Expert Systems with Applications 40 (2013) 385–393

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Clustering and visualization of bankruptcy trajectory using self-organizing map

Ning Chen [a,*], Bernardete Ribeiro [b], Armando Vieira [a], An Chen [a,c]

[a] GECAD, Instituto Superior de Engenharia do Porto, Instituto Politecnico do Porto, Rua Dr. Antonio Bernardino de Almeida, 431, Porto, 4200-072, Portugal
[b] CISUC, Department of Informatics Engineering, University of Coimbra, Rua Silvio Lima-Polo II, Coimbra, 3030-790, Portugal
[c] Institute of Policy and Management, Chinese Academy of Sciences, No. 15, Zhongguancun Beiyitiao, Beijing, 100190, PR China

ARTICLE INFO

Keywords:
Bankruptcy risk
Trajectory pattern
Self-organizing map
Visual clustering

ABSTRACT

Bankruptcy trajectory reflects the dynamic changes of financial situation of companies, and hence make possible to keep track of the evolution of companies and recognize the important trajectory patterns. This study aims at a compact visualization of the complex temporal behaviors in financial statements. We use self-organizing map (SOM) to analyze and visualize the financial situation of companies over several years through a two-step clustering process. Initially, the bankruptcy risk is characterized by a feature self-organizing map (FSOM), and therefore the temporal sequence is converted to the trajectory vector projected on the map. Afterwards, the trajectory self-organizing map (TSOM) clusters the trajectory vectors to a number of trajectory patterns. The proposed approach is applied to a large database of French companies spanning over four years. The experimental results demonstrate the promising functionality of SOM for bankruptcy trajectory clustering and visualization. From the viewpoint of decision support, the method might give experts insight into the patterns of bankrupt and healthy company development.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Bankruptcy prediction has long been a well studied topic in the literature, and remains more active in the face of the recent severe challenges of worldwide financial crisis. The increase of financial failure accelerates the economic deterioration and yields a lot of social problems. It becomes critically important to explore the potential bankruptcy behaviors and understand the implicit patterns from the perspective of early warning and decision support. To date, a large amount of research has been carried out using different methods, such as univariate and multivariate analysis, neural network, support vector machine, and rough set. Most of the prior studies are focused on financial distress prediction problem based on a static snapshot of financial situation. However, bankruptcy trajectory which characterizes the dynamic changes of financial situation receives little attention. To our knowledge, only a few works attempted to analyze the temporal sequence of financial statements (Jardin & Severin, 2011; Kiviluoto & Bergius, 1998; Schreck, Bernard, Tekusova, & Kohlhammer, 2009).

Self-organizing map (SOM) is a non-parametric neural network with the desirable combination of data abstraction and spatialization, and widely used for visual clustering in a wide range of applications. In this paper, we study the changes of financial situation of companies to examine the trajectory patterns through a two-step clustering process by extending our earlier work (Chen, Ribeiro, & Vieira, 2011). A self-organizing map clustering approach is proposed to analyze (and visualize) the effect of temporal evolution of some financial indictors in order to assess and establish eventual scenarios of bankruptcy. Initially, a feature self-organizing map (FSOM) is constructed to characterize the bankruptcy risk of companies. Afterwards, the instantaneous observations of temporal sequence are successively projected on the map and the positions are concatenated to a trajectory vector. The trajectory patterns are then learned by a trajectory self-organizing map (TSOM) and shown through appropriate visual representation. The proposed approach is applied to a large data set of French companies containing financial ratios in four consecutive years (2003–2006) and the final state in the following year (2007). The experimental results demonstrate the promising functionality of SOM for bankruptcy trajectory clustering and visualization. Taking the perspective of decision support, the described method might give experts insight into patterns of bankrupt or healthy company development.

The remainder of this paper is organized as follows. Section 2 reviews the related studies with the emphasis on bankruptcy prediction and trajectory mining. Section 3 presents the data set under exploration and the methodology of a SOM-based trajectory analysis approach. Further details on each phase of the approach are schematically illustrated. In Section 4, the experimental results are reported including the model parametrization, trajectory pattern analysis and component plan visualization. Lastly, the contributions and future remarks are discussed in Section 5.

* Corresponding author. Tel.: +351 228340500; fax: +351 228321159.
E-mail addresses: ningchen74@gmail.com (N. Chen), bribeiro@dei.uc.pt (B. Ribeiro), asv@isep.ipp.pt (A. Vieira), acn@isep.ipp.pt (A. Chen).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.07.047

2. Related work

Corporate bankruptcy prediction is a well-researched area in finance to predict the probability of business failure, given the historic data that describe the situation of a company over a given period. There has been a raising interest in seeking for more accurate prediction models able to better understand the financial data and prevent the sudden distress of companies. So far a large number of methods have been proposed following the research direction of statistical or intelligent approaches (Bellovary, Giacomino, & Akers, 2007). Due to the criticism on traditional statistical models, many recent efforts have been devoted into state-of-the-art intelligent approaches, which offer theories about how financial crises could be predicted. Various prediction models have been proposed using a wide range of intelligent methods including neural networks (NNs), fuzzy set theory (FS), decision tree (DT), case-based reasoning (CBR), support vector machines (SVMs), and soft computing (Ravi Kumar & Ravi, 2007). Neural networks have been the subject of many research activities in bankruptcy prediction yielding reasonably accurate models (Charalambous, Charitou, & Kaourou, 2000). Form the many studies existing in the literature, neural networks are generally superior to other methods (Atiya, 2001). A multi-layer perceptron (MLP) obtains desirable outcome on Taiwan and United States markets (Huang, Chen, Hsu, Chen, & Wu, 2004), and Iranian companies (Rafiei, Manzari, & Bostanian, 2011). In Chen, Vieira, Ribeiro, Duarte, and Neves (2011), a stable credit rating model based on Learning Vector Quantization (LVQ) neural network is successfully applied to corporate failure prediction and credit risk analysis of French companies. Likewise, SVM has shown its applicability to this problem compared with other core machine learning techniques (Chen, 2011; Tsai, 2008; Yang, You, & Ji, 2011). Some authors have mentioned that a well designed hybrid or ensemble is effective to boost the predictive capability of single predictor (Verikas, Kalsyte, Bacauskiene, & Gelzinis, 2010). Take for instance, the hybrid credit scoring model of ANN and CBR (Chuang & Hung, 2011) and the ensemble that aggregates DT, MLP and SVM (Hung & Chen, 2009).

Regardless of different techniques, most of the previous work share the common goal at an accurate prediction of the final state of companies, in other words, whether a company is bankrupt or not after a given period. In this viewpoint, they study the bankruptcy problem as a static result more than a dynamic process. Whereas from the perspective of decision support, a more important issue than accurate prediction might be the deep analysis and well understanding of the dynamic changes of financial situation of companies. Some questions have been raised: Which factors contribute mostly to the difference between the bankrupt companies and healthy ones? How to characterize the evolution of companies and identify (and visualize) the important trend? Bankruptcy trajectory analysis is a subject attempting to find solutions to these issues.

2.1. Self-organizing map and applications in bankruptcy prediction

Self-organizing map (SOM) is an unsupervised neural network proposed by Kohonen (1982) for visual cluster analysis. The neurons of the map are located on a regular grid embedded in a low (usually 2 or 3) dimensional space, and associated with the cluster prototypes by the connected weights. In the course of learning process, the neurons compete with each other through the best-matching principle in such way that the input is projected to the nearest neuron given a defined distance metric. The winner neuron and its neighbors on the map are then adjusted towards the input in proportion with the neighborhood distance, consequently the neighboring neurons likely represent the similar patterns of the input data space. Due to the data clustering and spatialization through the topology preserving projection, SOM is widely used in the context of visual clustering applications. Despite the unsupervised nature, the applicability of SOM is extended to classification tasks by means of a variety of ways, such as neuron labeling method, semi-supervised learning (Heikkonen, Koikkalainen, & Oja, 1993), and supervised learning vector quantization (LVQ) (Kohonen, 2001).

In the field of bankruptcy prediction, SOM is one of the frequently used models to analyze the high-dimensional financial data and understand the unwanted bankruptcy phenomenon. A wide range of research groups concentrate on the bankruptcy prediction problem, usually solved as a classification task to separate the companies into distress and healthy category (binary class) or a number of predefined credit rates (multi-class). The prediction models are constructed from the training data in terms of some criteria, such as the overall misclassification error, the Neyman–Pearson criterion (Kiviluoto, 1998), and the total misclassification cost (Chen, Vieira, Duarte, Ribeiro, & Neves, 2009). The capability of SOM and its supervised variants has been demonstrated in comparison with statistical and other intelligent methods. Recent examples are as follows. SOM is used to determine the credit class through a visual exploration (Merkevicius, Garsva, & Simutis, 2004). An enhanced version of LVQ can boost the prediction performance of MLP (Neves & Vieira, 2006). It is demonstrated that LVQ outperforms other neural networks, support vector machines and multivariate statistical methods on predicting the financial failure of Turkish banks (Boyacioglu, Kara, & Baykan, 2009). In cooperation with independent component analysis for dimensionality reduction, LVQ is able to recognize the distressed French companies (Chen & Vieira, 2009).

2.2. Trajectory mining and bankruptcy trajectory analysis

Trajectory data has attracted considerable attention in many applications where the object movements are routinely collected as time-dependent observation sequences, such as traffic monitoring, visual surveillance, robotic navigation, and stock prediction. Trajectory mining attempts to explore the implicit patterns from trajectory data. To achieve this, SOM has been introduced as a valuable tool in various applications. In a robotic navigational environment, SOM clusters the motion trajectory of moving objects and predicts the next instant position (Rajpurohit & Manohara, 2009). A SOM clustering model provides a powerful tool to visualize the dynamic behaviors of industrial processes for human supervision and fault detection (Fuertesa et al., 2010). In Schreck et al. (2009), a SOM-based framework is designed for general-purpose enabling users to visually monitor and interactively control the clustering process of trajectory data.

In the field of bankruptcy analysis, we may map the financial situation of a company to a 2-dimensional space, and observe the evolution of the financial situation as a trajectory in that space. These trajectories reflect the dynamic changes rather than a static snapshot of financial situation, and hence make possible to detect the time evolution of companies and recognize the trajectory patterns. SOM is regarded as an appealing information visualization technique for bankruptcy trajectory analysis due to its superiority on transforming the information into graphical representation so as to facilitate the pattern recognition and visual reasoning by decision makers. In Kiviluoto and Bergius (1998), a hierarchical SOM model is employed to explore the year-to-year trajectory of enterprises and view the time evolution of the situation. A recent study (Jardin & Severin, 2011) shows the advantages of using trajectory for medium-period forecast compared with the traditional single-period and multi-period forecast. They used SOM to design the trajectory of corporate collapse, and then forecasted the financial failure at horizons of three years. The main intention is to predict the failure of companies rather than the trajectory clustering and visualization. Our study in this paper follows the idea of Kiviluoto and Bergius (1998), examining the trajectories of companies and recognizing the influential factors relevant to bankruptcy risk. The main purpose is not to deliver a reliable classification or prediction regarding the bankruptcy risk as was done in most of the previous research. Instead, we aim at a compact visualization of the complex temporal behaviors in financial statement.

3. Methodology of bankruptcy trajectory clustering

The trajectory clustering procedure is basically a two-step consecutive learning process. Two self-organizing maps are integrated in the approach for different uses. The first SOM (FSOM) is trained without explicitly taking the temporal aspects of the data into account. The high-dimensional vectors of financial indicators in each time step are mapped to the respective best-matching units (BMUs) in the SOM grid. Their grid coordinates are then made up to a compressed two-dimensional representation of the original data, while the temporal progression for each company can be represented as the sequence of respective BMUs in the order of the time steps of previous years. In this manner, each company yields a trajectory in the SOM neighborhood grid. Afterwards, these trajectories are visualized and clustered by a second SOM (TSOM) to find the behavioral patterns in the set of trajectories.

Specifically, the trajectory clustering approach shown in Fig. 1 is composed of six phases, namely data preparation, feature clustering, trajectory representation, trajectory clustering, pattern representation, and visualization. Further details of each phase will be described in the following.

3.1. Data description

The original French database contains the financial statements of 110,723 companies in small or middle size during the year 2003–2006. From the financial statements, 29 numerical financial ratios are derived shown in Table 1. The class indicator gives the state of companies in the year 2007 in which 2792 observations went bankrupt eventually.

3.2. Data preparation

Data preprocessing is an integral part of clustering procedure, usually composed of sampling, logarithmization, and normalization. Since we consider the time evolution of companies during the year 2003 to 2006, only the companies which are in operation in all the years are preserved. (1) We chose all companies distressed in 2007 with at least 20% available values in each fiscal year, yielding a total of 718 distressed companies. It was known that the classification tends to favor the majority class (non-default companies) under the highly skewed distribution of the original database. To create a balanced data set, a random sampling was then employed to select a same number of healthy companies. Consequently, a balanced sample is made up of 1436 observation sequences. (2) A logarithmized operation was applied to mitigate the scatter distribution of the data as was sone in Chen et al. (2009). (3) The financial ratios vary greatly in scales, so we normalized the data set to zero mean and unit variance.

Table 1
Financial ratios of French database.

Variable description	
x_1 – Number of employees previous year	x_{16} – Cashflow/turnover
x_2 – Capital employed/fixed assets	x_{17} – Working capital/turnover days
x_3 – Financial debt/capital employed	x_{18} – Net current assets/turnover days
x_4 – Depreciation of tangible assets	x_{19} – Working capital needs/turnover
x_5 – Working capital/current assets	x_{20} – Added value per employee in k euros
x_6 – Current ratio	x_{21} – Total assets turnover
x_7 – Liquidity ratio	x_{22} – Operating profit margin
x_8 – Stock turnover days	x_{23} – Net profit margin
x_9 – Collection period days	x_{24} – Added value margin
x_{10} – Credit period days	x_{25} – Part of employees
x_{11} – Turnover per employee k euros	x_{26} – Return on capital employed
x_{12} – Interest/turnover	x_{27} – Return on total assets
x_{13} – Debt period days	x_{28} – EBIT margin
x_{14} – Financial debt/equity	x_{29} – EBITDA margin
x_{15} – Financial debt/cashflow	x_{30} – Class (bankrupt, healthy)

A preliminary analysis is performed on the experimental data. Fig. 2 illustrates the mean values of the financial ratios presented in Table 1 with respect to bankrupt companies and healthy companies. A t-test is performed on each ratio to evaluate the difference between two classes. A small p-value indicates that the two groups significantly differ in the mean of underlying financial ratio. As it may be observed, all financial ratios except x_8 ($p = 0.051$), x_9 ($p = 0.864$) and x_{19} ($p = 0.263$) are significantly different in the mean between the bankrupt and healthy companies at the level 5%.

3.3. Feature clustering

As demonstrated in the previous studies, one-year prior to the bankruptcy announcement is able to derive accurate bankruptcy models. We use the data composed of the financial ratios one year before bankruptcy to generate a feature SOM (FSOM) which characterizes properly the bankruptcy risk.

The network is trained by adjusting the prototypes iteratively on the foundation of winner-take-all principle until the termination condition (e.g., the maximum of iterations or the variation of prototypes) is fulfilled. For each input, only the winner and its neighbors modify their weights, and the change is determined by how close the input is to the winner, the learning rate $\alpha(t)$ and a neighborhood function $h(t)$. Assuming the map consists of m neurons each associated with a prototype w_p ($p = 1, \ldots, m$), the learning process is performed as follows:

1. For $p = 1, \ldots, m$, initialize the map with prototypes w_p;
2. For each input instance x, project it to the BMU c using Euclidean distance, defined as:

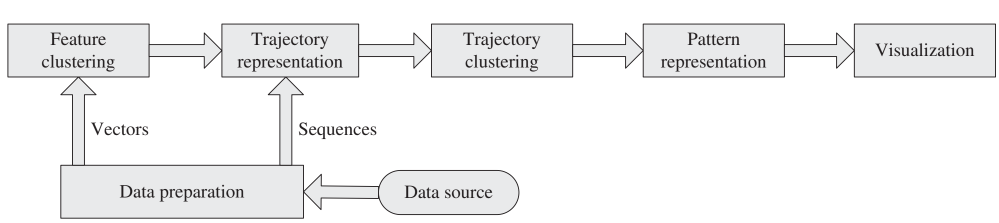

Fig. 1. Methodology of trajectory clustering.

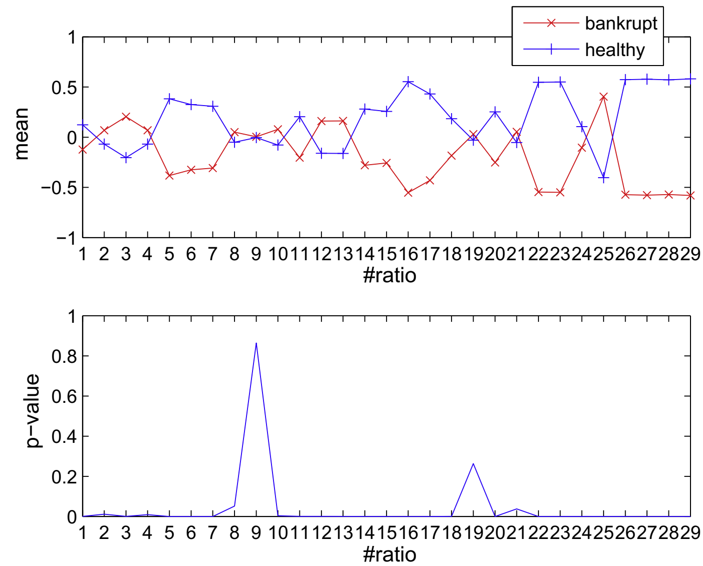

Fig. 2. The preliminary analysis of financial data.

$$c = \operatorname*{argmin}_{1 \leqslant i \leqslant m} \|x - w_i\|$$

3. For $p = 1, \ldots, m$, update the prototypes:

$$w_p(t+1) = w_p(t) + \alpha(t) h_{c,p}(t)(x - w_p(t))$$

4. Repeat from Step 2 a few iterations until the termination condition is satisfied.

A labeling operation is then applied to the learned map in a supervised fashion. By projecting the instances to the BMUs, the input space is divided into a finite number of Voronoi sets V_p ($p = 1, \ldots, m$), constituted by the instances projected on the underlying neuron:

$$V_p = \{x | \ p = \operatorname*{argmin}_{1 \leqslant i \leqslant m} \|x - w_i\|\}$$

Each neuron is associated with a percentage of default (PD) value defined as the ratio of bankrupt companies in the Voronoi set:

$$\mathrm{PD}(w_p) = \frac{|\{x \in V_p \mid \mathrm{Class}(x) = \mathrm{bankrupt}\}|}{|V_p|}$$

For the zero-hit neurons, the PD value is 0. By taking the value as a real number in [0,1], PD conveys the information of bankruptcy risk. The higher the value of PD, the higher the risk of a company being distressed in the following year. In this manner, the neurons can be labeled by the majority class indicator, in other words, the neuron is labeled as 'bankrupt' if the PD value is more than 0.5, and 'healthy' otherwise.

$$\mathrm{Label}(w_p) = \begin{cases} \mathrm{bankrupt} & \text{if } \mathrm{PD}(w_p) > 0.5 \\ \mathrm{healthy} & \text{otherwise} \end{cases}$$

As the class of companies is known in prior, a classification phase implemented by learning vector quantization (LVQ) is imposed on the resultant map to boost the discriminative performance. The learning process of LVQ is similar to SOM except that the class label is taken into consideration in the prototype update.

1. For each input instance x, project it to the BMU c using Euclidean distance:

$$c = \operatorname*{argmin}_{1 \leqslant i \leqslant m} \|x - w_i\|$$

2. Update the prototype w_c:

$$w_c(t+1) = \begin{cases} w_c(t) + \alpha(t)(x - w_c(t)) & \text{if } Label(w_c) = Class(x) \\ w_c(t) - \alpha(t)(x - w_c(t)) & \text{otherwise} \end{cases}$$

3. Repeat from Step 1 a few iterations until the termination condition is satisfied.

In summary, the feature clustering procedure proceeds with an unsupervised SOM followed by a supervised LVQ. The resultant map characterizes the bankruptcy risk of companies in terms of the percentage of default associated with the neurons.

3.4. Trajectory representation

For trajectory clustering, of primary concern is the meaningful interpretation of trajectory data. In this study, we consider a simple vector representation. As a projection-based approach, SOM preserves the topological property between the input space and the output space, thereby the feature vector can be represented by the position of BMUs on the map. The projection operation is applied to the financial vectors yearly for each company. Given a company, (p_{xi}, p_{yi}) denote the x-coordinate and y-coordinate values of its BMU on the map with respect to the ith feature vector. The concatenation of the coordinates of BMUs along the sequence yields the trajectory vector. As a result, we have a trajectory vector formatted as $\bar{x} = \langle p_{x1}, p_{y1}, \ldots, p_{xk}, p_{yk} \rangle$ where k is the number of years. Each trajectory vector is assigned with the final state of the company as the class label. In this manner, the temporal sequence of a company is transformed to a compact form of trajectory vector.

3.5. Trajectory clustering

In the same way as feature clustering, a trajectory SOM (TSOM) is then constructed taking as input the produced trajectory vectors. The learning is also conducted in an unsupervised manner so that the trajectories are grouped into a number of clusters based on the similarity.

In order to find the natural clusters, we firstly calculate a TSOM with a large number of neurons to cluster the trajectories and then use a k-means algorithm to partition the neurons to a small number of groups (Vesanto & Alhoniemi, 2000). The k-means algorithm is run multiple times for each value of k, and the best one is selected based on the sum of squared errors (for different initial means) and the Davies-Bouldin index (for different k values). Each trajectory vector is assigned to the cluster to which its BMU on TSOM belongs. Finally, the final clusters are labeled on the foundation of major voting scheme.

3.6. Pattern representation

On the basis of the clusters of trajectory clustering, the trajectory patterns can be extracted, taking the mean of the resulting clusters as the representative trajectory. For the sake of visualization, the mean values are rounded so as to indicate the position of neurons on FSOM.

3.7. Visualization

Compared with other clustering methods, one of the advantages of SOM is its capability of data exploration through a visual approach. Due to the topology-preserving projection, both the cluster structure and the non-linear correlation between variables can be detected from the link of neurons on the map (Hanafizadeh & Mirzazadeh, 2011). Some visualization techniques are listed as follows:

- *U-distance matrix*: The u-distance calculates the average distance between the prototype and its neighbors. It discloses the clustering structure under the assumption that the higher values usually identify the cluster boundary. Particularly, a part of the components can be used in calculating the u-distance matrix in order to compare the relative importance which contributes to make the cluster.
- *Component plane*: Each component plane shows the values of one variable in each map unit. Given this, the correlation between the variables can be analyzed.
- *Trajectory*: It shows the trajectory of time sequences on the map by connecting the BMUs. Through the trajectory visualization, the dynamic characters of the sequences can be detected.

In this present study, several kinds of visualization can be applied for the sake of visual trajectory inspection. (1) The PD values associated with each cluster are visualized stating the risk of bankruptcy. (2) The trajectory of a company projected on the map reflects the evolution of its financial situation. (3) The patterns derived from the trajectory clusters represent some typical behaviors such as the deterioration or amelioration associated with the bankruptcy risk. (4) The important financial ratios can be recognized which have significant influence on the state of companies. These visualizations will be shown in Section 4.

4. Experiments and results

In this section, we show the experimental results of the proposed approach on the French database, from which the bankruptcy trajectory patterns are detected and the influential factors are analyzed.

4.1. SOM parameterization

Since the rationale behind the trajectory representation is the topology-preservation projection of SOM, the topographic function can be used as the measurement for parameter specification and model evaluation. As was suggested in Kiviluoto (1996), resolution and topology preservation are two typical evaluation criteria of SOM. The former is measured by quantization error (qe), i.e., the average distance between each data vector and its BMU, and the latter is measured by topographic error (te), i.e., the proportion of all input vectors for which the first and second BMUs are not adjacent on the map. Additionally, the classification error (ce) measures the discriminative power of the learned model.

$$qe = \sum_{i=1}^{n} \|x_i, BMU(x_i)\|/n$$

$$te = \sum_{i=1}^{n} u(x_i)/n, \text{ where } u(x_i) = \begin{cases} 0 & \text{if BMU and 2-BMU are adjacent} \\ 1 & \text{otherwise} \end{cases}$$

In the experiments, we investigate the quality of FSOM in terms of the three qualitative measurements, namely qu, te and ce. Fig. 3 shows the results of different maps changing the size from $[3 \times 3]$ to $[10 \times 10]$. (We use square hexagonal topology with Gaussian neighborhood kernel according to the common practice.) As the map becomes larger, quantization error decreases at the cost of topographic error. Due to the fact that we intend to construct a map with both good topology preservation and discrimination ability, we used a $[5 \times 5]$ map for feature clustering, which has the lowest classification error (0.209), reasonable quantization error (0.39) and quantization error (3.44). Now that the size of FSOM is determined, the coordinate value of p_{xi} and p_{yi} is restricted to $\{1, 2, 3, 4, 5\}$.

The class label (left panel) and PD value (right panel) are visualized on the feature SOM in Fig. 4. The gray level denotes the PD value associated with the cluster, indicating that the higher risk is marked by a darker color. It can be seen that the top region has the high bankruptcy risk, and the bottom region has the low risk relatively.

As described in the previous section, the proposed approach consists of two self-organizing maps (see Fig. 5): the FSOM clusters the feature vectors $x = \langle x_1, x_2, \ldots, x_{29} \rangle$ composed of financial ratios of companies in 2006, and the TSOM clusters the trajectory vectors $\bar{x} = \langle p_{x1}, p_{y1}, p_{x2}, p_{y2}, p_{x3}, p_{y3}, p_{x4}, p_{y4} \rangle$ composed of the positions yearly on the FSOM. The resulting TSOM is subsequently fed to the k-means algorithm, for which the optimal k (number of clusters) is selected. In this study, 12 clusters are achieved as the result of k-means. In Fig. 6, the cluster distribution is visualized and the class label is marked in position of each neuron. It can be seen that the resulting clusters separate well the bankrupt companies from the healthy companies.

4.2. Trajectory pattern analysis

As was described in Section 3, the resultant trajectory clusters are represented by the mean, called trajectory pattern. In Fig. 7, the trajectory patterns, denoted as $P1, P2, \ldots, P12$, are shown on feature SOM (one map per cluster) by connecting the projected neurons of yearly vector. For each trajectory, the color indicates

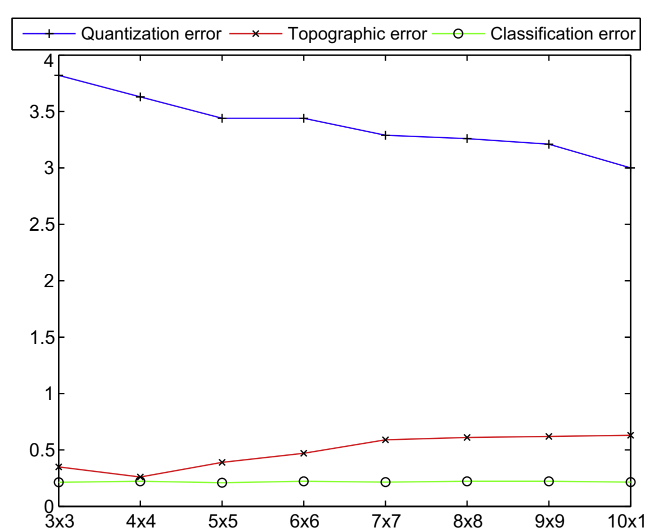

Fig. 3. The quality of FSOM vs. map size.

Fig. 4. The label and PD distribution of FSOM.

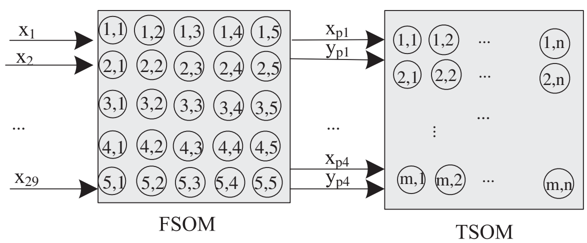

Fig. 5. The structure of feature SOM and trajectory SOM.

Fig. 6. The cluster distribution on TSOM ('bad': bankrupt, 'good': healthy).

the final state (red: 'bankrupt', green: 'healthy'), and the arrow denotes the trajectory from start point (2003) to end point (2006).

We can summarize the patterns into 4 typical types in Table 2, in which the first two types are risky and go bankrupt eventually. Among these patterns, type 'b' and 'c' are of primary interest, which denote the significant changes (deterioration or amelioration) of financial situation over the period.

- 'a' denotes the stably bad trajectory which always stays in the high risk region;
- 'b' denotes the significantly worse trajectory jumping from low risk region to high risk region, e.g., P5 has an increasing PD value, originating from the low risk region in the year 2003 and entering the high risk region in the year 2005;
- 'c' implies the significantly better trajectory jumping from the high risk region to the low risk region, e.g., P11 has a distinct improvement in the year 2004 compared with the year 2003;
- 'd' denotes the stably good trajectory which always stays in the low risk region during four years, indicating the perfect development of companies.

4.3. Component planes

Normally financial data is described by a large number of variables, in which some have direct relation with the bankruptcy risk. Feature selection is an issue of interest for the visualization of component planes. Strategies commonly used for explanatory variable selection include the empirical suggestion by previous studies or financial experts, univariate statistical test, and automatic search based on a predefined evaluation criterion (Jardin, 2010). In this study, we used several ranking algorithms to select the most relevant ratios as suggested in Aoyama, Fujiwara, Ikeda, Iyetomi, and Souma (2010). From the initial 29 financial ratios defined by COFACE, eight ratios are pinpointed as important factors with significant influence on the financial state, which are x_1 (No. of Employees), x_7 (Liquidity Ratio), x_{14} (Financial Debt/ Equity), x_{15} (Financial Debt/ Cashflow), x_{16} (Cashflow/ Turnover), x_{19} (Working Capital Needs/ Turnover), x_{21} (Total Assets Turnover) and x_{29} (EBITA Margin). Indeed as presented in Fig. 2, these selected financial ratios (components) except x_{19} have significantly different means between distressed companies and healthy companies.

In Fig. 8, the significance of the components with respect to the clustering are visualized by means of u-distance matrix. As was indicated in Kohonen (2001), the value of variables on the border of clusters changes more rapidly than the inner region, which gives the evidence to find the main factor in separating the cluster from the rest. Firstly, the u-distance matrix is calculated with respect to each component. The u-distance matrices are then shown in barchart (each bar represents the u-distance of single variable) and piechart (each pie represents the portion of u-distance incurred by single variable), indicating the relative importance of each variable in each map unit. The size of piecharts is scaled by the u-distance matrix calculated from all components so that the big pies likely indicate the border of clusters. As was shown in this figure, some variables have relatively large values on a part of borders, and the others have large values on the rest. It seems that the 8 variables selected have significant effect to make the clusters.

Component plane representation displays the scalar values of variables across the map using a multi-spectral image. We use a grey-level image in the way that the larger value is marked by a darker color. From the component planes, not only the mutual relation between financial ratios, but also the relation between univariate ratio and the bankruptcy risk are easily recognized. In Fig. 9, the first plane shows the PD value on the FSOM, and the others show the variable map of important factors (one map per factor). For simplicity of interpretation, four representative patterns

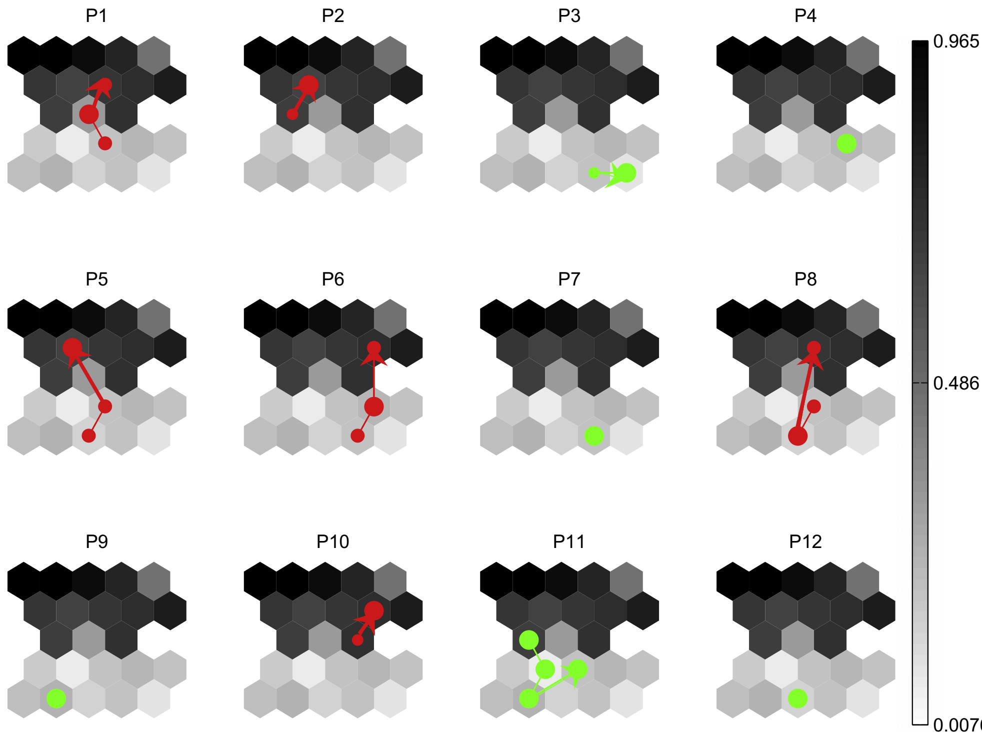

Fig. 7. Trajectory patterns visualization: 2003–2006. (Shown are the FSOM with PD value and trajectory patterns corresponding to each cluster.)

Table 2
Four types of trajectory patterns.

Type	Neurons	Description	Final state
a	P1, P2, P10	Stably bad	Bankrupt
b	P5, P6, P8	Significantly worse	Bankrupt
c	P11	Significantly better	Healthy
d	P3, P4, P7, P9, P12	Stably good	Healthy

indicating the deterioration or amelioration behavior are selected as example, namely P5, P6, P8, and P11. As can be seen in the planes, the variables x_{16}, x_{29} are inversely related to PD, in other words, they take apparently bigger values in the low risk region (the bottom region on the map) than the high risk region (the top region). It means these factors discriminate effectively the healthy companies and the bankrupt ones. Particularly, x_{16} and x_{29} are positively correlated due to the fact that when the former takes high value on the map the latter is also high in most cases, and vice verse. Consequently, it can be concluded that increasing the Cashflow turnover (x_{16}) and EBITA margin (x_{29}) contributes to the improvement of financial situation, yielding a low bankruptcy risk in successive years. In contrast, no direct correlation is presented between the value of other components and PD from their component planes. For example, there is a mixture of high and low

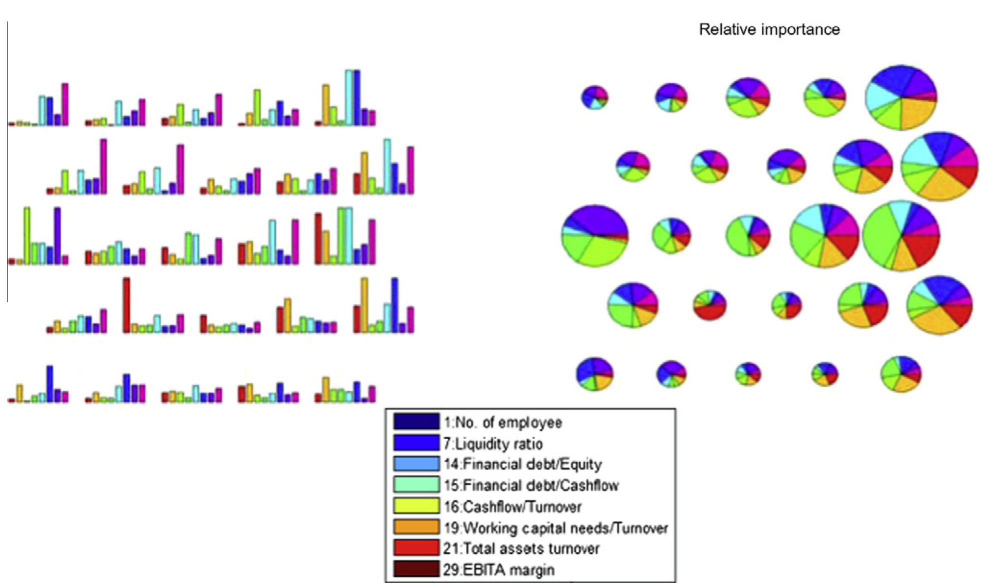

Fig. 8. Relative importance of variables.

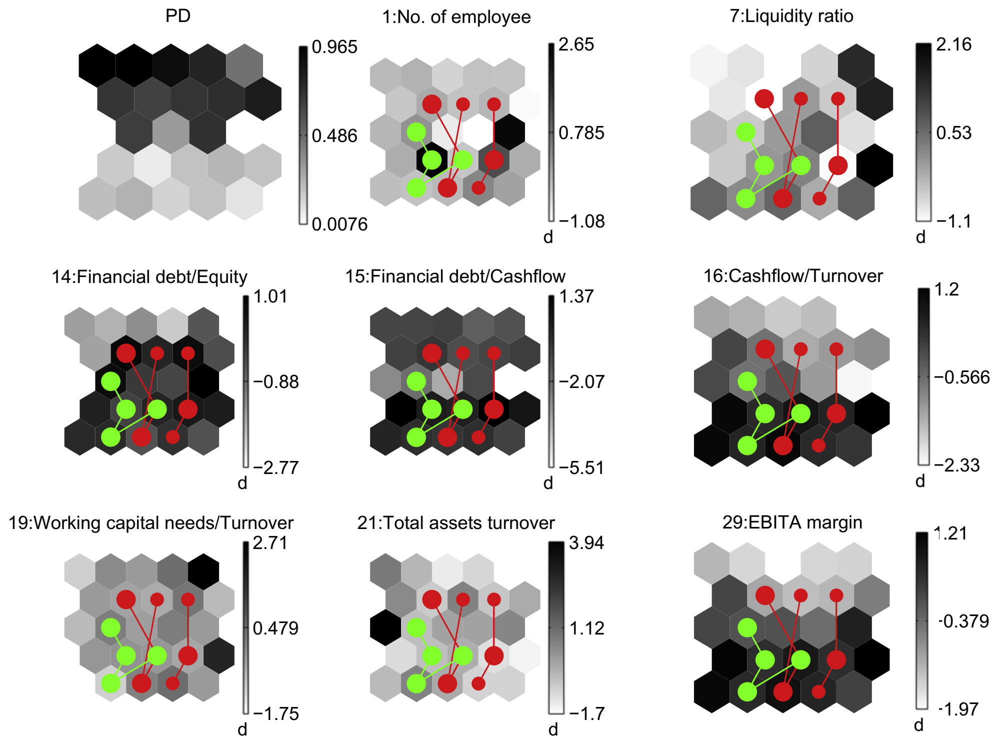

Fig. 9. Trajectory patterns on component planes: 2003–2006. (Shown are the FSOM with PD value and component planes with representative trajectory patterns.)

values for x_1 corresponding to the low risk region. Likewise, the companies positioned in the high risk region take a wide variety of values for x_{14}.

5. Conclusions

Self-organizing map is a useful visual data mining approach to explore a large amount of data. The projection defined from the high dimensional input data to a 2-dimensional visualization space makes the patterns graphically represented and easily recognizable. Despite the fact that SOM has been considerably used for financial failure prediction, its potential on trajectory analysis in finance field is not fully developed. In this paper, a self-organizing map clustering approach is used to cluster and visualize the character of the temporal progression of financial indicators. The idea is to assist experts with observing the development of companies over time and assessing their risk of corporate bankruptcy. The feature SOM clusters the financial vectors of observations and converts the temporal sequences to trajectory vectors by successive projection of the instantaneous statements to their BMUs. Sequentially, the trajectory SOM clusters the trajectory vectors and trajectory patterns can be detected from the resultant clusters. Through the visual interpretation, the important factors are analyzed qualitatively to understand the influence on bankruptcy risk. The results are presented using a data set of financial ratios from French companies, where the class label indicates whether the companies ended up bankrupt or not. The proposed SOM-based approach yields meaningful and appealing results, demonstrating the practicability on bankruptcy trajectory analysis. Some typical trajectory patterns are discovered reflecting the significant changes (deterioration or amelioration) of financial situation over the period. Moreover, we found increasing Cashflow turnover and EBITA margin is of benefit to the improvement of financial situation of companies.

The future study will consider the following research directors. Firstly, to verify the afore-mentioned findings presented in the study more experiments will be conducted on the comparison with existing methods, and the evaluation on different data sets. As was shown SOM provides a visual, interpretable way to the correlation between variables and bankruptcy risk, however a quantitative measurement is still helpful to denote the impact (Yang et al., 2011). Secondly, although the present study does not aim for the prediction models, the derived trajectory patterns can be used in corporate bankruptcy prediction by classifying the observed sequence to the best-matching cluster. The applicability of the model in prediction will be justified through empirical research. Thirdly, the proposed trajectory analysis method will improve the present AIRES (Advanced Intelligent Risk Evaluation System).[1] In fact, we have developed the SOM model for risk evaluation and visualization of French companies in AIRES. In the future, we intend to enhance the system by embedding the trajectory clustering and visualization techniques. Finally, the proposed framework provides a general way for trajectory analysis, and could be valuable to other trajectory-relevant data mining applications apart from the financial data.

References

Aoyama, H., Fujiwara, Y., Ikeda, Y., Iyetomi, H., & Souma, W. (2010). *Econophysics and companies: Statistical life and death in complex business networks*. Cambrige University Press.

Atiya, A. F. (2001). Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE Transactions on Neural Networks, 12*(4), 929–935.

Bellovary, J., Giacomino, D., & Akers, M. (2007). A review of bankruptcy prediction studies: 1930 to present. *Journal of Financial Education, 33*, 1–43.

Boyacioglu, M., Kara, Y., & Baykan, O. (2009). Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in turkey. *Expert Systems with Applications, 36*, 3355–3366.

Charalambous, C., Charitou, A., & Kaourou, F. (2000). Application of feature extractive algorithm to bankruptcy prediction. In *International joint conference on neural networks* (Vol. 5, pp. 303–308).

Chen, M.-Y. (2011). Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches. *Computers and Mathematics with Applications, 62*(12), 4514–4524.

[1] Web Site: http://aires.dei.uc.pt/AiresII.

Chen, N., & Vieira, A. (2009). Bankruptcy prediction based on independent component analysis. In *1st International conference on agents and artificial intelligence (ICAART09)* (pp. 150–155).

Chen, N., Vieira, A., Duarte, J., Ribeiro, B., & Neves, J. (2009). Cost-sensitive learning vector quantization for financial distress prediction. In *Lecture notes in artificial intelligence (LNAI 5816), 14th Portuguese conference on artificial intelligence (EPIA)* (pp. 374–385).

Chen, N., Ribeiro, B., & Vieira, A. S. (2011). Bankruptcy trajectory analysis on french companies using self-organizing map. In L. Antunes & H. Pinto (Eds.), *Lecture notes in artificial intelligence*. *EPIA 2011* (vol. 7026, pp. 407–417). Heidelberg: Springer.

Chen, N., Vieira, A., Ribeiro, B., Duarte, J., & Neves, J. (2011). A stable credit rating model based on learning vector quantization. *Intelligent Data Analysis, 15*(2), 237–250.

Chuang, C., & Hung, S. (2011). A hybrid neural network approach for credit scoring. *Expert Systems: The Journal of Knowledge Engineering, 28*(2), 185–196.

Fuertesa, J., Domingueza, M., Regueraa, P., Pradaa, M., Diazb, I., & Cuadrado, A. (2010). Visual dynamic model based on self-organizing maps for supervision and fault detection in industrial processes. *Engineering Applications of Artificial Intelligence, 23*(1), 8–17.

Hanafizadeh, P., & Mirzazadeh, M. (2011). Visualizing market segmentation using self-organizing maps and fuzzy delphi method-adsl market of a telecommunication company. *Expert Systems with Applications, 38*, 198–205.

Heikkonen, J., Koikkalainen, P., & Oja, E. (1993). Self-organizing maps for collision free navigation. In *World congress on neural networks, volume III of international neural network society annual meeting* (pp. 141–144).

Huang, Z., Chen, H., Hsu, C., Chen, W., & Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision Support Systems, 37*, 543–558.

Hung, C., & Chen, J. (2009). A selective ensemble based on expected probabilities for bankruptcy prediction. *Expert Systems with Applications, 28*(3), 5297–5303.

Jardin, P. (2010). Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy. *Neurocomputing, 73*, 2047–2060.

Jardin, P., & Severin, E. (2011). Predicting corporate bankruptcy using a self-organizing map: An empirical study to improve the forecasting horizon of a financial failure model. *Decision Support Systems, 51*(3), 701–711.

Kiviluoto, K. (1996). Topology preservation in self-organizing maps. In *IEEE international conference on neural networks* (Vol. 1, pp. 294–299).

Kiviluoto, K. (1998). Predicting bankruptcies with the self-organizing map. *Neurocomputing, 21*, 191–201.

Kiviluoto, K., & Bergius, P. (1998). Exploring corporate bankruptcy with two-level self-organizing maps. In *The 5th international conference on computational finance* (pp. 373–383).

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics, 43*(1), 141–152.

Kohonen, T. (2001). *Self-organizing maps* (3rd ed.). Springer Verlag.

Merkevicius, E., Garsva, G., & Simutis, R. (2004). Forecasting of credit classes with the self-organizing maps. *Informaciens Technologijos Ir Valsymas, 4*(33), 61–66.

Neves, J., & Vieira, A. (2006). Improving bankruptcy prediction with hidden layer learning vector quantization. *European Accounting Review, 15*(2), 253–271.

Rafiei, M., Manzari, S., & Bostanian, S. (2011). Financial health prediction models using artificial neural networks, genetic algorithm and multivariate discriminant analysis: Iranian evidence. *Expert Systems with Applications, 38*(8), 10210–10217.

Rajpurohit, V., & Manohara, P. (2009). Using self organizing networks for moving object trajectory prediction. *International Journal on Artificial Intelligence and Machine Learning, 9*(1), 27–34.

Ravi Kumar, P., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques – a review. *European Journal of Operational Research, 180*(1), 1–28.

Schreck, T., Bernard, J., Tekusova, T., & Kohlhammer, J. (2009). Visual cluster analysis of trajectory data with interactive Kohonen maps. *Information Visualization, 8*, 14–29.

Tsai, C.-F. (2008). Financial decision support using neural networks and support vector machines. *Expert Systems: The Journal of Knowledge Engineering, 25*(4), 380–393.

Verikas, A., Kalsyte, Z., Bacauskiene, M., & Gelzinis, A. (2010). Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: a survey. *Soft computing – a fusion of foundations, methodologies and applications, 14*(9), 995–1010.

Vesanto, J., & Alhoniemi, E. (2000). Clustering of the self-organizing map. *IEEE Transactions on Neural Networks, 11*(3), 586–600.

Yang, Z., You, W., & Ji, G. (2011). Using partial least squares and support vector machines for bankruptcy prediction. *Expert Systems with Applications, 38*(7), 8336–8342.