Barometr Regionalny
Tom 12 nr 3

Clustering and Visualization of Bankruptcy Patterns Using the Self-Organizing Maps

Andrzej Burda
University of Management and Administration in Zamość, Poland

Krzysztof Pancerz
University of Management and Administration in Zamość, Poland
University of Information Technology and Management in Rzeszów, Poland

Abstract

Pattern recognition of bankrupt or non-bankrupt enterprises may not only extend or confirm the knowledge in economics, but also deliver to experts, from the standpoint of the decision support, a view of the economic and financial situation of the audited enterprise. Therefore, it may be an effective tool for early warning of the bankruptcy risk of the enterprise. Such a tool is especially important for small and medium enterprises (SMEs) in the underdeveloped regions. The research described in the paper is intended for generation and visualization of the state of SMEs in the Podkarpacie region on the basis of information included in financial reports. A self-organizing map (SOM), often called the Kohonen net, has been used in the unsupervised modelling mode. Results of research show a high potential of the method to the stated objectives and the simplicity of the representation of knowledge transferred to entrepreneurs and financial analysts.

Keywords: Self-Organizing Map, clustering, visualization, small and medium enterprises, bankruptcy

Introduction

The economic development of highly developed countries is shaped mainly by small and medium enterprises, called further SMEs. It can simply be asserted that they employ up to 250 persons.[1] SMEs represent 99.8% of the enterprises operating in the European Union (EU) in the non-financial sector. They provide the workplace for more than 67% of all the people employed in the private sector (Schmiemann 2008). In each case, their bankruptcy poses a very high risk of unemployment in the area of their operation, especially in the less urbanized areas. Such areas include regions in Poland on the eastern border of EU, among others, the Podkarpacie region. Therefore, searching for efficient methods of assessment of their states is of great importance not only for us but also it fulfils an important social function.

1 Objectives and scope of the research

Generally, the goal of modelling the economic and financial situation of SMEs is prediction of their bankruptcy. One of the possible approaches to solve this problem is an attempt to discover patterns adequate for the assumptions of classification, mainly for the number and type of predefined classes. Reports in the literature in this area often show that machine learning methods, especially neural networks, are more effective than conventional statistical techniques (Lee, Booth, and Alam 2005; López-Iturriaga and Sanz 2014; Paliwal and Kumar 2009; Tseng and Hu 2010). Because of this advantage, as well as because of clear visualization of clustering results, we have used

1. In Poland, criteria of classification of enterprises according to their size are defined by the Economic Freedom Act.

self-organizing maps (SOMs), often called the Kohonen nets or Kohonen maps (Kohonen 1982), to discover hidden patterns of states of cases included in a database.

In statistical terms, SOMs belong to the group of non-parametric methods. They do not require assumptions about the parameters of the probability distribution of the input attributes. In addition to good clustering capabilities, they visualize at the plane, in the intuitive way, the spatial organization of the clusters obtained. An important feature of this projection is the topological neighbourhood taking into account the similarity of the patterns obtained. The goal of the conducted research is to discover and visualize patterns hidden in a database consisting of information about small and medium commercial enterprises from Podkarpackie Voivodship as well as to present them in the form useful in the business decision support processes.

2 Description of the research

2.1 The SME database

The raw database consists of 3544 cases — i.e., descriptions of enterprises of different legal status and economic activities, with the following distribution: 3501 **non-bankrupt** enterprises, 27 **in-liquidation** enterprises, and 16 **bankrupt** enterprises. Each case is described by means of 32 attributes (economic and financial indicators). The data come from the years 2000 to 2007.

In order to obtain a balanced data set, the whole number of cases has been reduced to 88 and the number of non-bankrupts to 45 using stratified sampling. The number of **in-liquidation** and **bankrupt** cases remained unchanged. According to the number of cases in the reduced data set, a number of descriptive attributes had to be significantly reduced too. The space of possible combinations of attributes has been searched by means of the genetic algorithm (Goldberg 1989), where a fitness function for a given subset was the accuracy of classification of a generalized regression neural network (GRNN) (Patterson 1996) applied for this subset. Finally, each object was described using five attributes, namely: Share of inventories in overall assets (SIOA), Cost level indicator (CLIR), Labor productivity (LAPR), Rate of sale change (RSCH), and Rate of employment change (RECH) (Kowerski 2006; Kowerski 2008; Nowak 2006).

2.2 Description of the method

The experiment has been performed using a self-organizing map (Kohonen net, Kohonen map). The architecture of the net for the 5-dimensional input space and the 2-dimensional map consisting of 16 neurons is shown in figure 1.

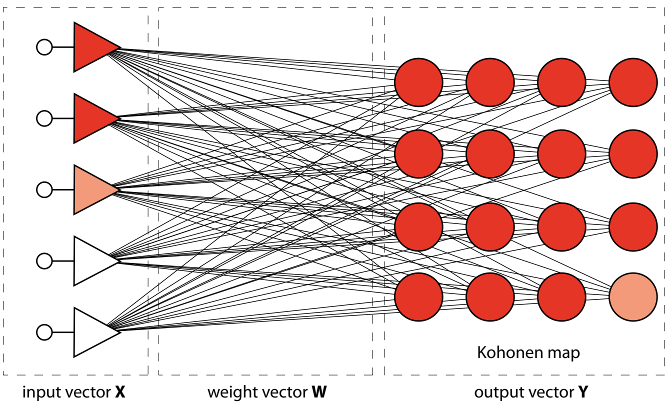

Fig. 1. The architecture of SOM used in the experiment

The net has been trained in the unsupervised mode using the Kohonen algorithm belonging to the group of WTM (Winner Takes Most) based algorithms (Haykin 1994). In the first step, responses of all neurons are calculated for given inputs and the winning neuron (the neuron with the highest output value) is selected. In the second step, the weights for the winning neuron and its neighbourhood are modified according to

$$W_i(t+1) = W_i(t) + \eta(t)G(i,x)[X(t) - W_i(t)], \tag{1}$$

where:
W_i is the weight vector of the i-th neuron,
X is the input vector,
t is the step index,
η is the learning rate,
$G(i,x)$ is the neighborhood function determined as in

$$G(i,x) = \begin{cases} 1 & \text{for } d(i,x) \le r \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

where r is a neighborhood radius and $d(i,x)$ denotes the Euclidean distance between weights of the i-th neuron and the winning neuron for the input x.

A change of weight vectors W of the winning neuron, as well as neurons from the neighborhood, causes the neurons to be more similar, in the sense of the vector similarity, to the input case. It means that, in the next epoch, these neurons react to this input vector X by the greater output.

To build a model, we have used Statistica 7.1. The following parameters have been set:

- the size of the Kohonen net: 4×4
- the learning rates and neighborhood radii
  - phase I (100 epochs): $\eta_{\max} = 0{,}1$ and $\eta_{\min} = 0{,}02$, $R_{\max} = 3$ and $R_{\min} = 1$
  - phase II (1000 epochs): $\eta_{\max} = 0{,}1$ and $\eta_{\min} = 0{,}01$, $R_{\max} = 0$ and $R_{\min} = 0$
- random initialization of weights (according to the Gauss distribution) from the interval (0; 1).[2]

Due to the assumption that obtained patterns are interpreted in the space of descriptive attributes (economic and financial indicators), the input vectors X were not normalized.

3 Results

Experiments performed according to the presented methodology enabled us to distinguish, on the map, 16 patterns existing in the available database (see section 2.1). Performing the test for all cases in the available database, we have determined the incidence of particular categories of cases (non-bankrupt, in-liquidation, bankrupt). The maps obtained for each category are shown in figure 2.

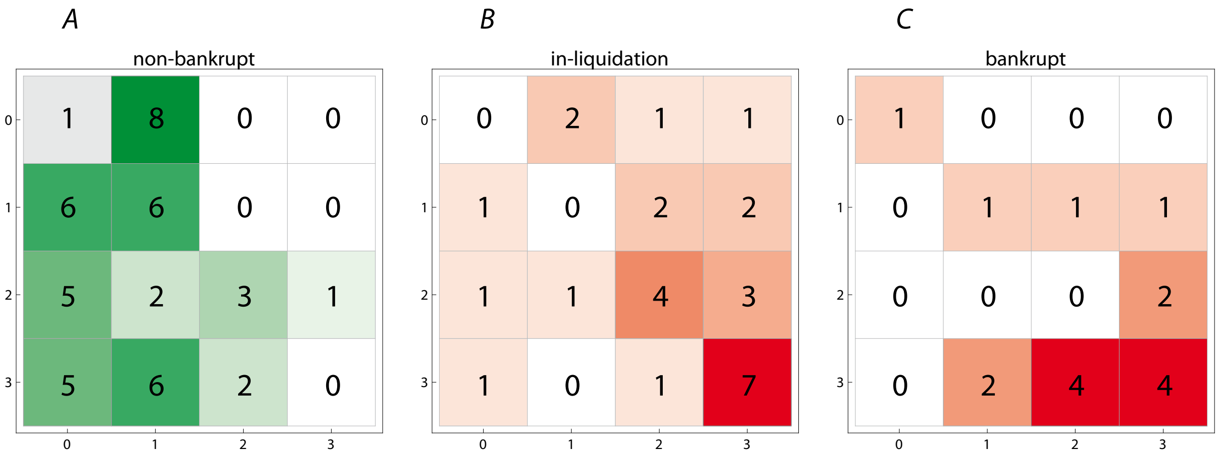

Fig. 2. The incidence of particular categories of cases: *A*—non-bankrupt, *B*—in-liquidation, *C*—bankrupt on the Kohonen map

2. [In the journal (in both Polish and English texts) European practice of number notation is followed — for example, 36 333,33 (European style) = 36 333.33 (Canadian style) = 36,333.33 (US and British style). Furthermore in the International System of Units (SI units), fixed spaces rather than commas are used to mark off groups of three digits, both to the left and to the right of the decimal point. — Ed.]

In further considerations, in view of the similarity of topological maps (figures 2*B* and 2*C*), categories **in-liquidation** and **bankrupt** were not distinguished. For the sake of the convention used by most researchers in the field of economy, both categories were covered by the common name **bankrupt**. The Kohonen map after this change is shown in figure 3.

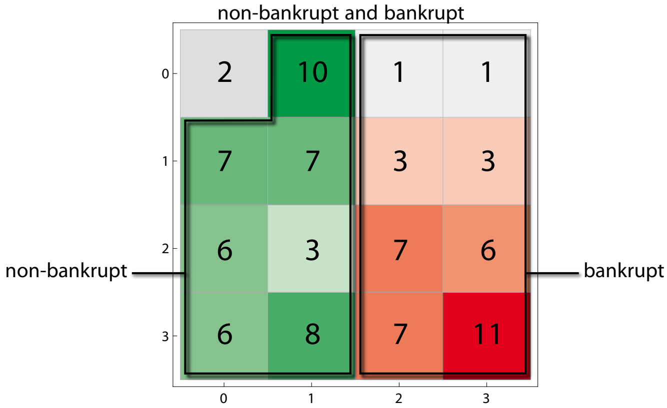

Fig. 3. The incidence of categories of cases: **non-bankrupt** and **bankrupt** on the Kohonen map

Such a defined clustering process, with the notable presence of outliers, probably unusual, or corrupted data (clusters 0, 2, and 3) is characterized by classification statistics, calculated for the whole available database, shown in table 1.

Tab. 1. Classification statistics of the model

Cases	Non-bankrupt n	Non-bankrupt %	Bankrupt n	Bankrupt %
Correctly classified	2209	62,33	31	72,09
Incorrectly classified	815	23,00	8	18,60
Unknown	520	14,67	4	9,30
All	3544	100,00	43	100,00

Each neuron of the Kohonen map (with numbers $n = 0, ..., 15$) is characterized by its weight vector W_n that is a pattern (prototype) of all cases recognized by it. Recognition is expressed by the greatest output value of the neuron for the input vector (attribute values describing the case).

With the help of this model, we can separate 13 patterns (described by W_n) belonging to two categories of enterprises: **non-bankrupt** and **bankrupt**. For the **non-bankrupt** category, the most common pattern is the pattern represented by cluster 1, whereas, for the **bankrupt** category, the pattern represented by cluster 15. To emphasize differences between patterns, all patterns are shown together on the bar graph (fig. 4). Due to domination of the attribute Share of inventories in overall assets (SIOA) over the remaining attributes, a similar graph, only for the remaining attributes, is shown in figure 5.

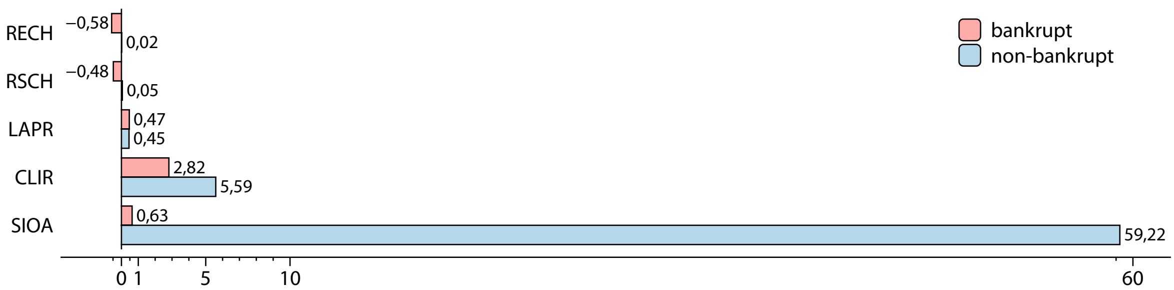

Fig. 4. Vectors of pattern features, where: RECH — Rate of employment change, RSCH — Rate of sale change, LAPR — Labor productivity, CLIR — Cost level indicator, and SIOA — Share of inventories in overall assets

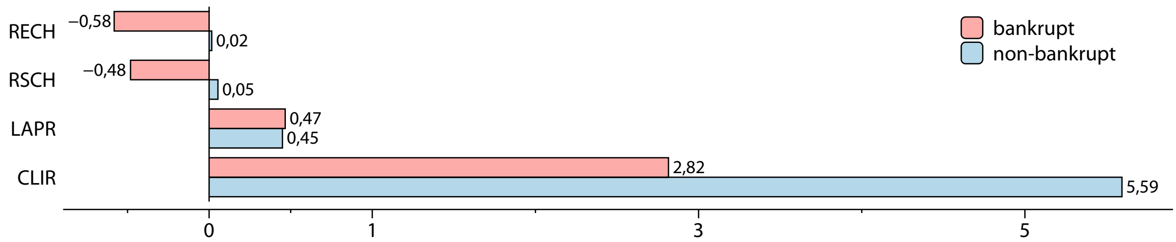

Fig. 5. Vectors of pattern features, where: RECH — Rate of employment change, RSCH — Rate of sale change, LAPR — Labor productivity, and CLIR — Cost level indicator

4 Discussion

The generated SOM model has been tested on the available database. In fact, with a limited number of cases belonging to the bankrupt category, the test was practically resubstitution (Burda and Hippe 2010; Reich and Barai 1999). Despite these limitations, the classification error, equal to 18,6% for the bankrupt category and 23% for the non-bankrupt category, enables us to accept the model as reliable.

The results of the experiment allow us to separate 6 patterns of SMEs of the bankrupt category and 7 patterns of SMEs of the non-bankrupt category. Within each of these categories, there are the majority patterns. Pattern 1 represents 22% of the total number of cases of the bankrupt category, whereas pattern 15 represents as many as 26% of cases of the non-bankrupt category. Differences between them shown in Figures 4 and 5 are statistically insignificant for the attribute Labor productivity (LAPR), but dominant for Share of inventories in overall assets (SIOA). Enterprises of the bankrupt category, unlike the non-bankrupt category, have additionally smaller values of Cost level indicator (CLIR), Rate of sale change (RSCH), and Rate of employment change (RECH). The last ones take negative values.

Conclusions

Our research shows that, within the available database, small and medium enterprises (in-liquidation or bankrupt) are located in the same region of the Kohonen map. It means that these cases do not show significant differences in applied 5-dimensional space of descriptive attributes. It enables us to carefully state that, in the current and further research, both categories can be joined in one category — i.e., bankrupt, without exposing us to a large loss of classification accuracy. The results of the performed experiment enable us to separate and describe the patterns of SMEs that can be classified to the bankrupt category or to the non-bankrupt category. Dynamic recognition of these patterns, provided in the next stage of our research, will probably enable us to discover bankruptcy or survival scenarios of small and medium commercial enterprises in the Podkarpacie region.

References

Burda, A., and Z.S. Hippe. 2010. Uncertain Data Modeling. The Case of Small and Medium Enterprises. In *3rd International Conference on Human System Interaction*, edited by T. Pardela and B. Wilamowski. Rzeszów.

Goldberg, D.E. 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning.* Reading, MA: Addison-Wesley Pub. Co.

Haykin, S.S. 1994. *Neural Networks. A Comprehensive Foundation.* New York-Toronto: Macmillan.

Kohonen, T. 1982. "Self-Organized Formation of Topologically Correct Feature Maps." *Biological Cybernetics* (43):59–69.

Kowerski, M. 2006. "Koncepcja badań sektora małych i średnich przedsiębiorstw w projekcie *System przeciwdziałania powstawaniu bezrobocia na terenach słabo zurbanizowanych.*" *Barometr Regionalny* no. 6:1–5.

Kowerski, M. 2008. "Assessment of the Economic Condition of Small Enterprises with Logit Micro-Macro Models. The Case of the Lubelskie Voivodship." *2008 Conference on Human System Interactions, Vols 1 and 2*:375–380.

Lee, K., D. Booth, and P. Alam. 2005. "A Comparison of Supervised and Unsupervised Neural Networks in Predicting Bankruptcy of Korean Firms." *Expert Systems with Applications* no. 29 (1):1–16. doi: 10.1016/j.eswa.2005.01.004.

López-Iturriaga, F.J., and I.P. Sanz. 2014. "Bankruptcy Visualization and Prediction Using Neural Networks: A Study of U.S. Commercial Banks." *Expert Systems with Applications*: Available online 25 November 2014, In Press, Uncorrected Proof – Note to users. doi: 10.1016/j.eswa.2014.11.025.

Nowak, E. 2006. "Propozycje zmiennych oceniających kondycję ekonomiczno-finansową przedsiębiorstw." *Barometr Regionalny* no. 6:35–41.

Paliwal, M., and U.A. Kumar. 2009. "Neural Networks and Statistical Techniques: A Review of Applications." *Expert Systems with Applications* no. 36 (1):2–17. doi: 10.1016/j.eswa.2007.10.005.

Patterson, D.W. 1996. *Artificial Neural Networks. Theory and Applications*. Singapore-New York: Prentice Hall.

Reich, Y., and S.V. Barai. 1999. "Evaluating Machine Learning Models for Engineering Problems." *Artificial Intelligence in Engineering* no. 13 (3):257–272. doi: 10.1016/S0954–1810(98)00021–1.

Schmiemann, M. 2008. "Enterprises by Size Class—Overview of SMEs in the EU." *Statistics in Focus* no. 31:1–8.

Tseng, F.-M., and Y.-C. Hu. 2010. "Comparing Four Bankruptcy Prediction Models: Logit, Quadratic Interval Logit, Neural and Fuzzy Neural Networks." *Expert Systems with Applications* no. 37 (3):1846–1853. doi: 10.1016/j.eswa.2009.07.081.