Knowledge-Based Systems 26 (2012) 196–206

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

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

The prediction for listed companies' financial distress by using multiple prediction methods with rough set and Dempster–Shafer evidence theory

Zhi Xiao [a], Xianglei Yang [a,*], Ying Pang [a], Xin Dang [b]

[a] School of Economics and Business Administration, Chongqing University, Chongqing 400044, PR China
[b] Department of Mathematics, University of Mississippi, MS 38677, USA

ARTICLE INFO

Article history:
Received 17 March 2011
Received in revised form 7 August 2011
Accepted 8 August 2011
Available online 12 August 2011

Keywords:
Financial distress prediction
Multiple prediction methods
Rough set
Dempster–Shafer evidence theory
Weight

ABSTRACT

It is critical to build an effective prediction model to improve the accuracy of financial distress prediction. Some existing literatures have demonstrated that single classifier has limitations and combination of multiple prediction methods has advantages in financial distress prediction. In this paper, we extend the research of multiple predictions to integrate with rough set and Dempster–Shafer evidence theory. We use rough set to determine the weight of each single prediction method and utilize Dempster–Shafer evidence theory method as the combination method. We discuss the research process for the financial distress prediction based on the proposed method. Finally, we provide an empirical experiment with Chinese listed companies' real data to demonstrate the accuracy of the proposed method. We find that the performance of the proposed method is superior to those of single classifier and other multiple classifiers.

1. Introduction

Financial distress of a company not only makes the company suffer a huge economic loss but also makes investors and government suffer great economic loss. So it is important to build prediction models with high accuracy and effectiveness in financial distress prediction. Consequently, how to predict companies' financial distress effectively and timely has become a hot research topic in academic world.

Researchers in western countries have begun to study financial distress since 1966. Beaver [1] was the first one who applied a univariate model on financial ratios to predict corporate bankruptcy. Altman [2] proposed the method of multiple discriminant analysis (MDA) to perform bankruptcy prediction. He concluded that corporate bankruptcy could be explained by five financial ratios. Ohlson [3] introduced a logistic regression (Logit) model to predict financial distress. Later, Zmijewski [4] proposed a new financial distress prediction method of Probit.

In recent several decades, various artificial intelligent methods were employed to predict financial distress. Frydman et al. [5] used a method of decision tree for financial distress prediction. Tam [6] presented a neural network (NN) approach to bank failures prediction and showed the proposed method was an effective method for evaluating the financial condition of a bank. Salchenberger et al. [7] used a neural network model to achieve a higher degree of prediction accuracy of financial distress. Sun and Shenoy [8] applied Bayesian networks for bankruptcy prediction. Hu [9] presented a novel multi-player perceptron approach by using a non-additive decision making method to the financial distress prediction. Bryant [10] established a case-based reasoning (CBR) bankruptcy prediction system and compared the prediction accuracy with Logit model. Park and Han [11] proposed an analogical reasoning structure for feature weighting. They applied the proposed approach to bankruptcy prediction and showed good results. Li and Sun [12] applied CBR to financial distress prediction with financial ratios. Varetto [13] used genetic algorithm to analyze the insolvency risk. Shin and Lee [14] applied genetic algorithm (GA) to bankruptcy prediction and illustrated how GA can be applied to bankruptcy prediction. Their results showed that using GA for bankruptcy prediction was promising. Dimitras et al. [15] and Mckee [16] used rough set theory for the prediction of business failure and bankruptcy respectively. Kaski et al. [17] used self-organizing map (SOM) to explore financial statements of enterprises. Min and Lee [18] applied support vector machine (SVM) for the bankruptcy prediction problem. They compared its accuracy with those of MDA, Logit and back propagation NN (BPNN). The results showed that SVM outperformed those methods. Shin et al. [19] investigated the efficacy of applying SVM to bankruptcy prediction problem and showed the superior of SVM method to BPNN in the bankruptcy prediction. Ko and Lin [20] introduced an evolutionary approach with modularized evaluation functions to forecast financial distress. Ravisankar and Ravi [21] used group method of data handling (GMDH), counter

* Corresponding author.
E-mail address: xiangleiy@126.com (X. Yang).

0950-7051/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.knosys.2011.08.001

propagation neural network (CPNN) and fuzzy adaptive resonance theory map (fuzzy ARTMAP) to study the financial distress prediction for banks. They found that the GMDH method performed better than other two methods and the methods in previous studies used in their paper. Sun and Li [22] implemented a data mining method for listed companies' financial distress prediction. Later, they [23] indicated that existing methods had disadvantages in dealing with dynamic sample data. Therefore, they explored a new model of dynamic financial distress prediction based on longitudinal data streams. Their results showed that the dynamic model had better performance than static models.

Up to now, much attention has been paid to financial distress prediction methods based on a single method or combination of two methods, and few literatures paid attention to combination of multiple methods. However, some existing literatures have showed that single classifier had limitations and multiple classifiers improved the prediction accuracy in the financial distress prediction. Jo and Han [24] proposed a new hybrid model based on CBR, NN and MDA. The new model achieved higher prediction accuracy than each individual model used. Lin and McClean [25] demonstrated that a combined model of several classification models yielded higher prediction accuracy than those of individual classifiers. Sun and Li [26,27] used combination of multiple classifiers for financial distress prediction. They found that the method based on the combination of multiple classifiers could largely improve the average prediction accuracy and stability by giving an empirical experiment with Chinese listed companies' real world data. Also the empirical experiment indicated that the method based on combination of multiple classifiers was more suitable for the financial distress prediction than the single classifiers. Nanni and Lumini [28] investigated the performance of several systems based on ensemble of classifiers. They found that the ensemble system of Random Subspace performed better than all other models that were used in the bankruptcy prediction.

Therefore, the advantages of combination of multiple methods can not be neglected. Some literatures [29,30] in other fields also demonstrated that combination of multiple methods can reduce the variance of estimated error and hence improve the whole recognition performance.

It is of great importance to select an effective combination method in using combination of multiple methods. Existing methods commonly take the well-known methods including majority voting, Borda count, Bayesian, behavior-knowledge space (BKS) as the combination methods. Besides, some literatures took neural networks, fuzzy algorithm as combination methods [31–33]. Majority voting method is intuitively simple and needs no extra memory. But it has a limitation that all classifiers are treated equally regardless of the characteristic of each classifier [34]. The same limitation also exists in the Borda count method [34]. The Bayesian method has two major limitations [34,35]. One is that it requires the assumption of mutual independencies among multiple classifiers, which is usually not satisfied in real applications. The other one is that it can not model imprecision about uncertain measurements. Although BKS dose not require independence of individual classifiers, its space complexity is prohibitively high. Furthermore, the generalization performance of BKS becomes poorly when the class distribution is not uniform [36]. In this paper, we take Dempster–Shafer (D-S) evidence theory as combination method for the financial distress prediction. D-S evidence theory, which is a powerful method to combine information [37], is an important combination method in multiple classifiers. D-S evidence theory has advantages in dealing with imprecision and uncertain information. It also has advantages in dealing with any union of classes, which is useful to deal with classification problems objectively. For classification problems, D-S evidence theory can treat individual classifiers differently according to the evidences that have been collected, and it does not require prior probability. It can effectively distinguish between "unknown" and "uncertainty" for classification problems. It can reduce the size of hypothesis set according to the information collected, thus it decreases the complexity of knowledge space and hence computation complexity. In addition, some other literatures demonstrated the advantages of D-S evidence theory for dealing with classification problems. Xu et al. [38] used D-S evidence theory to combine individual methods with an application on handwriting recognition. They compared it with three types of combination methods, which were Bayesian formalism, voting principle and neural networks. They found that the combination method based on D-S evidence theory could obtain high recognition and reliability rates as well as high robustness. Therefore, the combination method based on D-S evidence theory was taken as their first recommendation.

Up to now, although D-S evidence theory has been applied in the financial distress prediction [39], the benefit of application of D-S evidence theory in multiple predictions for financial distress has been neglected. Therefore, in this paper, we propose a research model for listed companies' financial distress based on multiple prediction methods by taking D-S evidence theory as combination method. The distinction between this paper and [38] is that we use different decision rule and we incorporate with rough set to weight significance of each single prediction method.

Different single prediction methods have different individual performance for a specific problem and each single prediction method has its own uncertainty. Therefore, different prediction methods play different roles in financial distress prediction. All prediction methods can not be treated equally. D-S evidence theory can measure the significance of single prediction methods according to the collected evidences. However, it is not enough to use D-S evidence theory to measure the significance of single prediction methods. Conventional D-S evidence theory has an unavoidable disadvantage that it performs poorly when fusing high conflict information [40]. By reassigning weight factors before fusing, Sun et al. [40] reduced the influence of conflict evidence and got more reasonable fusing results than conventional D-S evidence theory. To measure the significance of single evidence more accurately, Liu et al. [41] proposed a new method of weighted D-S evidence combination. They also found that the weighted D-S evidence combination had more precision of fusion than D-S evidence combination in the field of the instance of diesel engine state evaluation. To measure the significance of single prediction method more accurately, in this paper we also give different weights to different prediction methods before using D-S evidence theory to combine outputs of single predictions. The well-known methods of determining weights include methods of expert assessment, correlation analysis, minimum sum of errors, fuzzy sets, grey theory, neural networks, genetic algorithm and wavelet analysis. The disadvantage of expert assessment and fuzzy sets is that the weight depends on empirical knowledge of experts. Thus, the result is subjective. The methods of correlation and minimum sum of errors are based on the theory of statistics. When they have a greater number of single prediction models, they would have a larger amount of calculation even sometimes need the approximate calculation by numerical computing methods. The methods based on grey theory, neural networks, genetic algorithm and wavelet analysis usually need solve minimum sum of errors [42]. In this paper, we use rough set to determine weight of each single prediction method. Rough set theory can effectively measure the degree of significance of single prediction method according to the outputs of single prediction without needing extra prior information. Therefore, it makes the results of determining weights more objectively. And it avoids solving linear or nonlinear extremum problem and hence avoids a large amount of calculation. Rough set has been widely applied into determining weights [42–44].

We firstly determine weight of each single prediction method according to outputs of single prediction method of training samples by using rough set. Then we combine the prediction results of single prediction method using combination method of D-S evidence theory. Finally, we give an empirical example to demonstrate the efficiency of the proposed method using data of Chinese listed companies.

The rest of this paper is organized as follows. Section 2 gives a brief introduction of rough set and D-S evidence theory. Section 3 presents the prediction process based on multiple prediction methods, rough set and evidence theory and gives a simple example to explain the research process of the proposed method. In Section 4, we provide an empirical example of financial distress prediction with data of Chinese listed companies and compares the proposed model with Logit, NN, SVM, multiple classifiers based on majority voting, multiple classifiers based on Bayesian combination method and multiple classifiers based on BKS. We conclude and discuss possible future work in Section 5.

2. Brief introduction to rough set and D-S evidence theory

2.1. Rough set

Rough set theory, which was firstly proposed by Pawlak [45], is a good method to deal with problem containing uncertain information. Rough set theory has been successfully applied in many fields such as machine learning, pattern recognition, decision analysis and knowledge discovery.

Consider an information system denoted as a pair $\Re = (U, C \cup D)$, where $U = \{u_1, \ldots, u_t\}$ is a nonempty finite set of objects, $C = \{c_1, \ldots, c_m\}$ is a nonempty finite set of condition attributes, and $D = \{d_1, \ldots, d_n\}$ is a nonempty finite set of decision attributes. The equivalence relationship is defined as follows.

$$R_C = \{(x,y) \in U \times U | c_j(x) = c_j(y),\ \forall c_j \in C\} \tag{1}$$

$$R_{c_j} = \{(x,y) \in U \times U | c_i(x) = c_j(y),\ \forall c_i \in C, c_i \neq c_j\},\quad j = 1, \ldots, m \tag{2}$$

$$R_D = \{(x,y) \in U \times U | d_j(x) = d_j(y),\ \forall d_j \in D\}. \tag{3}$$

It is obvious that R_C, R_{c_j} and R_D are all equivalence relationship on U. U/R_C and U/R_{c_j} are called knowledge systems on the basis of condition attributes. And U/R_D is called knowledge system on the basis of decision attributes.

Definition 1. The dependence of D to C is defined as:

$$H(R_D/R_C) = -\sum_{[x]\in U/R_C} p[x] \sum_{[y]\in U/R_D} p([y]/[x]) \log(p([y]/[x])) \tag{4}$$

where $p[x] = \frac{card[x]}{card[U]}, p([y]/[x]) = \frac{card([y]\cap[x])}{card[x]}$. The value of $H(R_D/R_C)$ is the dependence of D to C.

Definition 2. The significance of C_j can be defined as follow.

$$\omega(c_j, C, D) = |H(R_D/R_{c_j}) - H(R_D/R_C)|, \tag{5}$$

$$H(R_D/R_{C_j}) = -\sum_{[x]\in U/R_{C_j}} p[x] \sum_{[y]\in U/R_D} p([y]/[x]) \log(p([y]/[x])),\quad j = 1, \ldots, m. \tag{6}$$

Definition 3. The weight of condition attribute is defined as follow.

$$w_j = \frac{\omega(c_j, C, D)}{\sum_{j=1}^{m} \omega(c_j, C, D)} \tag{7}$$

It is noted that all of Definitions 1–3 are based on the theory of entropy.

2.2. D-S evidence theory

D-S evidence theory which originated from the upper and lower probabilities by Dempster [37] is a new important reasoning method and is a powerful method to combine accumulative evidence of changing prior opinions in the light of new evidences. Because of its power to synthesize information, D-S evidence theory has been applied in many fields such as fault diagnosis [46], multi-class classification [47] and supplier selection [48]. Also it has been applied in intuitionistic fuzzy sets to solve multiple criteria decision making problem [49]. Recently D-S evidence theory is frequently used in the fusion of decision-making layer [50,51].

2.2.1. Basic probability assignment

Let Θ be a finite nonempty set of mutually exclusive alternatives, and be called the frame of discernment. For any proposition A in any problem, they all belong to the power set 2^{Θ}. On 2^{Θ} we can define the basic probability assignment function (BPAF) m: $2^{\Theta} \to [0,1]$, such that m satisfies: $m(\Phi) = 0$ and $\sum_{A\subseteq\Theta} m(A) = 1$, where Φ is the empty set.

$m(A)$ is called basic probability assignment of proposition A. If $m(A) > 0$, the subset A is called focal element.

2.2.2. Belief function

The belief function of proposition A, denoted as $Bel(A)$, is defined as follows:

$$Bel(A) = \sum_{B\subseteq A} m(B),\quad \forall A \subseteq \Theta. \tag{8}$$

The belief function represents the minimal support of A, which means the total trust of proposition A.

2.2.3. Plausibility function

The plausibility function can be defined as follows:

$$Pl(A) = 1 - Bel(\bar{A}) = \sum_{B\cap A\neq\Phi} m(B),\quad \forall A \subseteq \Theta. \tag{9}$$

The plausibility function represents the maximal support of A.

2.2.4. D-S rule of combination

Let Bel_1 and Bel_2 be two belief functions, and m_1 and m_2 be the corresponding basic probability assignment functions. The evidences are $A_1, A_2, \ldots, A_m$ with respect to m_1 and $B_1, B_2, \ldots, B_n$ with respect to m_2. If $\sum_{A_i\cap B_j=\Phi} m_1(A_i)m_2(B_j) < 1$, we have

$$m(A) = m_1 \oplus m_2(A) = \frac{1}{1-K} \sum_{A_i\cap B_j=A} m_1(A_i)m_2(B_j),\quad \forall A \subseteq \Theta,\quad A \neq \Phi;$$

$$m(A) = m_1 \oplus m_2(A) = 0,\quad A = \Phi. \tag{10}$$

where

$$K = \sum_{A_i\cap B_j=\Phi} m_1(A_i)m_2(B_j).$$

K, which reflects the conflict between the evidences, is called the conflict probability. Coefficient $\frac{1}{1-K}$ is called normalized factor.

3. Research process

In this section, we will present the research process for financial distress prediction. We first input financial ratios to each of single prediction methods. And then we calculate the weight of each single prediction method according to the outputs of single prediction based on the training samples. Finally, we use D-S evidence theory to combine the outputs of each single prediction method. Fig. 1 demonstrates the research process for financial distress prediction proposed in this paper.

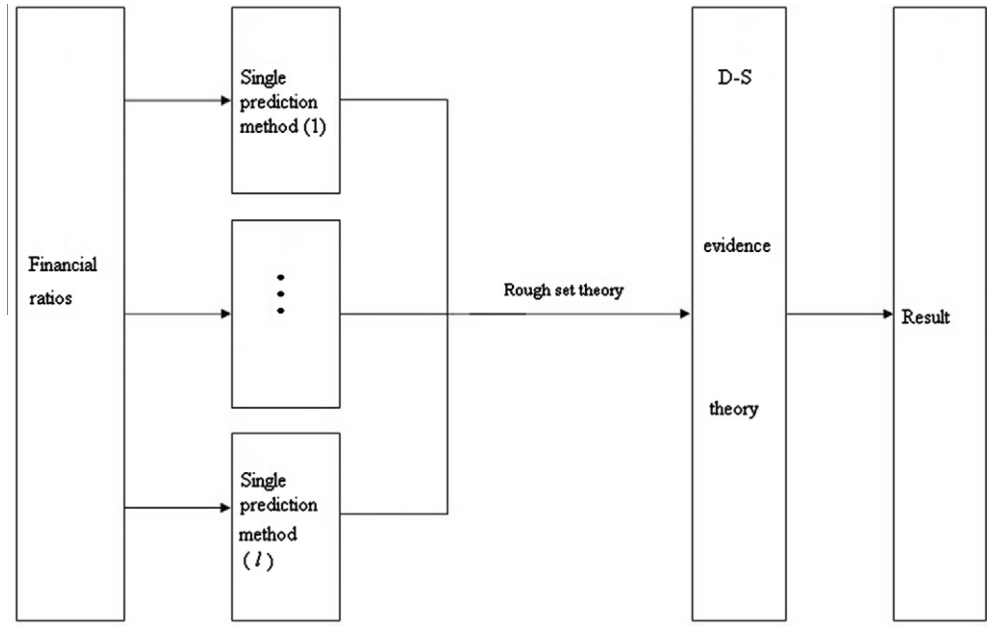

Fig. 1. Process of using multiple prediction methods for financial distress prediction with rough set and D-S evidence theory.

More specified, the steps of financial distress prediction based on the proposed method are described as follows.

Step 1. Predict using single methods.

The data obtained from the real world are usually different from each other in unit and scale due to the criteria. Therefore, it is of great importance to normalize the data to eliminate the difference before applying the data to single prediction. The function of normalization is defined as follow.

$$x'_{ij} = \frac{x_{ij} - \min_j}{\max_j - \min_j} \tag{11}$$

where x_{ij} means the attribute value of the variable j for the ith company. $\max_j$ and $\min_j$ represent the maximal value and minimal value of the variable j cross all companies respectively.

Let $y_{method(j)}(A_i)$ represents the output of single prediction. Where $method(j)$ represents single prediction method j, and A_i represents the ith company. Therefore, $y_{method(j)}(A_i)$ represents prediction value of company i using the single prediction method j. The matrix form of outputs is as follow.

$$\begin{bmatrix} y_{method(1)}(A_1) & y_{method(1)}(A_2) & \cdots & y_{method(1)}(A_n) \\ y_{method(2)}(A_1) & y_{method(2)}(A_2) & \cdots & y_{method(2)}(A_n) \\ \vdots & \vdots & \vdots & \vdots \\ y_{method(l)}(A_1) & y_{method(l)}(A_2) & \cdots & y_{method(l)}(A_n) \end{bmatrix} \tag{12}$$

where n is the number of companies and l represents the number of single prediction methods we use.

To satisfy the requirement of D-S rule of combination and to achieve more accuracy results of financial distress prediction, we have all of the values of outputs be between 0 and 1. That is $y_{method(j)}(A_i) \in [0,1]$.

Step 2. Determine the weight of each single prediction method we use.

Before using rough set to determine weights, we need to draw the characteristic of each attribute with the corresponding output in order to calculate the degree dependence of each decision attribute to condition attributes. There are four common methods for characterization which are Equal distance quartile, Naïve Scaler algorithm, SsmiNaive Scaler algorithm and Boolean calculation.

For the financial distress prediction problem, $U = \{A_1, \ldots, A_n\}$, $C = \{method(1), \ldots, method(l)\}$ and $D = \{d\}$, where d is the decision attribute that diagnoses whether one company is in financial distress or not. If the ith company is in financial distress, $d(A_i) = 1$ and 0 otherwise. We can construct the knowledge system as Table 1. In Table 1, $C'_{method(j)}(A_i)$ is the value after characterizing $y_{method(j)}(A_i)$.

According to the Eqs. (4)–(6), we can obtain

$$\omega(method(j), C, D) = |H(R_D/R_{method(j)}) - H(R_D/R_C)|$$

Therefore, the weight of method j can be calculated as follow.

$$w_{method(j)} = \frac{\omega(method(j), C, D)}{\sum_{j=1}^{l} \omega(method(j), C, D)}. \tag{13}$$

The future financial condition of companies is unknown. But we can know whether the current condition of companies is in financial distress or not. Therefore, in this step, $d(A_i)$ comes from training samples and $y_{method(j)}(A_i)$ comes from outputs of training samples. In this step, sample size n represents the number of companies of training sample.

Step 3. Combine the outputs of single prediction method.

Before using D-S evidence theory to combine the outputs of single prediction method, we need to transform the outputs into basic probability assignments. It is implemented by function (14) and function (15).

$$m_{method(j)}(A_i) = \frac{w_{method(j)} \cdot y_{method(j)}(A_i)}{\sum_{i=1}^{n} w_{method(j)} \cdot y_{method(j)}(A_i) + 1} \tag{14}$$

$$m_{method(j)}(\Theta) = \frac{1}{\sum_{i=1}^{n} w_{method(j)} \cdot y_{method(j)}(A_i) + 1} \tag{15}$$

Table 1
Knowledge system for financial distress prediction.

U	$method(1)$	$\cdots$	$method(l)$	d
A_1	$C'_{method(1)}(A_1)$	$\cdots$	$C'_{method(l)}(A_1)$	$d(A_1)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_n	$C'_{method(1)}(A_n)$	$\cdots$	$C'_{method(l)}(A_n)$	$d(A_n)$

where n represents the number of companies, which can be the number of training samples or the number of testing samples, and A_i represents the ith company. $y_{method(j)}(A_i)$ represents the outputs of the ith company by using single prediction method j. $w_{method(j)}$ is the weight of prediction method j,which is calculated by Eq. (13) in step 2.

The matrix form of basic probability assignments is presented below.

$$\begin{bmatrix} m_{method(1)}(A_1) & m_{method(1)}(A_2) & \cdots & m_{method(1)}(A_n) & m_{method(1)}(\Theta) \\ m_{method(2)}(A_1) & m_{method(2)}(A_2) & \cdots & m_{method(2)}(A_n) & m_{method(2)}(\Theta) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ m_{method(l)}(A_1) & m_{method(l)}(A_2) & \cdots & m_{mehod(l)}(A_n) & m_{method(l)}(\Theta) \end{bmatrix} \tag{16}$$

After transforming outputs into basic probability assignments, we combine the basic probability assignments using D-S rule of combination according to Eq. (10). Then a final basic probability assignment $m(A_i)$ for the i th company is obtained. They are $m(A_1), m(A_2)\ldots m(A_n)$.

Step 4. Make decision.

Firstly, convert $m(A_i)$ into decision-making value by using equation (17). According to the basic probability assignment function (14), we define Eq. (17) as follow.

$$y'(A_i) = m(A_i)\cdot\left(\sum_{i=1}^{n} y'(A_i) + 1\right), \quad i = 1, 2, \ldots, n \tag{17}$$

In matrix form it can be expressed as:

$$\begin{bmatrix} 1-m(A_1) & -m(A_1) & \cdots & -m(A_1) \\ -m(A_2) & 1-m(A_2) & \cdots & -m(A_2) \\ \vdots & \vdots & \vdots & \vdots \\ -m(A_n) & -m(A_n) & \cdots & 1-m(A_n) \end{bmatrix} \begin{bmatrix} y'(A_1) \\ y'(A_2) \\ \vdots \\ y'(A_n) \end{bmatrix} = \begin{bmatrix} m(A_1) \\ m(A_2) \\ \vdots \\ m(A_n) \end{bmatrix} \tag{18}$$

According to Eq. (18), we can calculate $y'(A_i)$.

Then a criteria needs to be set to diagnose whether the companies are in financial distress or not. Suppose CR is the criteria.

If $y'(A_i) \geqslant CR$, we say that the ith company is in financial distress ($d(A_i) = 1$).

If $y'(A_i) < CR$, we say that the ith company is not in financial distress ($d(A_i) = 0$).

Obviously, if the Eq. (18) has more than one solution or has no solution, we may have difficulty to take a prediction. For example, suppose that $CR = 0.5$ and the solution of $y'(A_i)$ has two values 0.8 and 0.1, then we are not able to make decision on company A_i. Fortunately, such scenarios would not happen. We can prove that the Eq. (18) has a unique solution.

Proposition 1. Eq. (18) *has a unique solution.*

Proof. Eq. (18) is of the form $AX = F$,

$$\text{where } A = \begin{bmatrix} 1-m(A_1) & -m(A_1) & \cdots & -m(A_1) \\ -m(A_2) & 1-m(A_2) & \cdots & -m(A_2) \\ \vdots & \vdots & \vdots & \vdots \\ -m(A_n) & -m(A_n) & \cdots & 1-m(A_n) \end{bmatrix},$$

$$X = \begin{bmatrix} y'(A_1) \\ y'(A_2) \\ \vdots \\ y'(A_n) \end{bmatrix}, \quad F = \begin{bmatrix} m(A_1) \\ m(A_2) \\ \vdots \\ m(A_n) \end{bmatrix}.$$

A is called coefficient matrix of X. It is well known that if $|A| \neq 0$ then Eq. (18) has a unique solution. We have

$$|A| = \begin{vmatrix} 1-m(A_1) & -m(A_1) & \cdots & -m(A_1) \\ -m(A_2) & 1-m(A_2) & \cdots & -m(A_2) \\ \vdots & \vdots & \vdots & \vdots \\ -m(A_n) & -m(A_n) & \cdots & 1-m(A_n) \end{vmatrix} \tag{19}$$

$$= \begin{vmatrix} 1-\sum_{i=1}^{n} m(A_i) & 1-\sum_{i=1}^{n} m(A_i) & \cdots & 1-\sum_{i=1}^{n} m(A_i) \\ -m(A_2) & 1-m(A_2) & \cdots & -m(A_2) \\ \vdots & \vdots & \vdots & \vdots \\ -m(A_n) & -m(A_n) & \cdots & 1-m(A_n) \end{vmatrix} \tag{20}$$

$$= 1-\sum_{i=1}^{n} m(A_i) \begin{vmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{vmatrix} = 1-\sum_{i=1}^{n} m(A_i) \tag{21}$$

By the function (15), we have $m_{method(j)}(\Theta) \neq 0$ for $j = 1, \ldots, l$.

Then $\sum_{i=1}^{n} m(A_i) = 1 - m(\Theta) \neq 1$. And thus $|A| = 1 - \sum_{i=1}^{n} m(A_i) \neq 0$, which yields the result.

Here, a simple example is given to describe the research process. □

Example 1. Let us consider ten companies. The financial condition of ten companies is given in Table 2. In Table 2, "1" represents the company being in financial distress, and "0" represents the company which is not in financial distress. Three prediction methods are used to predict financial distress. The value of financial ratios and the step of single prediction are not given in this example.

Table 2
Condition of ten companies.

Companies	Actual financial condition
1	1
2	1
3	0
4	1
5	0
6	0
7	1
8	1
9	1
10	0

Table 3
Outputs of single prediction.

Companies	Method 1	Method 2	Method 3
1	0.65	0.48	0.76
2	0.53	0.62	0.88
3	0.42	0.67	0.78
4	0.92	0.71	0.82
5	0.14	0.25	0.28
6	0.16	0.21	0.15
7	0.34	0.46	0.39
8	0.49	0.41	0.43
9	0.59	0.65	0.72
10	0.12	0.26	0.31

Table 4
Knowledge system after characterizing the outputs of single prediction.

Companies	Method 1	Method 2	Method 3	Actual financial condition (d)
1	1	0	1	1
2	1	1	1	1
3	0	1	1	0
4	1	1	1	1
5	0	0	0	0
6	0	0	0	0
7	0	0	0	1

Table 5
Actual financial condition and prediction financial condition.

Companies	Actual financial condition	Prediction financial condition
1	1	1
2	1	1
3	0	0
4	1	1
5	0	0
6	0	0
7	1	0
8	1	0
9	1	1
10	0	0

The sample of ten companies is categorized into a training sample and a testing sample. Companies 1–7 are used as the training sample, and companies 8–10 are used as the testing sample. The outputs of single prediction by using three methods are listed in Table 3.

In Step 2, by characterizing the outputs of single prediction of the training sample, we can construct knowledge system that is given in Table 4. Then we can get the following results.

$H(R_D/R_C) = 0.2728.$
$H(R_D/R_{method(1)}) = 0.5456, \quad H(R_D/R_{method(2)}) = 0.2728,$
$H(R_D/R_{method(3)}) = 0.2728.$
$\omega(method(1), C, D) = 0.2728, \quad \omega(method(2), C, D) = 0,$
$\omega(method(3), C, D) = 0.$
$w_{method(1)} = 1, \quad w_{method(2)} = 0, \quad w_{method(3)} = 0.$

In Step 3, by the functions (14) and (15), we can get the matrix form of basic probability assignments as follow.

$$\begin{bmatrix} m_{method(1)}(A_1) & \cdots & m_{method(1)}(A_7) & m_{method(1)}(\Theta) \\ m_{method(2)}(A_1) & \cdots & m_{method(2)}(A_7) & m_{method(2)}(\Theta) \\ m_{method(3)}(A_1) & \cdots & m_{method(3)}(A_7) & m_{method(3)}(\Theta) \end{bmatrix} = \begin{bmatrix} 0.1563 & 0.1274 & 0.1010 & 0.2212 & 0.0337 & 0.0385 & 0.0817 & 0.2404 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$
$$\begin{bmatrix} m_{method(1)}(A_8) & m_{method(1)}(A_9) & m_{method(1)}(A_{10}) & m_{method(1)}(\Theta) \\ m_{method(2)}(A_8) & m_{method(2)}(A_9) & m_{method(2)}(A_{10}) & m_{method(2)}(\Theta) \\ m_{method(3)}(A_8) & m_{method(3)}(A_9) & m_{method(3)}(A_{10}) & m_{method(3)}(\Theta) \end{bmatrix} = \begin{bmatrix} 0.2227 & 0.2682 & 0.0545 & 0.4545 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The results of combining are:

$m(A_1) = 0.1563, \quad m(A_2) = 0.1274, \quad m(A_3) = 0.1010,$
$m(A_4) = 0.2212, \quad m(A_5) = 0.0337, \quad m(A_6) = 0.0385,$
$m(A_7) = 0.0817, \quad m(A_8) = 0.2227, \quad m(A_9) = 0.2682,$
$m(A_{10}) = 0.0545.$

In Step 4, according to the Eq. (17) we have

$y'(A_1) = 0.6507, \quad y'(A_2) = 0.5304, \quad y'(A_3) = 0.4205,$
$y'(A_4) = 0.9209, \quad y'(A_5) = 0.1403, \quad y'(A_6) = 0.1603,$
$y'(A_7) = 0.3401, \quad y'(A_8) = 0.4899, \quad y'(A_9) = 0.5900,$
$y'(A_{10}) = 0.1199.$

In this example, we set the criteria $CR = 0.5$. That means if $y'(A_i) \geqslant 0.5$, company i is in financial distress. If $y'(A_i) < 0.5$, company i is not in financial distress. The actual financial condition and prediction financial condition of companies in this example are listed in Table 5.

Only company 7 and company 8 are predicted incorrectly.

4. Empirical experiment

4.1. Samples

In this study, according to the benchmark whether the listed company is specially treated (ST) by China Securities Supervision and Management Committee (CSSMC), we categorize Chinese listed companies into two classes, which are companies in health situation and companies in financial distress. If one company is specially treated, it is a company in financial distress. Otherwise, it is a healthy company. Chinese listed companies are specially treated by CSSMC if they have had negative net profit in recent continuous two years or they have purposely published financial statements with serious false and misstatement. In our study, we consider ST companies as the companies that have had negative net profit in recent continuous two years.

We randomly collect 276 samples listed in Shenzhen Stock Exchange and Shanghai Stock Exchange. Among them, two thirds are healthy companies while other one third ST companies. By eliminating the sample companies in case of missing financial ratios data, the final samples include 92 ST companies that were specially treated between 2007 and 2009 and 161 healthy companies during the same period. There is no proportion between the remaining ST companies and the remaining healthy companies. Ohlson [3] indicated that there was no appreciate criteria to be used to match distressed companies and healthy companies. Later, some researchers also did not use matched sample [5,52,53]. Therefore, we did not add ST companies or healthy companies to a certain proportion any more. Thirty-nine financial ratios (see Table 6.) covering profit ability, debt ability, activity ability, growth ability, cash ability and shareholder profit ability are selected as initial features.

For Chinese listed companies' financial distress prediction, some literatures used data at year ($t-2$), which represents two years before, to predict financial distress at year ($t-0$) [26,27].

Table 6
Definition of variables.

Variables	Definition
X1	Net operating income rate
X2	Net profit margin of total assets
X3	Return on equity
X4	Return on total assets
X5	Return on invested capital
X6	Operating margin
X7	Profit margin
X8	Current ratio
X9	Quick ratio
X10	Cash ratio
X11	Asset-liability ratio
X12	Tangible net debt ratio
X13	Working capital ratio
X14	Working capital to total assets ratio
X15	Working capital to net assets ratio
X16	Equity ratio
X17	Long-term debt ratio
X18	Equity to liability ratio
X19	Interest coverage ratio
X20	Account receivable turnover
X21	Account payable turnover
X22	Inventories turnover
X23	Current assets turnover
X24	Fixed assets turnover
X25	Total assets turnover
X26	Working capital turnover
X27	Sales growth rate of major operation
X28	Growth ratio of net profit
X29	Capital maintenance and appreciation
X30	Growth ratio of total assets
X31	Cash flow to current liability
X32	Cash to main business income ratio
X33	Net operating cash flow per share
X34	Net cash flow of investing activities per share
X35	Cash ratio to sales
X36	Operating income per share
X37	Earnings per share
X38	Net assets per share
X39	Price-to-book ratio

And some other literatures used data at year $(t-3)$ to predict financial distress at year $(t-0)$ [54]. In this paper, we use data at year $(t-2)$ and data at year $(t-3)$ to predict financial distress at year $(t-0)$ respectively. All of the data used in this study are obtained from China Stock Market & Accounting Research (CSMAR) database.

4.2. Selection of single prediction method

Single prediction methods play an important role in the performance of multiple prediction methods. Every single prediction method has its own advantages and disadvantages. On the one hand, we expect to utilize advantages of all of single prediction methods. On the other hand, too many methods will increase the multiple prediction system's complexity. Meanwhile, the proposed method in this paper requires the prediction values of single prediction methods in the range of $[0,1]$ rather than classification labels $\{0,1\}$ (See Step 1.). Therefore, some classification methods which have been applied for financial distress prediction can not be selected such as decision trees (DT) and MDA. Logistic regression, neural network (NN) and support vector regression (SVR) are good methods for prediction. And they have been demonstrated to be good methods for financial distress prediction [3,6,7,18,19]. To achieve more prediction accuracy and reduce the complexity of prediction, in this paper, we take logistic regression, neural network and support vector regression as basic prediction methods.

4.3. Experiment and comparison

4.3.1. Experiment

To evaluate the prediction accuracy and stability of the proposed method in this paper, we use 10- folder cross-validation to perform our experiment. The samples collected are randomly divided into ten datasets. On each validation, we compare the proposed method with other methods. The maximum and minimum values of each variable are listed in Tables 7 and 8.

t-test and stepwise logistic regression are used to reduce features from 39 initial financial ratios in Table 6 using software SPSS 16.0 based on training datasets. We first use t-test to remove some

Table 7
Max and Min values of variables $X1-X29$.

Variables	Values at year $(t-2)$	Values at year $(t-3)$
X1	*Max* = 0.604626	*Max* = 0.479729
	Min = −4.846423	*Min* = −2.147677
X2	*Max* = 0.325284	*Max* = 0.190511
	Min = −0.620457	*Min* = −0.154485
X3	*Max* = 0.550259	*Max* = 0.729280
	Min = −1.538052	*Min* = −0.997526
X4	*Max* = 0.501713	*Max* = 0.266862
	Min = −0.524037	*Min* = −0.149110
X5	*Max* = 0.725510	*Max* = 0.544860
	Min = −1.498759	*Min* = −0.444843
X6	*Max* = 0.677344	*Max* = 0.545405
	Min = −3.564519	*Min* = −2.091958
X7	*Max* = 4.175723	*Max* = 1.384188
	Min = −2.353363	*Min* = 0.463575
X8	*Max* = 29.85033	*Max* = 6.708822
	Min = 0.267057	*Min* = 0.251585
X9	*Max* = 25.175949	*Max* = 6.011356
	Min = 0.045613	*Min* = 0.064571
X10	*Max* = 2.059608	*Max* = 0.000000
	Min = 0.000000	*Min* = 0.000000
X11	*Max* = 0.935909	*Max* = 0.914882
	Min = 0.009122	*Min* = 0.109977
X12	*Max* = 23.175897	*Max* = 11.378543
	Min = 0.011628	*Min* = −492.337748
X13	*Max* = 0.966500	*Max* = 0.850943
	Min = −2.744513	*Min* = −2.974794
X14	*Max* = 0.755015	*Max* = 0.657885
	Min = −0.552717	*Min* = −0.479246
X15	*Max* = 1.735940	*Max* = 2.165512
	Min = −5.120481	*Min* = −5.269909
X16	*Max* = 0.990878	*Max* = 0.890023
	Min = 0.064091	*Min* = 0.085118
X17	*Max* = 0.837292	*Max* = 0.805384
	Min = 0.000000	*Min* = 0.000000
X18	*Max* = 108.619697	*Max* = 8.092819
	Min = 0.068479	*Min* = 0.093037
X19	*Max* = 217.224049	*Max* = 300.000000
	Min = −209.381408	*Min* = −146.282660
X20	*Max* = 404.759402	*Max* = 214.412132
	Min = 0.785612	*Min* = 0.510858
X21	*Max* = 61.74442	*Max* = 30.947160
	Min = 0.167578	*Min* = 0.929975
X22	*Max* = 675.652448	*Max* = 219.886561
	Min = 0.023957	*Min* = 0.083881
X23	*Max* = 9.598082	*Max* = 11.173971
	Min = 0.024046	*Min* = 0.066029
X24	*Max* = 69.187856	*Max* = 35.359740
	Min = 0.158047	*Min* = 0.107633
X25	*Max* = 2.884861	*Max* = 2.754135
	Min = 0.015332	*Min* = 0.038685
X26	*Max* = 317.651402	*Max* = 487.615776
	Min = −373.043107	*Min* = −271.965189
X27	*Max* = 7.670736	*Max* = 16.275289
	Min = −0.912951	*Min* = −0.754743
X28	*Max* = 394.669323	*Max* = 221.108103
	Min = −303.312789	*Min* = −49.980694
X29	*Max* = 3.162923	*Max* = 3.211478
	Min = 0.164317	*Min* = 0.359860

Table 8
Max and Min value of variables $X30-X39$.

Variables	Values at year $(t-2)$	Values at year $(t-3)$
$X30$	$Max = 2.604411$	$Max = 1.004208$
	$Min = -0.558090$	$Min = -0.410402$
$X31$	$Max = 0.800984$	$Max = 2.218627$
	$Min = -10.676984$	$Min = -0.642190$
$X32$	$Max = 1.554536$	$Max = 10.557602$
	$Min = -4.371405$	$Min = -0.975485$
$X33$	$Max = 3.887462$	$Max = 2.144374$
	$Min = -2.385782$	$Min = -0.940413$
$X34$	$Max = 1.153775$	$Max = 0.547787$
	$Min = -3.539853$	$Min = -3.806722$
$X35$	$Max = 91.468199$	$Max = 78.722321$
	$Min = -16.275194$	$Min = -262.444061$
$X36$	$Max = 46.937359$	$Max = 51.541205$
	$Min = 0.043860$	$Min = 0.111267$
$X37$	$Max = 2.984278$	$Max = 1.712317$
	$Min = -2.113392$	$Min = -1.537526$
$X38$	$Max = 15.674792$	$Max = 6.698431$
	$Min = 0.000000$	$Min = 0.118087$
$X39$	$Max = 133.571600$	$Max = 59.193650$
	$Min = 0.000000$	$Min = 0.691025$

Table 9
Selected features using data at year $(t-2)$.

Validation	Features
1	Working capital turnover ($X26$), Earnings per share ($X37$)
2	Working capital turnover ($X26$), Earnings per share ($X37$)
3	Inventories turnover ($X22$), Sales growth rate of major operation ($X27$), Earnings per share ($X37$), Price-to-book ratio ($X39$)
4	Quick ratio ($X9$), Working capital turnover ($X26$), Earnings per share ($X37$)
5	Net operating income rate ($X1$), Working capital turnover ($X26$), Earnings per share ($X37$)
6	Working capital turnover ($X26$), Earnings per share ($X37$)
7	Return on equity ($X3$), Quick ratio ($X9$), Working capital to total assets ratio ($X14$), Net operating cash flow per share ($X33$), Operating income per share ($X36$), Earnings per share ($X37$)
8	Working capital turnover ($X26$), Earnings per share ($X37$)
9	Working capital turnover ($X26$), Earnings per share ($X37$)
10	Net profit margin of total assets ($X2$), Return on equity ($X3$), Earnings per share ($X37$)

features on the significance at 5%. Then we use stepwise logistic regression to further reduce the remaining features. The selected features are listed in Tables 9 and 10.

In the first step, the selected features are input into each single prediction method for independent financial distress prediction. In this study, the radial basis function (RBF) is used as the basic kernel function of SVR prediction. Grid-search and cross-validation are used to search for optimal parameters values of RBF based on training datasets. We use BPNN as the NN algorithm. Matlab 7.0 software is used to implement single NN prediction. We run twenty experiments for NN prediction and we select the optimal set of experiment results as the outputs of NN prediction.

Table 10
Selected features using data at year $(t-3)$.

Validation	Features
1	Working capital to net assets ratio ($X15$), Growth ratio of total assets ($X30$), Cash flow to current liability ($X31$)
2	Net profit margin of total assets ($X2$), Return on equity ($X3$), Working capital to net assets ratio ($X15$), Account payable turnover ($X21$), Net operating cash flow per share ($X33$)
3	Return on equity ($X3$), Return on total assets ($X4$), Working capital to net assets ratio ($X15$), Long-term debt ratio ($X17$), Sales growth rate of major operation ($X27$), Cash flow to current liability ($X31$)
4	Return on total assets ($X4$), Asset-liability ratio ($X11$), Net operating cash flow per share ($X33$)
5	Net profit margin of total assets ($X2$), Return on equity ($X3$), Working capital to net assets ratio ($X15$), Account payable turnover ($X21$), Net operating cash flow per share ($X33$)
6	Net profit margin of total assets ($X2$), Return on equity ($X3$), Working capital to net assets ratio ($X15$), Cash to main business income ratio ($X32$)
7	Return on total assets ($X4$), Tangible net debt ratio ($X12$), Sales growth rate of major operation ($X27$), Cash flow to current liability ($X31$)
8	Return on total assets ($X4$), Asset-liability ratio ($X11$), Sales growth rate of major operation ($X27$), Cash to main business income ratio ($X32$), Cash ratio to sales ($X35$)
9	Net profit margin of total assets ($X2$), Return on equity ($X3$), Working capital to net assets ratio ($X15$), Account payable turnover ($X21$), Net operating cash flow per share ($X33$)
10	Return on total assets ($X4$), Asset-liability ratio ($X11$), Equity to liability ratio ($X18$), Cash to main business income ratio ($X32$)

In the second step, we take the Equal distance quartile method for characterization. In this paper, we define that the attribute value of company in financial distress is equal to 1, and the attribute value of healthy company is equal to 0. Hence, the range of attribute values is divided into two intervals which are marked by 0 and 1 respectively.

In this paper, we set 0.5 as the criteria that diagnoses whether one company is in financial distress or not. In other words, if $y'(A_i) \geqslant 0.5$, we say that the ith company is in financial distress. The prediction accuracy of the proposed method in this paper is given in Tables 11 and 12.

4.3.2. Comparison

We compare prediction accuracy of the proposed method with independent logistic regression, SVM classification and NN classification, multiple classifications based on majority voting method, multiple classifications based on Bayesian and multiple classifiers based on BKS. For a two-class problem, the Borda count is equivalent to the majority vote. Therefore, we do not compare the proposed method with multiple classifications based on Borda count. To make the comparison meaningful, we use the same way to deal with the comparative models. Independent SVM classification takes the same basic kernel function as that of proposed method. And we take the same method to search for optimal

Table 11
Results of 10-folder cross-validation using data at year ($t-2$).

Validation	Prediction accuracy of training datasets (%)							Prediction accuracy of testing datasets (%)						
	Logit	NN	SVM	Multiple classifiers based on majority voting	Multiple classifiers based on Bayesian	Multiple classifiers based on BKS	Proposed method in this paper	Logit	NN	SVM	Multiple classifiers based on majority voting	Multiple classifiers based on Bayesian	Multiple classifiers based on BKS	Proposed method in this paper
1	84.10	85.09	88.16	84.65	86.84	87.72	87.28	84.00	80.00	84.00	84.00	88.00	88.00	88.00
2	84.50	82.82	88.55	84.14	87.67	88.11	88.99	80.77	80.77	76.92	80.77	80.77	76.92	84.62
3	82.40	82.02	86.40	85.65	85.96	87.28	86.40	92.00	92.00	92.00	92.00	92.00	92.00	92.00
4	85.90	82.02	90.79	88.16	90.35	91.67	88.60	80.00	80.00	84.00	80.00	84.00	84.00	88.00
5	86.70	83.26	89.43	87.67	87.67	88.99	88.55	73.08	80.77	84.62	76.92	76.92	80.77	84.62
6	84.10	82.89	87.28	85.09	86.40	86.84	88.16	84.00	84.00	88.00	84.00	88.00	92.00	88.00
7	85.80	85.90	88.43	85.90	88.99	89.43	88.11	80.77	84.62	80.77	80.77	80.77	80.77	80.77
8	85.09	87.28	86.84	83.77	84.65	85.96	87.72	92.00	88.00	92.00	92.00	96.00	96.00	92.00
9	83.80	85.09	86.84	85.96	87.28	88.16	86.84	84.00	88.00	80.00	88.00	84.00	80.00	88.00
10	85.90	83.77	88.16	85.96	87.28	88.16	88.60	84.00	80.00	92.00	88.00	92.00	92.00	92.00
Mean	84.83	84.01	88.09	85.70	87.31	88.23	87.93	83.46	83.82	85.43	84.65	86.25	86.25	87.80
Variance	1.660	3.090	1.784	1.971	2.475	2.458	0.721	31.244	18.322	29.389	26.980	36.061	43.172	13.633
Coefficient of variation	0.020	0.037	0.020	0.023	0.028	0.028	0.008	0.374	0.219	0.344	0.319	0.418	0.501	0.155

Table 12
Results of 10-folder cross-validation using data at year ($t-3$).

Validation	Prediction accuracy of training datasets (%)							Prediction accuracy of testing datasets (%)						
	Logit	NN	SVM	Multiple classifiers based on majority voting	Multiple classifiers based on Bayesian	Multiple classifiers based on BKS	Proposed method in this paper	Logit	NN	SVM	Multiple classifiers based on majority voting	Multiple classifiers based on Bayesian	Multiple classifiers based on BKS	Proposed method in this paper
1	70.20	70.18	70.18	70.61	68.86	71.93	69.30	60.00	60.00	52.00	64.00	56.00	60.00	64.00
2	70.50	66.52	67.84	67.84	69.60	71.81	70.04	69.23	61.54	61.54	57.69	69.23	73.08	69.23
3	69.70	63.60	74.56	71.05	68.86	71.89	75.00	68.00	76.00	68.00	84.00	76.00	84.00	70.61
4	68.00	67.54	66.67	70.18	66.67	70.18	67.54	52.00	64.00	52.00	52.00	72.00	60.00	60.00
5	67.40	64.76	71.81	69.16	66.96	73.57	71.37	61.54	73.08	69.23	69.23	65.38	61.54	69.23
6	64.50	66.23	68.86	65.79	66.67	70.18	67.54	64.00	64.00	72.00	64.00	64.00	68.00	76.00
7	65.20	66.96	67.40	66.96	65.20	68.28	67.40	69.23	69.23	69.23	69.23	73.08	69.23	73.08
8	70.20	64.04	70.18	68.42	70.18	74.56	66.67	76.00	80.00	84.00	64.00	68.00	68.00	80.00
9	70.60	66.67	68.42	69.74	69.74	71.05	71.49	64.00	68.00	64.00	68.00	60.00	60.00	68.00
10	69.30	68.86	69.74	69.30	67.54	70.18	68.42	56.00	60.00	52.00	52.00	56.00	56.00	60.00
Mean	68.56	66.54	69.57	68.91	68.03	71.36	69.48	64.00	67.59	64.40	64.42	65.97	65.99	69.02
Variance	4.953	4.220	5.403	2.768	2.720	3.299	6.589	49.422	48.367	107.911	88.533	48.993	68.939	41.927
Coefficient of variation	0.072	0.063	0.078	0.040	0.040	0.046	0.095	0.772	0.716	1.676	1.374	0.743	1.045	0.607

parameters. For independent NN classification, we also take the same optimal set among 20 experiment results as the results of independent NN classification. For multiple classifications based on majority voting method, multiple classifications based on Bayesian and multiple classifiers based on BKS, we also take the same optimal set among 20 experiment results as the results of single NN prediction.

Prediction results of the training datasets and the corresponding testing datasets using data at year $(t-2)$ are summarized in Table 11.

Financial prediction method proposed by this paper has lower average prediction accuracy for training datasets than SVM and multiple classifiers based on BKS. However, both of the variance and coefficient of variation based on proposed method are the lowest for training datasets. For testing datasets, our method has the highest prediction accuracy on all validations except validation 6, validation 7 and validation 8. Besides, the average prediction accuracy of the proposed method is also the highest for testing datasets. Also, the proposed method has the lowest variance and coefficient of variation for testing datasets.

Table 12 provides prediction results of the training datasets and the corresponding testing datasets by using data at year $(t-3)$.

The proposed method has higher average prediction accuracy for training dataset than Logit, NN, multiple classifiers based on majority voting and multiple classifiers based on Bayesian. Although the variance and coefficient of variation are the highest for training datasets, its variance and coefficient of variation are the lowest for testing datasets. Besides, the proposed method has the highest prediction accuracy on validation 1, validation 6, validation 7, validation 9 and validation 10 for testing datasets. Meanwhile, it has the highest average prediction accuracy for testing datasets.

4.4. Discussion

Experimental results show that financial distress prediction based on the proposed method has higher average prediction accuracy and lower variance and coefficient of variation than any other single classifiers and multiple classifiers for the testing sample, so it can greatly improve the prediction accuracy and prediction stability. It can combine outputs of single classifiers effectively. Hence, the multiple prediction method proposed in this paper is an effective and an excellent method for financial distress prediction.

5. Conclusions

The combination of multiple prediction methods has advantages in financial distress prediction. In this paper, we extended the research of combination of multiple prediction methods for financial distress. We used multiple prediction methods for financial distress prediction incorporating with rough set and D-S evidence theory. We applied rough set theory to determine the weight of each single prediction method and we took D-S evidence theory as the combination method. We gave a simple example to illustrate the research process of the proposed method. We used real world data of Chinese listed companies to evaluate performance of the proposed method and compared it with single classification methods and some other multiple classifiers. Results showed that proposed method had the highest prediction accuracy and the lowest variance and coefficient of variation for testing samples. Hence, multiple prediction method proposed in this paper is an effective and an excellent method for financial distress prediction.

The proposed method has some limitations. In this paper, although proposed method has good prediction results for testing samples, the prediction results for training samples are not so good. The criteria in Step 4, which diagnoses whether one company is in financial distress or not, may influence the prediction performance. If the value of criteria is too large, it may lower the accuracy for the prediction of companies in financial distress. If the value of criteria is too low, it may lower the accuracy for the prediction of healthy companies. A proper criteria is important to improve the prediction accuracy of financial distress. It will be interesting to search the optimal criteria such that both training accuracy and testing accuracy are high. Continuation of this work could take the direction.

Acknowledgements

We thank the editor and referees for their constructive remarks that helped to improve the clarity and the completeness of this paper.

References

[1] W.H. Beaver, Financial ratios and predictions of failure, Journal of Accounting Research 4 (1966) 71–111.
[2] E.I. Altman, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, Journal of Finance 23 (1968) 589–609.
[3] J.A. Ohlson, Financial ratios and the probabilistic prediction of bankruptcy, Journal of Accounting Research 18 (1980) 109–131.
[4] H.J. Zmijewski, Methodological issues related to the estimation of financial distress prediction models, Journal of Accounting Research 22 (1984) 59–82.
[5] H. Frydman, E.I. Altman, D. Kao, Introducing recursive partitioning for financial classification: the case of financial distress, Journal of Finance 40 (1985) 269–291.
[6] K.Y. Tam, Neural network models and the prediction of bank bankruptcy, Omega 19 (1991) 429–445.
[7] L. Salchenberger, C. Mine, N. Lash, Neural networks: a tool for predicting thrift failures, Decision Sciences 23 (1992) 899–916.
[8] L. Sun, P.P. Shenoy, Using Bayesian networks for bankruptcy prediction: some methodological issues, European Journal of Operational Research 180 (2007) 738–753.
[9] Y. Hu, Incorporating a non-additive decision making method into multi-layer neural networks and its application to financial distress analysis, Knowledge-Based Systems 21 (2008) 383–390.
[10] S.M. Bryant, A case-based reasoning approach to bankruptcy prediction modeling, International Journal of Intelligent Systems in Accounting, Finance and Management 6 (1997) 195–214.
[11] C.S. Park, I. Han, A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction, Expert Systems with Applications 23 (2002) 255–264.
[12] H. Li, J. Sun, Ranking order case-based reasoning for financial distress prediction, Knowledge-Based Systems 21 (2008) 868–878.
[13] F. Varetto, Genetic algorithm application in the analysis of insolvency risk, Journal of Banking and Finance 22 (1998) 1421–1439.
[14] K.S. Shin, Y.J. Lee, A genetic algorithm application in bankruptcy prediction modeling, Expert Systems with Applications 23 (2002) 321–328.
[15] A.I. Dimitras, R. Slowinski, R. Susmaga, C. Zopounidis, Business failure prediction using rough sets, European Journal of Operational Research 114 (1999) 263–280.
[16] T.E. Mckee, Developing a bankruptcy prediction model via rough sets theory, International Journal of Intelligent Systems in Accounting, Finance and Management 9 (2000) 59–173.
[17] S. Kaski, J. Sinkkonen, J. Peltonen, Bankruptcy analysis with self-organizing maps in learning metrics, IEEE Transaction on Neural Networks 12 (2001) 936–947.
[18] J.H. Min, Y.C. Lee, Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters, Expert Systems with Applications 28 (2005) 603–614.
[19] K.S. Shin, T.S. Lee, H.J. Kim, An application of support vector machine in bankruptcy prediction model, Expert Systems with Applications 28 (2005) 127–135.
[20] P. Ko, P. Lin, An evaluation-based approach with modularized evaluations to forecast financial distress, Knowledge-Based Systems 19 (2006) 84–91.
[21] P. Ravisankar, V. Ravi, Financial distress prediction in banks using group method of data handling neural network, counter propagation neural network and fuzzy ARTMAP, Knowledge-Based Systems 23 (2010) 823–831.
[22] J. Sun, H. Li, Data mining method for listed companies' financial distress prediction, Knowledge-Based Systems 21 (2008) 1–5.

[23] J. Sun, K.Y. He, H. Li, SFFS-PC-NN optimized by genetic algorithm for dynamic prediction of financial distress with longitudinal data streams, Knowledge-Based Systems 24 (2011) 1013–1023.
[24] H. Jo, I. Han, Integration of case-based forecasting, neural network, and discriminant analysis for bankruptcy, Expert Systems with applications 11 (1996) 415–422.
[25] F.Y. Lin, S. McClean, A data mining approach to the prediction of corporate failure, Knowledge-Based Systems 14 (2001) 189–195.
[26] J. Sun, H. Li, Listed companies' financial distress prediction based on weighted majority voting combination of multiple classifiers, Expert Systems with Applications 35 (2008) 818–827.
[27] J. Sun, H. Li, Financial distress prediction based on serial combination of multiple classifiers, Expert Systems with Applications 36 (2009) 8659–8666.
[28] L. Nanni, A. Lumini, An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring, Expert Systems with Applications 36 (2009) 3028–3033.
[29] E. Kim, W. Kim, Y. Lee, Combination of multiple classifiers for customer's purchase behavior prediction, Decision Support Systems 34 (2002) 167–175.
[30] M. Kim, S. Min, I. Han, An evolutionary approach to the combination of multiple classifiers to predict a stock price index, Expert Systems with Applications 31 (2006) 241–247.
[31] S.B. Cho, J.H. Kim, Multiple network fusion using fuzzy logic, IEEE Transactions on Neural Networks 6 (1995) 497–501.
[32] F. Yamaoka, Y. Lu, A. Shaout, M. Shridhar, Fuzzy integration of classification results in handwritten digit recognition systems, in: Proceeding of the fourth IWFHR, pp. 255–264.
[33] Y.S. Huang, K. Liu, C.Y. Suen, A neural approach for multi-classifier recognition systems, in: Proceeding of 4th IWFHR, (1994), pp. 235–244.
[34] E. Kim, W. Kim, Y. Lee, Combination of multiple classifiers for the customer's purchase behavior prediction, Decision Support Systems 34 (2002) 167–175.
[35] S.L. Hegarat-Mascle, I. Bloch, D. Vidar-Madjar, Application of Dempster–Shafer evidence theory to unsupervised classification in multisource remote sensing, IEEE Transactions on Geoscience and Remote Sensing 35 (1997) 1018–1031.
[36] L. Yang, J. Zhang, Improved behavior knowledge space combination method based on observational learning, in: IMACS: Multiconference on Computional Engineering in Systems Applications, vol. 1–2, (2006), pp. 1952–1956.
[37] A.P. Dempster, Upper and lower probabilities introduced by multivalued mappings, Annals of Mathematical Statistics 38 (1967) 325–339.
[38] L. Xu, A. Krzyżak, C.Y. Suen, Methods of combining multiple classifiers and their applications to handwriting recognition, IEEE Transactions on Systems, Man, and Cybernetics 22 (1992) 418–435.
[39] M.J. Beynon, A novel technique of object ranking and classification under ignorance: an application to the corporate failure risk problem, European Journal of Operation Research 167 (2005) 493–517.
[40] R. Sun, H.Z. Huang, Q. Miao, Improved information fusion approach based on D-S evidence theory, Journal of Mechanical Science and Technology 22 (2008) 2417–2425.
[41] B.Y. Liu, G.Y. Ou, H.B. Chang, Study of weighted D-S evidence combination based on rough set theory and its application, Computer Engineering and Design 27 (2006) 3940–3942.
[42] B. Zhong, Z. Xiao, A new method of combination forecast based on rough set theory, Statistical Research 11 (2002) 37–39.
[43] Z. Xiao, S. Ye, B. Zhong, C. Sun, BP neural network with rough set for short term load forecasting, Expert Systems with Applications 36 (2009) 273–279.
[44] W. Li, Y. Shang, Study of weigh of determination based on rough set theory in de-sulfuration projects decision, in: Proceeding of IEEE International Symposium on Knowledge Acquisition and Modeling Workshop, Vol. 1–2, (2008), pp. 305–307.
[45] Z. Pawlak, Rough sets, International Journal of Computer and Information Sciences 11 (1982) 341–356.
[46] B. Yang, K.J. Kim, Application of Dempster–Shafer theory in fault diagnosis of induction motors using vibration and current signals, Mechanical Systems and Signal Processing 20 (2006) 403–420.
[47] Y. Liu, Y. Jiang, X. Liu, S. Yang, CSMC: a combination strategy for multi-class classification based on multiple association rules, Knowledge-Based Systems 21 (2008) 786–793.
[48] D. Wu, Supplier selection in a fuzzy group setting: a method using grey related analysis and Dempster–Shafer theory, Expert Systems with Applications 36 (2009) 8892–8899.
[49] L. Dymova, P. Sevastjanov, An interpretation of intuitionistic fuzzy sets in terms of evidence theory: decision making aspect, Knowledge-Based Systems 23 (2010) 772–782.
[50] R. Setia, G.S. May, Run-to-run failure detection and diagnosis using neural networks and Dempster–Shafer theory: an application to excimer laser ablation, IEEE Transactions of Electronics Packaging and Manufacture 29 (2006) 42–49.
[51] L. Ai, J. Wang, X. Wang, Multi-features fusion diagnosis of tremor based on artificial neural network and D-S evidence theory, Signal Processing 88 (2008) 2927–2935.
[52] R.C. Lacher, K.C. Pamela, S.C. Sharma, L.F. Fant, A neural network for classifying the financial health of a firm, European Journal of Operation Research 85 (1995) 53–63.
[53] F.Y. Lin, S. McClean, A data mining approach to the prediction of corporate failure, Knowledge-Based Systems 14 (2001) 189–195.
[54] H. Yang, L. Tai, Predicting financial distress of listed corporations based on fuzzy support vector machine, Journal of Management Sciences in China 12 (2009) 102–110.