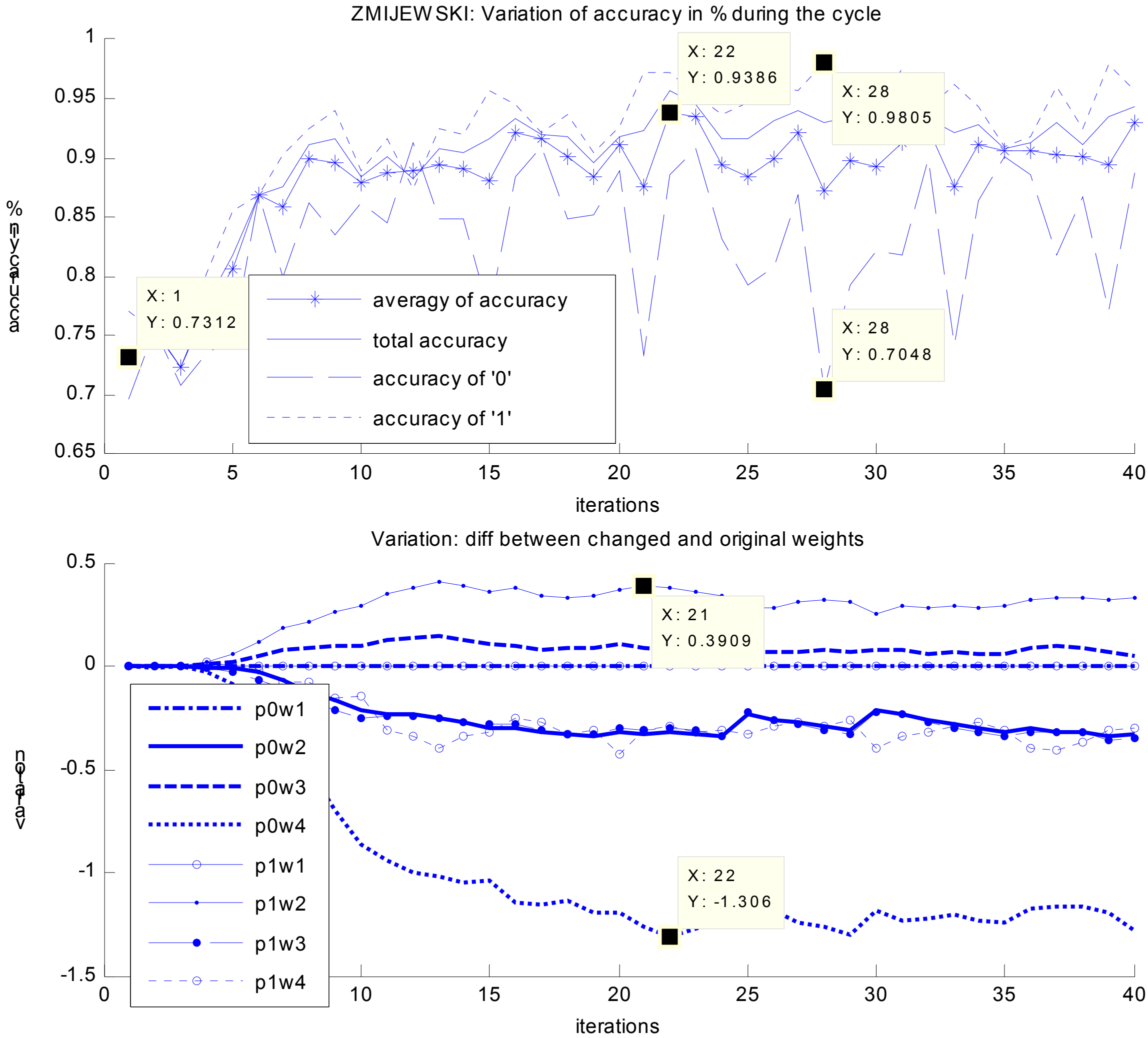


Figure 2. Improving of accuracy: ZMIJEWSKI

- Increase was influenced mainly under changing of variable p1w2 (Net income/Total assets (-4.513)) and p0w4 (Short-term assets/Short-term liabilities (0.004)) as shown in the second graph of the figure 2 – after 22 iteration differences between changed and original weight (p0w4) seeks -1.306.
- Accuracy of ‘1’ increased up to 98.05% after 28 iteration; otherwise, very weak prediction of ‘0’ (70.48%) do not let to determine this structure of SOM as optimal.
- We consider, the optimal structure of the weights was reached after the 21 iterations. Average of accuracy aspects calculates to 93.86%.
- Confusion matrix gives more detailed information about considered optimal structure of SOM in the table 2.

Results of experiments with the SHUMWAY can be seen in the figure 3.

Table 2. Confusion matrix of ZMIJEWSKI

Actual vs Predicted (Confusion matrix)			
	Predicted (by model)		
	0	1	Total (units)
Actual 0 (%)	88.56	11.44	201
Actual 1 (%)	2.71	97.29	923
Total accuracy (%)			95.73
Average of all aspects of accuracy (%)			93.86

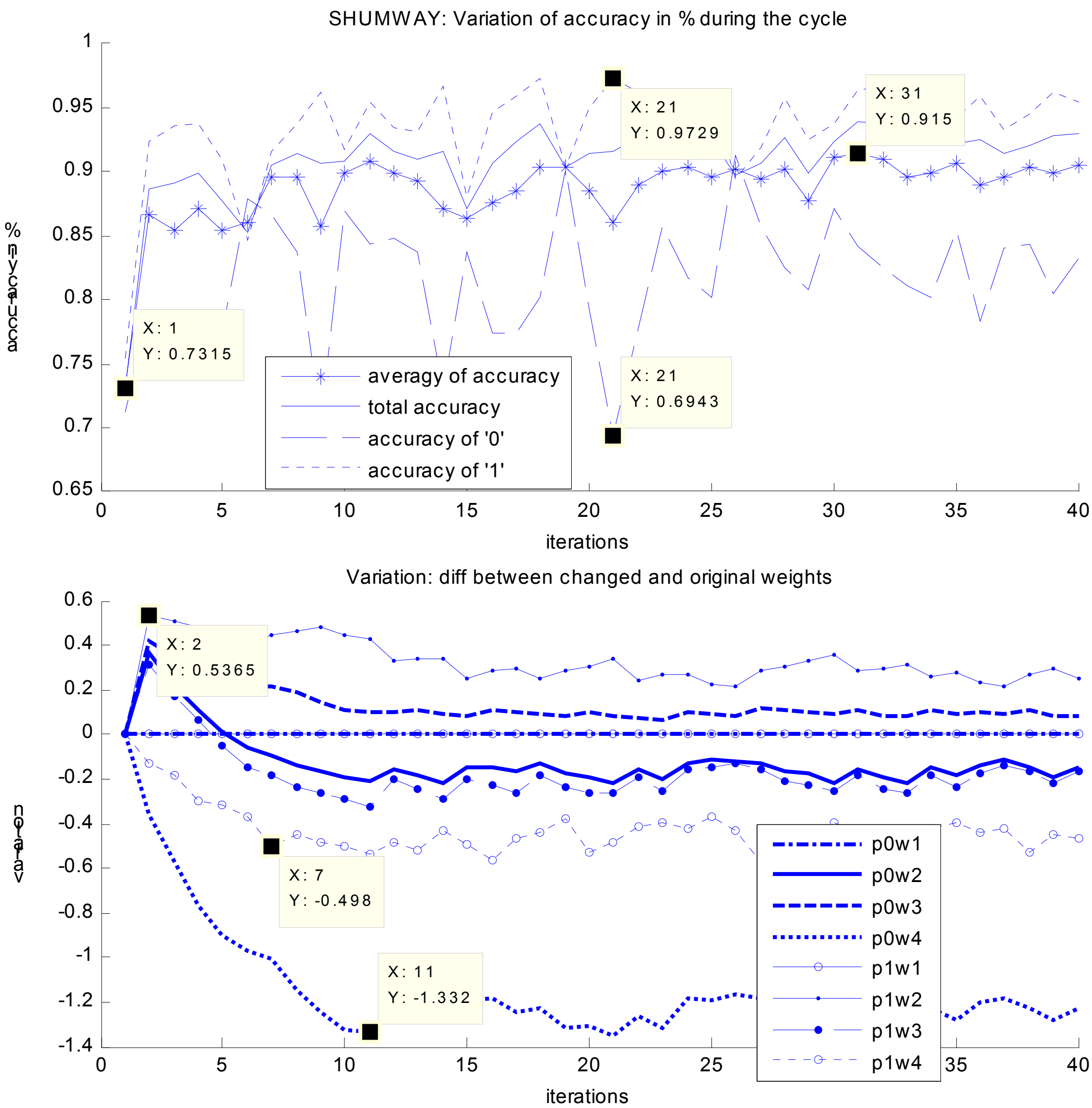


Figure 3. Improving of accuracy: SHUMWAY

Conclusions of experiments with SHUMWAY are very similar to results of ZMIJEWSKI. It is, because both models use the same ratios. The figure 3 presents results as follows:

- Rapidly increase of accuracy from 73.15% to 91.50% after 31 iterations;
- The main influence to the performance of accuracy make the same ratios suchlike in ZMIJEWSKI: p1w2 (Net income/Total assets (-4.513)) and p0w4 (Short-term assets/ Short-term liabilities (0.004)).

Confusion matrix presented in the table 3.

Table 3. Confusion matrix

Actual vs Predicted (Confusion matrix)			
	Predicted (by model)		
	0	1	Total (units)
Actual 0 (%)	84.19	15.81	234
Actual 1 (%)	3.60	96.40	916
Total accuracy (%)			93.91
Average of all aspects of accuracy (%)			91.50

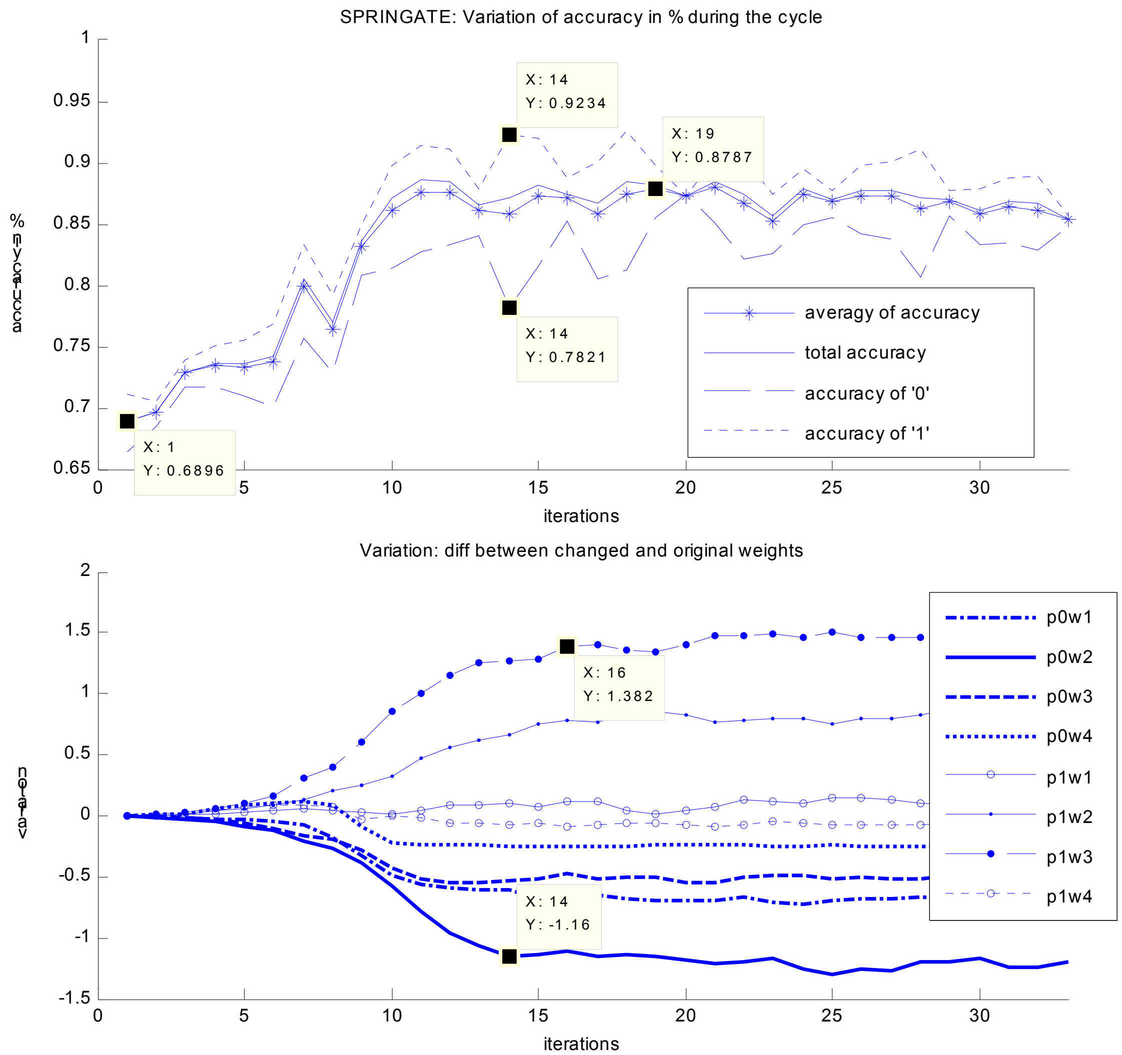


Figure 4. Improving of accuracy: SPRINGATE

Some interesting moments are remarked during analysis of this graph:

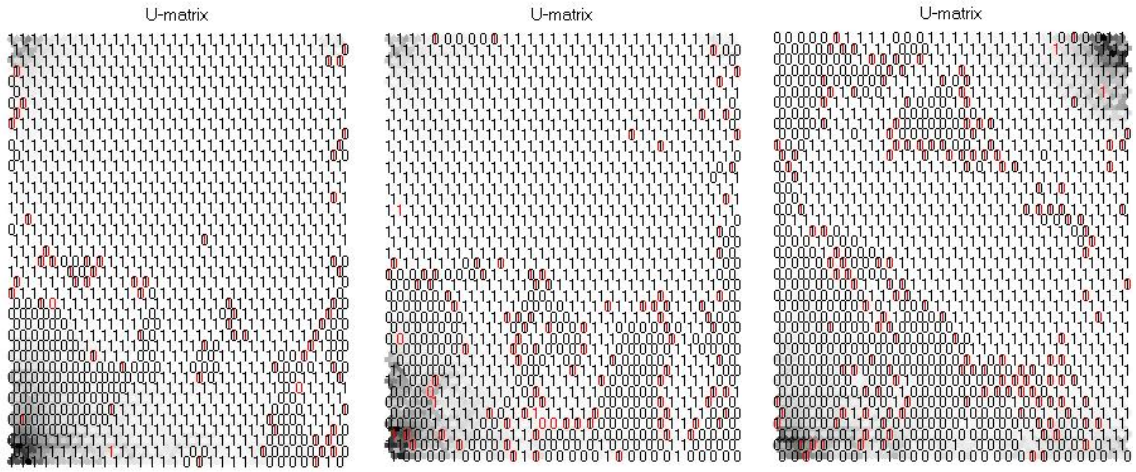
- Accuracy of prediction increased from 68.96% to 87.87% after 15 iterations of the cycle. Increase was influenced mainly under changing of variable p1w3 (EBT/Short-term liabilities (0.66)) as shown in the second graph of the figure 4 – after 6 iteration difference between changed and original weight (p1w3) seeks 1.382. The variable p0w2 (EBIT/Total assets (3.07)) makes also deep influence to the improvement of accuracy (-1.16 after 14 iterations). After further iterations of the cycle, differences between changed and original variables do not increase and remains at the same level. We consider, the optimal structure of the weights was reached after the 15 iterations.

Table 4 presents Confusion matrix of the results.

Figure 5 let us to make a visual comparison of generated SOM`s based on the different bankruptcy models. The weights for the inputs are determined according to the best results of accuracy and established optimal structure of SOM as described above.

Table 4. Confusion matrix

Actual vs Predicted (Confusion matrix)			
	Predicted (by model)		
	0	1	Total (units)
Actual 0 (%)	85.51	14.49	421
Actual 1 (%)	10.16	89.84	728
Total accuracy (%)			88.25
Average of all aspects of accuracy (%)			87.86



ZMIJEWSKI SHUMWAY SPRINGATE

Figure 5. Comparison of SOM`s based on the different bankruptcy models

Un-corresponding nodes are presented in SOM`s as different color and on the top each-other. Several conclusions are remarked via visual exploration of presented U-matrix of SOM`s as follows:

- The most amounts of un-corresponding nodes are observed in the SOM of SPRINGATE.
- The main difference between SOM`s of ZMIJEWSKI and SHUMWAY and SOM of SPRINGATE is the amount of nodes labeled as “0” (negative changes of financial state). The SOM of SPRINGATE is structured as 421 units of “0” and 728 unites of “1”, whereas labels “0” and “1” are distributed in the SOM`s of ZMIJEWSKI and SHUMWAY with the ratio app. 200:900.

4. Conclusions

Comparing experiments described above, it can be make some conclusions:

- The most significant increase of accuracy of prediction is achieved after 5-10 iterations via changing the weights of ratios (app. +20%). The use of feed-forward neural network gives a short way to achieve the good results of prediction. Execute of further steps of the cycle can determine the optimal structure of the SOM (good prediction of both classes).
- Visualization of SOM by U-Matrix style gives good capabilities to explore the distribution of train- and test labels.
- The Model discriminate specifically the ratios that have most important impact to the determination of bankruptcy class.

References

[1] **A.F. Atiya.** Bankruptcy prediction for credit risk using neural networks: a survey and new results. *IEEE Transactions on Neural Networks*, *Vol.*12, *No.*4, 2001, 929-935.

[2] **E. Altman.** Financial Ratios, Discrimination Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance*, 1968, 589-609.

[3] **E. Altman.** Predicting Financial Distress of Companies: Revisiting the Z-Score and ZETA® Models, 2000. *Working paper at http://pages.stern.nyu.edu/~ealtman/Zscores.pdf* .

[4] **G. Deboeck.** Self-Organizing Maps Facilitate Knowledge Discovery in Finance. *Financial Engineering* News, 1998.

[5] EDGAR Online, Inc. 1995-2006. *http://pro.edgar-online.com*.

[6] **K. Kiviluoto.** Predicting bankruptcies with the self-organizing map. *Neurocomputing*, *Vol.*21, 1998, 191–201.

[7] **T. Kohonen.** The Self-Organizing Map. *Proceedings of the IEEE*, 78, 1997, 1464-1480.

[8] **K. Martin-del-Prio, Serrano-Cinca.** Self-Organizing Neural Network: The Financial State of Spanish Companies. *Neural Networks in Finance and Investing. Using Artificial Intelligence to Improve Real-World Performance. R.Trippi, E.Turban, Eds. Probus Publishing*, 1993, 341-357.

[9] **E. Merkevičius, G. Garšva, S. Girdzijauskas.** A Hybrid SOM-Altman Model for Bankruptcy Prediction. *International Conference on Computational Science* (4), *Lecture Notes in Computer Science*, 3994, 2006, ISSN 0302-9743, 364-371.

[10] **E. Merkevičius, G. Garšva, S. Simutis.** Neuro-discriminate Model for the Forecasting of Changes of Companies Financial Standings on the Basis of Self-organizing Maps. *International Conference on Computational Science* (5) *Part* II, *Lecture Notes in Computer Science*, 4488, 2007, 439–446.

[11] **T.Shumway.** Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *Journal of Business*, *Vol.* 74, *No.*1, 2001, 101-124.

[12] **G.L.V. Springate.** Predicting the Possibility of Failure in a Canadian Firm. *Unpublished M.B.A. Research Project, Simon Fraser University, January* 1978. *In: Insolvency Prediction, E. Sands & Associates Inc. http://www.sands-trustee.com/insolart.htm.*

[13] **J. Vesanto, J. Himberg, E. Alhoniemi, J. Parhankangas**. SOM toolbox for Matlab 5. *Technical report* A57, *Helsinki University of Technology, Finland*, 2000.

[14] **M.E. Zmijewski.** Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research* 24 (*Supplement*), 1984, 59-82.

Received March 2007.