Knowledge-Based Systems 85 (2015) 52–61

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

The performance of corporate financial distress prediction models with features selection guided by domain knowledge and data mining approaches

Ligang Zhou [a,*], Dong Lu [b], Hamido Fujita [c]

[a] School of Business, Macau University of Science and Technology, Taipa, Macau
[b] School of Business, SiChuan Normal University, SiChuan Province, PR China
[c] Faculty of Software and Information Science, Iwate Prefectural University, Iwate, Japan

ARTICLE INFO

Article history:
Received 18 April 2014
Received in revised form 17 March 2015
Accepted 20 April 2015
Available online 27 April 2015

Keywords:
Financial distress prediction
Features selection
Domain knowledge
Data mining

ABSTRACT

Experts in finance and accounting select feature subset for corporate financial distress prediction according to their professional understanding of the characteristics of the features, while researchers in data mining often believe that data alone can tell everything and they use various mining techniques to search the feature subset without considering the financial and accounting meanings of the features. This paper investigates the performance of different financial distress prediction models with features selection approaches based on domain knowledge or data mining techniques. The empirical results show that there is no significant difference between the best classification performance of models with features selection guided by data mining techniques and that by domain knowledge. However, the combination of domain knowledge and genetic algorithm based features selection method can outperform unique domain knowledge and unique data mining based features selection method on AUC performance.

1. Introduction

Corporate financial distress prediction (CFDP) is very important for investors, credit lenders and company's partners, such as suppliers or retailers. The investors and credit lenders need to evaluate the financial distress risk of a company before they make any investment or credit granting decisions on the company in order to avoid suffering a great loss. A company's suppliers or retailers always conduct credit transaction with the company and they also need to fully understand the company's financial status and make decisions on the credit transaction.

To correctly predict a company's financial distress is a great concern for many stake holders of a company. This practical significance has driven a lot of studies on the issue of corporate financial distress prediction. Most of these studies often focused on introducing or improving the quantitative approaches from statistics and data mining discipline to develop corporate financial distress prediction models (CFDPM) with the objective of increasing the prediction accuracy. The preliminary study of CFDPM with a multivariate framework proposed by Altman [1] was based on the discriminant analysis approach. Thereafter, many other complex statistical and data mining methods were introduced to develop the CFDPM, such as neural networks [2,3], decision trees [4], and support vector machines [5]. In addition, the fuzzy theory can also be used for developing CFDPM [6,7]. Most recent research mainly focuses on the development of hybrid models with the combination of two or more than two methods [8–10]. Although the empirical results in these studies often showed that hybrid models could outperform the single models, the computation always consumes more time and the theory or reason for the combinations is not always known and explained, which prevent their wide applications in practice to some degree.

The problem of corporate financial distress prediction is to take advantage of all currently available information related to the company to predict if it will fall into the condition of default or financial difficulty. Consequently, the performance of the CFDPM is determined not only by the model or methods that is used for the prediction but also by the selection of available information. In practice, some credit rating agencies just use their experiences and judgments to select the relevant information to evaluate the credit risk of a particular company or individual with a simple scorecard instead of complex statistical models [11]. However, the information related to a company is huge, including macroeconomic situations, company characteristics, financial status and market information, and most studies have demonstrated that financial and marketing information is the most effective in

* Corresponding author.
E-mail addresses: mrlgzhou@gmail.com (L. Zhou), dlu@sicnu.edu.cn (D. Lu), issam@iwate-pu.ac.jp (H. Fujita).

http://dx.doi.org/10.1016/j.knosys.2015.04.017
0950-7051/© 2015 Elsevier B.V. All rights reserved.

financial distress prediction. What financial and marketing information should be considered in the development of corporate financial distress prediction models?

There are often two research streams in the feature subset selection for corporate financial distress prediction models. One is based on the domain knowledge from financial and accounting theory. The main characteristic of the features selected by domain knowledge is that the effect of the features on the financial distress can be evaluated to some degree in terms of financial and accounting theory. Altman [1] investigated a set of twenty-two financial and economic ratios in the prediction of corporate bankruptcy and found that the subset of the following variables is useful for financial distress prediction: working capital/total assets, retained earnings/total assets, earnings before interest and taxes/total assets, market value equity/book value of total debt. Altman et al. [12] observed the distinct difference in the accounting procedures and the quality of financial documents between the firms in China and those in the western world, and considered variables that were widely accepted in China and deemed contributive in previous studies. They investigated fifteen variables that reflect various aspects of a company, such as profitability, liquidity and solvency, and asset management efficiency and capital structure and financial leverage. After considering a large number of combinations of the 15 characteristic variables, they found that the following feature subset yielded the best performance: total liabilities/total assets, net profit/average total assets, working capital/total assets, and retained earnings/total assets. Shumway [13] developed a simple hazard model and compared the performance of Altman's variables [1] and Zmijiewski's variables [14] and a new set of variables including accounting and three market-driven variables. The empirical result shows that the new accounting and market-driven variables set outperforms other two alternative models in out-of-sample forecasts. The accounting and market-driven feature subset includes: net income/total asset, total liabilities/total asset, relative size (market capitalization/total size of the corresponding market), the firm's past excess returns and the idiosyncratic standard deviation of the firm's stock returns. Ravi and Ravi [15] reviewed 128 papers in bankruptcy prediction and listed more than 500 different variables used by these different papers. Almost all of these 128 papers used different subsets of features. It is perhaps natural that different experts have different opinions in determining what information should be considered in the prediction of financial distress of a company.

Another stream in feature subset selection is based on data mining techniques. Adherents to the data mining stream view believe that data will tell everything, and the approach uses some features selection methods in data mining to identify which feature subset can improve the prediction performance without considering the financial and accounting meanings of the features. Tsai [16] compared five well-known features selection methods used in bankruptcy prediction and used multi-layer perceptron neural networks to construct the prediction model, and found the *t*-test features selection method performs better than others. du Jardin [17] introduced a neural network based model using a set of variables selected by a criterion being adapted to the network for the bankruptcy prediction problem. Drezner et al. [18] reported that a tabu search based variables selection model can increase the predictability of corporate bankruptcy by up to 10 percentage points in comparison to Altman's *Z*-Score [1] model. Although most researchers in this stream like Cho, Mays, et al. [10,19] noticed that there were hundreds of financial variables and the model performance was affected by input variables selection, they only investigated a very small subset of variables guided by previous studies in the data set for empirical study without taking good advantage of the original data set from which the sample for training and testing model was retrieved. Few previous studies in financial distress prediction compare the performance of features selection with domain knowledge and data mining, together with investigating the difference of feature subset found by domain knowledge and data mining [2–4,8–10].

The contribution of this study is twofold. First, it compares the performance of domain knowledge and data mining based features selection methods in financial distress prediction on a data set with more than three hundred variables. The experimental result shows that the features selected by data mining methods can perform as well as those selected by domain knowledge of experts in finance or accounting. Second, it considers the combination of domain knowledge and data mining features selected approach in order to take good advantage of the experts' professional knowledge and the powerful mining capability of data mining techniques. The experimental result shows that the performance of the combined method can outperform unique domain knowledge and unique features selection method.

The outline of this paper is as follows. Section 2 introduces the important domain knowledge and data mining feature subset selection methods for financial distress prediction. Section 3 reports the empirical results and Section 4 gives the conclusion.

2. Domain knowledge vs. data mining in features selection

2.1. Features selection by domain knowledge

Financial ratio analysis is an important way to analyze financial statements. There are often hundreds of financial ratios measuring different aspects of a company, such as liquidity, long-term solvency, asset management, profitability, and market value. The meaning and usage of the financial variables has been widely discussed in finance [20,21]. It is impossible to investigate all financial ratios suggested for CFDPM by the researchers from finance and accounting. Only the ratios that are widely accepted and have been verified with great performance and have been taken as a benchmark in most previous research are considered. Therefore, a classical group of features selected from domain knowledge is based on the work from Altman [1], Altman [12] and Shumway [13]. The feature subset employed by Altman [1], Altman [12] and Shumway is denoted as FA1, FA2, and FS respectively. The union of these three feature subsets is denoted by FAAS. The detail of the ten features in FAAS is briefly described as follows.

1. Working capital to total assets (WCTA) measures the firm's liquidity or short-term solvency. High WCTA shows that the firm can match its account payable obligation on time and a low WCTA indicates that the firm may be unable to pay its suppliers and creditors.
2. Retained earnings to total assets (RETA) reflects a firm's strategy on its net earnings. If a firm needs more funds for the increase of business and it prefers to raise funds from inside, the firm would like to keep a higher RETA.
3. Earnings before interest and taxes to total assets (EBTITA) is an important measures of a firm's profitability. Higher EBITTA indicates higher profitability of a firm.
4. Sales to total assets (STA) is also a measures of a firm's profitability. A low ratio indicates that the total assets of the firm cannot provide adequate revenue.
5. Net income to total assets (NITA) is also known as return on assets (ROA). It indicates how efficient a firm's management is at using its assets to generate earnings. It is another important measure of a firm's profitability.
6. Total liabilities to total assets (TLTA) measures a firm's long-term solvency. It indicates a firm's financial risk by determining what ratio of company's assets is financed by debt. Higher TLTA means higher financial risk.

7. Market equity to total liabilities (METL) is the ratio of total market value of all the outstanding shares (market capitalization) to the total liabilities. It shows market reaction to financial status.
8. Excess annual return over the market (EAR) measures the difference between the stock return and the market return.
9. Firm's market capitalization to total market capitalization (FMC2MC) measures the firm's size related to the whole market.
10. Standard deviation of day's stock return (Sigma) indicates the fluctuation of a firm's stock daily return.

2.2. Data mining methods for features selection

Feature selection techniques have been widely studied in data mining. Guyon and Elisseeff [22] point out the potential benefits of feature selection: facilitating data understanding, reducing computational time and defying the "curse of dimensionality" to improve prediction performance.

The features selection methods can be generally categorized into three groups:

1. Filter methods that select variables by ranking them with information generated from data, such as relative entropy and absolute value two-sample *t*-test with pooled variance estimate.
2. Wrapper methods that assess feature subset according to their performance to a given model. This contains a searching procedure to search the space of possible feature subset and evaluate each subset in terms of the performance of the given model on the subset. If there are 20 features, the number of different feature subset will be $2^{20}-1$ (the null subset is excluded). The practical computation time taken in exhaustive searching is always unacceptable, and the heuristic searching methods, such as simulated annealing, genetic algorithms, and tabu search are better alternatives.
3. Embedded methods, which incorporate features selection as part of the training process of the model, such as the Least Absolute Shrinkage and Selection Operator (Lasso) method for contrasting a linear model.

In this study, different feature subset selection methods from each of above three categories are introduced or developed.

Suppose Y is dependent variable indicating whether the firm falls into financial distress or not and the feature values of an observation $\boldsymbol{x} = (x_1, \ldots, x_m)$ is an assignment of a set of features $\boldsymbol{X} = (X_1, X_2, \ldots, X_m)$. There are in total N observations. $Y \in \{1,\ 0\}$ and y is an assignment of values of Y. $y = 1$ means that the firm has financial distress, $y = 0$ means that the firm's finance is normal.

2.2.1. Filter methods

For the filter methods, three different criteria for ranking the features are employed: entropy [23], *t*-test [24] and receiver operating characteristic (ROC) [25].

1. Entropy ranking
   Information theory measures entropy as follows:
   $$H(Y) = -\sum_{y\in Y} p(y)log_2(p(y)) \tag{1}$$
   where $p(y)$ is the marginal probability density function for the random variable $Y = y$. If the observed Y in the training dataset is portioned according to the value of a feature X_i, the entropy of Y after observing X_i is given:
   $$H(Y \mid X_i) = -\sum_{x_i\in X_i} p(x_i) \sum_{y\in Y} p(y \mid x_i) log_2(p(y \mid x_i)) \tag{2}$$
   where $p(y|x_i)$ is the conditional probability of y given x_i. Therefore $H(Y \mid X_i)$ is defined as the entropy of a feature i, $i = 1, 2, \ldots, m$.
2. *t*-test
   Suppose $v_j^1 = var(X_j \mid y_i = 1), v_j^0 = var(X_j \mid y_i = 0)$, where $\mathrm{var}(\cdot)$ is the variance of a group of values, $m_j^1 = mean(X_j \mid y = 1),\ m_j^0 = mean(X_j \mid y = 0)$, where $mean(\cdot)$ is the mean of a group of values, features weighting strategy based on *t*-test on the training dataset is defined as following [19,24]:
   $$z_j = \frac{\left|m_j^1 - m_j^0\right|}{\sqrt{v_j^1 \big/ n^1 + v_j^0 \big/ n^0}} \tag{3}$$
   where n^1, and n^0 denote the number of observations with $y = 1$ or $y = 0$ respectively.
3. ROC
   The idea of ROC based features ranking is to rank the features in terms of the area under the convex hull of the ROC curve. The ROC curve can easily be constructed by sweeping the threshold and computing percentages of wrong and correct classifications over the available training feature vectors [19,25].

Each feature in the data set can be ranked in terms of any one of the above criteria. The top ranked features are always selected, but they may exhibit highly dependence. Therefore we develop the following filter method for feature subset selection which introduces the Variance Inflation Factor (VIF) to prevent high dependency among features. The VIF of a feature i is denoted by V_i is calculate as follows:

$$V_i = \frac{1}{1 - R_i^2} \tag{4}$$

where R_i^2 is the R-square of regression equation $X_i = \beta_0 + \boldsymbol{\beta X}'$, where $\boldsymbol{X}' = \boldsymbol{X} - \{X_i\}$.

The algorithm based on the filter method and VIF (FVIF) for feature subset selection is given as follows:

```
Algorithm FVIF
Input: The training sample set $S_r$ with $m$ features $X_1, X_2, \ldots,$
  $X_m$ and $Y$; the number of features $m^*$ to be selected.
Output: feature subset $F^*$
1. $F^* = \emptyset$;
2. Rank the $m$ features in terms of entropy or t-test or ROC
   criteria in descending order, add the first feature to $F^*$
   and then calculate VIF of the second feature according to
   the feature in $F^*$. If the VIF value of the second feature is
   less than 10 (empirical value suggested in [26]), then the
   second feature will be added to $F^*$; otherwise, move to
   the next feature, and so on, until the number of total fea-
   tures in $F^*$ is $m^*$.
```

The algorithm FVIF taking entropy, *t*-test or ROC ranking criteria is denoted by FV-en, FVIF-tt, and FVIF-roc respectively.

2.2.2. Wrapper methods

There are various searching approaches that can be used in wrapper methods. The simple searching approach is to use sequential forward selection (SFS) which adds new features from a candidate subset if the introduction of the new feature to the model can reduce the error of the model. One disadvantage of SFS is that once a feature is selected, even if it becomes obsolete after the addition of other features, it cannot be removed. Heuristic searching methods can overcome this shortcoming of SFS.

Since the feasible feature subset can be denoted by a vector of binary numbers, for example, if there are a total of five features, i.e. $m = 5$, a feasible feature subset instance with X_2 and X_5 selected can be coded as (0, 1, 0, 0, 1). This vector can be naturally taken as a genome in a genetic algorithm. The final optimal genome found by a genetic algorithm can easily be transformed to the optimal feature subset. Therefore, in this study, the genetic algorithm is selected for wrapper methods.

The following wrapper method is developed, based on genetic algorithm (WMBGA) for feature subset selection. Each individual in a population is denoted by its genome, a serial of m gene like $(g_1, g_2, \ldots, g_m), g_i \in \{0, 1\}$.

Algorithm WMBGA

```
Input: $S_r$: the training sample set with $m$ features $X_1, X_2, \ldots, X_m$ and $Y$;
$m^*$: the number of features to be selected;
$K$: the number of individuals in the population;
$G_{max}$: the maximum number of generations;
$I_u$: the maximum tolerable number of iterations that the fitness function has not improved;
$r_1$: The proportion of individuals in the current population that will be admitted to the next generation unchanged;
$r_2$: The proportion of individuals in the current population (which proportion will be mutated).
Output: feature subset $F^*$
 1. $i = 1$, Create initial population $P_i$ with $K$ individuals. Each individual is generated by setting $m^*$ randomly selected genes to 1. $u = 0; f_{min} = 1$ (the initial value of the fitness value).
 2. while $i \leqslant G_{max}$ and $u \leqslant I_u$
   2.1 Evaluate the fitness of each individual in the population $P_i$ and sort their performance in ascending order.
   2.2 if the fitness value of the top individual $f_{top} \leqslant f_{min}$, then
         $ind^* = ind_{top}$ (the top individual genome), $f_{min} = f_{top}, u = 1$,
       else
         $u = u + 1$
       endif
   2.3 The last selected $[K \times (1 - r_1)]$ ranked individuals in the current population for crossover. $[\cdot]$ is a function to take the roundup.
   2.4 Randomly selected $[K \times r_2]$ individual in the current population for mutation.
   2.5 $i = i + 1$,
endwhile
 3. $F^*$ contains all $X_j$ where $g_j = 1$ in $ind^* = (g_1, g_2, \ldots, g_m)$.
The fitness function on an individual $(g_1, g_2, \ldots, g_m)$ is defined as follows.
 1. Select the corresponding $m^*$ features where $g_k = 1$, $k = 1, 2, \ldots, m$, which is denoted by $X'_1, X'_2, \ldots, X'_{m^*}$.
 2. Construct the following prediction model with the selected data from the training sample set $S_r$.
      $\widehat{Y}_i = f(\boldsymbol{X}'_i)$    (5)
    where $\boldsymbol{X}'_i = (1, X'_{1i}, X'_{2i}, \ldots, X'_{m^*i}),\ i = 1, 2, \ldots, N$.
 3. The fitness function value is defined as
      $\varepsilon = \frac{\sum_{i=1}^{N} e_i}{N}$    (6)
    where
      $e_i = \begin{cases} 0 & \text{if } \widehat{Y}_i = Y_i \\ 1 & \text{if } \widehat{Y}_i \neq Y_i \end{cases}$
```

2.3. Embedded methods

Lasso is a regularization technique that can reduce the number of features in a generalized linear model and both identify and separate important features from redundant features. In this study, the elastic net (EN) [27] which is a hybrid of ridge regression and lasso regularization is introduced to feature subset selection. Empirical studies suggest that the elastic net technique can outperform lasso data with highly correlated features [27,28]. The elastic net for generalized linear models is defined as follows [28].

$$\min_{\beta_0, \boldsymbol{\beta}} \left(\frac{Dev(\beta_0, \boldsymbol{\beta})}{N} + \lambda P_\alpha(\boldsymbol{\beta}) \right) \tag{7}$$

where

$$Dev(\beta_0, \boldsymbol{\beta}) = \sum_{i=1}^{N} \left(y_i - \beta_0 - \boldsymbol{x}_i^T \boldsymbol{\beta} \right)^2 \tag{8}$$

$$P_\alpha(\boldsymbol{\beta}) = \sum_{j=1}^{m} \left(\frac{(1-\alpha)}{2} \beta_j^2 + \alpha |\beta_j| \right) \tag{9}$$

$$\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_m) \tag{10}$$

λ is a nonnegative regularization parameter and α is number strictly between 0 and 1.

2.4. The framework of experimental design

The framework of the experiment to test the performance of different features selection methods on the data set is shown as Fig. 1.

3. Empirical study

3.1. The data set

In China, the stock exchanges give a company "special treatment (ST)" to indicate risk warning to investors in the stock markets. The rules for giving ST to a company are usually based on the financial performance of a company in the past fiscal years. If a company receiving ST is not able to recover in the following specified years, the company will face the risk of being suspended or

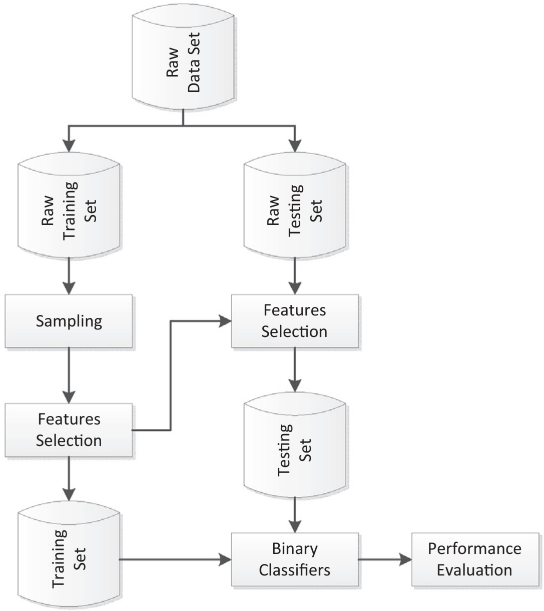


Fig. 1. The framework of the experimental design.

delisted. Therefore, the companies receiving ST can be taken as companies with financial distress.

The data set is collected from a commercial dataset GuoTaiAn (GTA) – China Stock Market and Accounting Research Database (CSMAR) which contains the ST records and a wide variety of financial data and market data of the listed companies in China. The original dataset has 208 different financial ratios measuring short-term solvency, asset management or turnover, long-term solvency, profitability, capital structure, stock holder's earning profitability, cash management and development capability. The financial ratios whose missing values take more than 20% of the total company-year observations are excluded. Finally, there are 166 financial ratios and 3 three market variables used by Shumway [13], retrieved and calculated from the database. To consider the effect of the change of each financial ratio on financial distress, for each financial ratio, the following features are also introduced:

$$\Delta X_i^t = \frac{X_i^t - X_i^{t-1}}{X_i^{t-1}} \tag{11}$$

where X_i^t denotes the value X_i in the fiscal year t.

The three market variables introduced by Shumway [13] are defined as follows in this dataset.

1. *Excess annual return over the market (EAR)*: daily return of the stock minus the corresponding index return of Shanghai A-shares market or Shenzhen A-shares market cumulated to obtain the yearly return. If the stock is listed in Shanghai stock market, the index of Shanghai A-shares market will be chosen for the computation. Likewise, for the stock listed in Shenzhen market, the index of Shenzhen A-shares will be chosen.
2. *Firm's market capitalization to total market capitalization (FMC2MC)*: the log value of the market capitalization of the stock in the last trading day of the fiscal year to the market capitalization of the total market capitalization of Shanghai A&B-shares and ShenZhen A&B shares.
3. *Standard deviation of day's stock return (σ)*:

$$\sigma = \sqrt{\frac{\sum_{t=D_l-59}^{D_l}\left(R_{stock}^t - \overline{R}_{stock}\right)^2}{60-1}} \tag{12}$$

where D_l is the last trading day of the fiscal year;

$R_{stock}^t = \frac{P_{stock}^t - P_{stock}^{t-1}}{P_{stock}^{t-1}}$, where P_{stock}^t is the last price on day t.

$\overline{R}_{stock}$ is the average return of the last sixty trading days of the stock.

The final sample has a total of 338 features and 10,365 company-year observations for ST companies and Non-ST companies. The description of the features can be found in [29]. The missing value will be filled by the mean of the corresponding value of companies in the same industry in the same fiscal year. The 338 features include the features used by Altman [1], Altman [12] and Shumway [13]. In the China stock market, the listed companies always disclose their financial statements for the last fiscal year around April in a year. The special treatment is always given after the disclosure of the financial statements, and the prediction of ST before the disclosure of financial statements is important for investors. Therefore, to predict the ST for a company in a calendar year T before the disclosure of financial statements, only the variables observed in or before the fiscal year $T-2$ can be used. The companies observed in each year can be categorized into two groups: ST companies and Non-ST companies. Therefore, any classification method can be used for the prediction of corporate financial distress. In this study, 6 different common prediction modes: logistic regression (LR), k-nearest neighbors (kNN), decision tree C4.5 (C4.5), ripper (RP), neural networks (NN) and support vector machines (SVM) are tested under different features selection approaches.

There are, in total, 287 ST companies in the final sample and Fig. 2 shows the number of ST companies by calendar year. It can be observed that the maximum number of STs occurred in year 2006 and 2007 which is very close to the subprime crisis which happened in USA in 2007. Although it is difficult to provide evidence to show the relationship between the number of ST companies and the subprime crisis, the USA is China's largest trade partner and the global business places China and USA "in the same boat" when facing economic crisis from any partner.

3.2. Experiment settings

The number of Non-ST companies by calendar year is shown in Fig. 3. From Figs. 2 and 3, it can be observed that the number of Non-ST companies is much greater than ST companies in each year and it makes the prediction of corporate financial distress a highly imbalanced classification problem. In this study, the regression model is estimated with the Ordinary Least Squares method which estimates the parameters in the model with the object of minimizing the sum of squared errors. If the training sample is highly imbalanced, the model may have bias in classifying all instances into Non-ST and still keeping very small error. Therefore, the sampling strategy for the training sample set and test sample set is as follows [30]:

1. *Training sample set*: select all 180 ST company instances in and before the calendar year 2006, and randomly select the same number of Non-ST company instances as that of ST company instance in each year in and before 2006. Then the total number of instances is 360 in the training sample.
2. *Test sample set*: since in practice, the financial distress prediction model should be used for all listed companies with available information, to test the capability of the model in a practical situation, all valid instances observed after 2006 (from

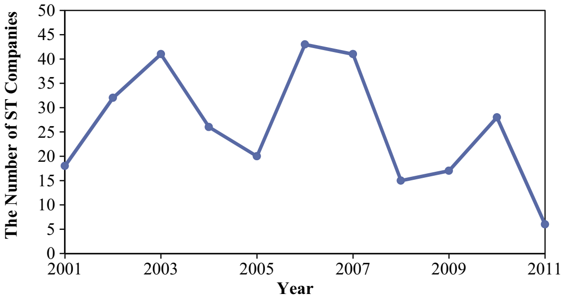


Fig. 2. The number of ST companies by calendar year.

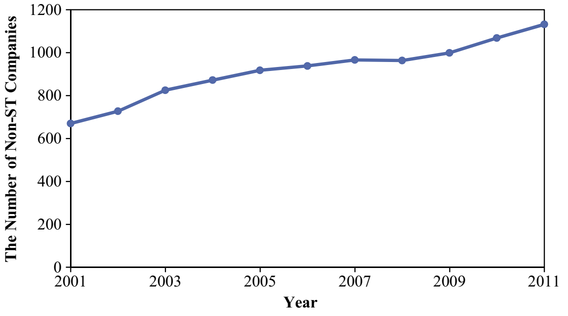


Fig. 3. The number of Non-ST company by year.

Table 1
Average of performance of classification models with different feature subset selection methods.

Approaches	LOG				kNN			
	Sen	Spe	Acc	AUC	Sen	Spe	Acc	AUC
FV-en	0.9389 (0.0148)	0.9738 (0.0106)	0.9397 (0.0143)	**0.9733** (0.0008)	0.9294 (0.0108)	0.9841 (0.0045)	0.9305 (0.0105)	0.9680 (0.0016)
FV-tt	0.9376 (0.0141)	0.9748 (0.0108)	0.9384 (0.0136)	**0.9732** (0.0008)	0.9269 (0.0113)	0.9841 (0.0045)	0.9281 (0.0110)	0.9698 (0.0020)
FV-roc	0.9342 (0.0189)	0.9729 (0.0112)	0.9350 (0.0183)	**0.9727** (0.0013)	0.9236 (0.0094)	0.9794 (0.0106)	0.9247 (0.0093)	0.9654 (0.0054)
SFS	0.9155 (0.0127)	0.8561 (0.0529)	0.9143 (0.0117)	0.9485 (0.0113)	0.8174 (0.0365)	0.6738 (0.0439)	0.8145 (0.0350)	0.8202 (0.0083)
WMBGA	0.9374 (0.0104)	0.9701 (0.0131)	0.9380 (0.0100)	**0.9725** (0.0022)	0.9157 (0.0218)	0.9673 (0.0212)	0.9167 (0.0216)	0.9601 (0.0106)
EN	0.9212 (0.0216)	0.8701 (0.0645)	0.9202 (0.0206)	0.9438 (0.0205)	0.9212 (0.0196)	0.8654 (0.0475)	0.9201 (0.0196)	0.9384 (0.0149)
FA1	0.9206 (0.0232)	0.9168 (0.0444)	0.9205 (0.0227)	0.9581 (0.0103)	0.8722 (0.0780)	0.5121 (0.2723)	0.8648 (0.0766)	0.7605 (0.1408)
FA2	0.9131 (0.0129)	0.8879 (0.0458)	0.9126 (0.0118)	0.9601 (0.0056)	0.8224 (0.0312)	0.6551 (0.0380)	0.8190 (0.0301)	0.8045 (0.0102)
FS	0.9186 (0.0188)	0.9617 (0.0213)	0.9195 (0.0181)	**0.9725** (0.0020)	0.8678 (0.0276)	0.8458 (0.0276)	0.8673 (0.0268)	0.9236 (0.0065)
FAAS	0.9310 (0.0121)	0.9178 (0.0260)	0.9307 (0.0115)	0.9670 (0.0033)	0.8518 (0.0366)	0.7430 (0.0265)	0.8496 (0.0353)	0.8806 (0.0099)
	C4.5				Ripper			
FV-en	0.9822 (0.0030)	0.9311 (0.0131)	0.9322 (0.0127)	0.9567 (0.0051)	0.9832 (0.0039)	0.9298 (0.0133)	0.9309 (0.0129)	0.9565 (0.0050)
FVIF-tt	0.9832 (0.0039)	0.9272 (0.0138)	0.9283 (0.0135)	0.9543 (0.0065)	0.9841 (0.0045)	0.9233 (0.0184)	0.9246 (0.0180)	0.9522 (0.0099)
FVIF-roc	0.9832 (0.0039)	0.9277 (0.0140)	0.9289 (0.0136)	0.9546 (0.0066)	0.9860 (0.0066)	0.9113 (0.0341)	0.9129 (0.0333)	0.9549 (0.0071)
SFS	0.9832 (0.0086)	0.8974 (0.0327)	0.8992 (0.0320)	0.9274 (0.0325)	0.9748 (0.0177)	0.9179 (0.0180)	0.9191 (0.0175)	0.9493 (0.0087)
WMBGA	0.9720 (0.0153)	0.9270 (0.0122)	0.9279 (0.0122)	0.9493 (0.0184)	0.9636 (0.0213)	0.9109 (0.0171)	0.9119 (0.0167)	0.9363 (0.0154)
EN	0.9355 (0.0589)	0.8922 (0.0269)	0.8931 (0.0263)	0.9178 (0.0338)	0.9832 (0.0059)	0.9080 (0.0193)	0.9095 (0.0189)	0.9492 (0.0131)
FA1	0.9804 (0.0103)	0.9013 (0.0322)	0.9029 (0.0316)	0.9336 (0.0267)	0.9813 (0.0076)	0.9163 (0.0216)	0.9176 (0.0211)	0.9437 (0.0177)
FA2	0.9187 (0.0392)	0.9043 (0.0358)	0.9046 (0.0345)	0.9259 (0.0206)	0.9327 (0.0341)	0.9005 (0.0316)	0.9012 (0.0307)	0.9172 (0.0160)
FS	0.9776 (0.0118)	0.9288 (0.0147)	0.9298 (0.0143)	0.9532 (0.0075)	0.9813 (0.0000)	0.9205 (0.0139)	0.9217 (0.0136)	0.9509 (0.0069)
FAAS	0.9776 (0.0118)	0.9288 (0.0147)	0.9298 (0.0143)	0.9532 (0.0075)	0.9776 (0.0118)	0.9273 (0.0148)	0.9283 (0.0144)	0.9533 (0.0075)
	NN				SVM			
FV-en	0.9309 (0.0152)	0.9804 (0.0053)	0.9319 (0.0148)	0.9696 (0.0052)	0.9348 (0.0077)	0.9766 (0.0079)	0.9356 (0.0076)	**0.9713** (0.0007)
FVIF-tt	0.9149 (0.0219)	0.9822 (0.0053)	0.9163 (0.0214)	0.9664 (0.0050)	0.9211 (0.0106)	0.9804 (0.0030)	0.9223 (0.0104)	**0.9713** (0.0008)
FVIF-roc	0.9150 (0.0257)	0.9720 (0.0117)	0.9162 (0.0250)	0.9584 (0.0083)	0.9128 (0.0100)	0.9804 (0.0030)	0.9142 (0.0098)	**0.9701** (0.0013)
SFS	0.8551 (0.0595)	0.9084 (0.0466)	0.8561 (0.0578)	0.9098 (0.0439)	0.7062 (0.1018)	0.9290 (0.0565)	0.7108 (0.0988)	0.9282 (0.0086)
WMBGA	0.9020 (0.0231)	0.9542 (0.0446)	0.9031 (0.0221)	0.9594 (0.0165)	0.9345 (0.0071)	0.9748 (0.0063)	0.9353 (0.0069)	**0.9721** (0.0009)
EN	0.8373 (0.0515)	0.8907 (0.0466)	0.8384 (0.0502)	0.9297 (0.0202)	0.8626 (0.1372)	0.7879 (0.2199)	0.8611 (0.1308)	0.9621 (0.0068)
FA1	0.8937 (0.0397)	0.9262 (0.0518)	0.8944 (0.0390)	0.9366 (0.0337)	0.8558 (0.0434)	0.9449 (0.0142)	0.8576 (0.0426)	0.9612 (0.0110)
FA2	0.8823 (0.0255)	0.9505 (0.0237)	0.8836 (0.0247)	0.9472 (0.0207)	0.8624 (0.0111)	0.8692 (0.0268)	0.8625 (0.0106)	0.9414 (0.0074)
FS	0.8588 (0.0281)	0.9907 (0.0099)	0.8615 (0.0274)	0.9530 (0.0203)	0.8551 (0.0255)	0.9748 (0.0117)	0.8575 (0.0249)	**0.9716** (0.0021)
FAAS	0.9079 (0.0157)	0.9570 (0.0193)	0.9089 (0.0155)	0.9545 (0.0078)	0.7937 (0.0749)	0.9290 (0.0399)	0.7964 (0.0726)	0.9421 (0.0087)

2007 to 2011) are used. There are a total of 107 ST company-year instances and 5128 Non-ST company-year instances in the test sample.

The general measures of performance of classification models as follows are used.

1. Sensitivity (sen) = $\frac{NN}{NN+NS}$
2. Specificity (spe) = $\frac{SS}{SN+SS}$
3. Accuracy (acc) = $\frac{NN+SS}{NN+NS+SS+SN}$

where *NN*: the number of Non-ST companies correctly classified as Non-ST companies, *NS*: the number of Non-ST companies

Table 2
Comparison of average rank on AUC among the 10 different feature subset selection methods.

Models	FV-en	FV-tt	FV-roc	SFS	WMBGA	EN	FA1	FA2	FS	FAAS
LOG	**2.4**	**3.1**	**3.3**	9.1	**3.0**	9.1	7.7	7.7	**3.3**	6.3
kNN	**2.0**	**1.4**	**2.9**	8.3	**3.9**	**5.2**	8.6	9.3	6.1	7.3
C4.5	**1.5**	**3.7**	**4.2**	7.1	**4.9**	8.2	7.0	8.4	**4.5**	**5.5**
Ripper	**2.1**	**3.9**	3.2	**6.0**	7.9	**5.1**	**6.2**	9.7	6.4	**4.5**
NN	**1.8**	**3.0**	**5.0**	8.8	**4.2**	8.5	6.6	6.1	**5.2**	5.8
SVM	**2.6**	**3.3**	**4.6**	9.8	**2.2**	6.3	6.5	8.7	**2.6**	8.4
Average	2.1 (0.40)	3.1 (0.89)	3.9 (0.85)	8.2 (1.40)	4.4 (1.98)	7.1 (1.76)	7.1 (0.90)	8.3 (1.29)	4.7 (1.52)	6.3 (1.38)

Table 3
Comparison of average rank on AUC among the six different classification methods.

Classification methods	LOG	kNN	C4.5	Ripper	NN	SVM
Average rank	1.60 (0.9101)	4.71 (1.3729)	4.31 (1.2847)	4.30 (1.4320)	3.64 (1.5144)	2.44 (1.1748)

classified as ST companies; *SS*: the number of ST companies correctly classified as ST companies, *SN*: the number of ST companies classified as Non-ST companies.

4. *Area under ROC curve (AUC)*: ROC graph is a two-dimensional graph in which sensitivity is plotted on the *Y* axis and 1-specificity is plotted on *X* axis. An ROC graph depicts relative trade-off between benefits (true positives) and costs (false positives), and it is useful for visualizing the performance of classification models [31]. AUC is a good performance measure especially for the highly imbalanced test sample.

To balance the computation time and performance of WMBGA, the parameters in WMBGA are set empirically as follows. $K = 20, G_{max} = 100, I_u = 30, r_1 = 0.5, r_2 = 0.3$. The logistic regression model shown as Eq. (5) is used to verify the performance of the different feature subset selection methods. All experiments are conducted with Matlab and the EN method is implemented by the Matlab's statistics toolbox. The LOG, kNN ($k = 10$), NN, and SVM are also implemented by Matlab with the default settings. C4.5 and ripper are implemented by Weka [32] under the default settings.

3.3. Empirical results

3.3.1. Performance of feature subset selection methods and classification models

To remove bias from a random selected 180 Non-ST company-year instances in the training sample, 10 groups of different training sample are generated and each group includes the same set of ST company-year instances but a different set of Non-ST company-year instances.

In this study, the performance of a feature subset selection method actually refers to the classification performance of classification models on the test sample with features selected from the training sample by the feature subset selection method. The average performance spe, sen, acc and AUC on the test sample of all the different feature subset selection methods applied on 10 groups of training sample is given in Table 1. The group with the best AUC performance without significant difference is marked in bold. The number in the bracket is the standard deviation of the performance measures on the 10 iterations of test. Since the performance of filter methods and wrapper method based on genetic algorithm is influenced by the parameter of the number of features selected m^*, these two types of methods are trained and tested with different m^* taking values in {1, 2, 3, ..., 10, 15, 20}. The maximum value of m^* is set to 20, because the total number of significant features in regression models reported in previous research [15] is always less than 20 and more variables subsequently introduced to the regression model cannot improve the model performance. Table 1 only presents the best average AUC performance of FVIF and WMBGA with the different values for m^*. For the elastic net method, the parameter α is set to the values in set {1, 0.8, 0.5, 0.1} and parameter λ is optimized by a simple exhaustive searching in the range of [0, maxLamda] with 100 steps and the maxLamda is estimated to be just sufficient to produce all zero coefficients β [28]. The performance of elastic net method reported in Table 1 is the performance of the models which has the greatest AUC among models generated by the EN method, with all different settings of parameters α and λ.

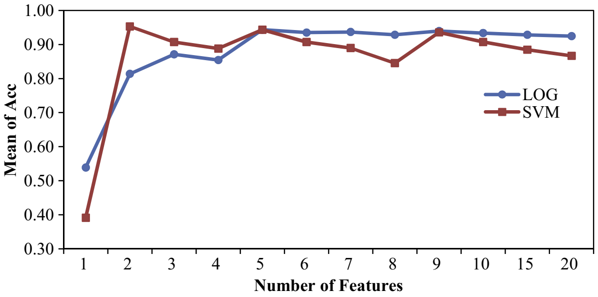


Fig. 4. Mean of Acc of FV-en against different value of parameter m^*.

Since AUC is the most popular performance measure in the area of classification on an imbalanced test sample set, the comparison of AUC performance of different models with different feature subset selection approaches is conducted. A non-parametric equivalent of the repeated-measures ANOVA without assumptions of normal distributions or homogeneity of variance, is used to conduct the global hypothesis test whether the AUC of all feature subset selection methods have difference. The *p* values of Friedman test on the AUC performance for each model under the 10 different features selection approach are approximately equal to 0, so the null hypothesis that the AUC of all feature subset selection methods have no difference is rejected. Then the Nemenyi test can be used to compare all features selection methods with each other. The performance of two treatments is significantly different if the corresponding average ranks differ by at least the criteria difference [33].

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}} \tag{13}$$

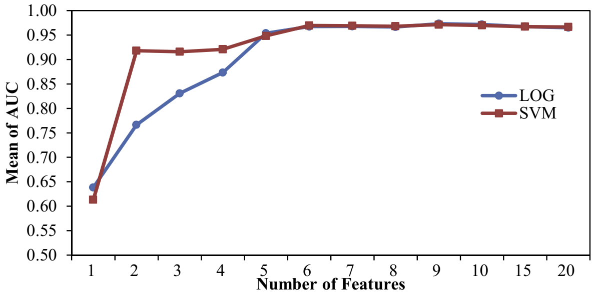


Fig. 5. Mean of AUC of FV-en against different value of parameter m^*.

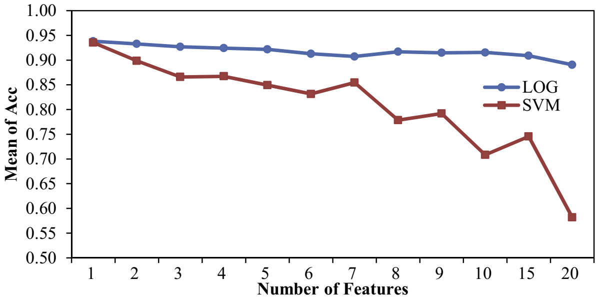


Fig. 6. Mean of Acc of WMBGA against different value of parameter m^*.

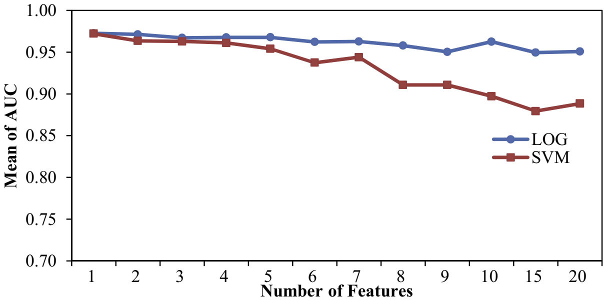


Fig. 7. Mean of AUC of WMBGA against different value of parameter m^*.

where q_α is the critical value for the two-tailed Nemenyi test with significance level α; k is the number of classifier; N is the number of datasets.

The average ranks of AUC for all 10 features selection methods by different models are shown in Table 2. At $\alpha = 0.05, CD = 4.28$. In terms of Nemenyi test, the features selection approaches having no significant difference from the best features selection method are marked in bold for each model. It is interesting to observe that the AUC performance of three different filter methods have no significant difference for all six different models, which shows that they have good robustness. The average rank of FV-en, FV-tt and FV-roc is 2.1, 3.1 and 3.9 respectively and the standard deviation of the rank on $6 \times 10 = 60$ groups of training and tests is no more than 1. Although the average ranks of WMBGA, FFS and FAAS are not significantly different from the three filter methods, they have larger standard deviation on the performance for different models. In average, WMBGA ranks 2.2 on SVM model but ranks 7.9 on Ripper model. FS ranks 2.6 on SVM model but 6.4 on Ripper model.

The average ranks of AUC for six different classification models are shown in Table 3. To compare the AUC performance among the six different classification models, the Nemenyi test is used. Each model has been trained and tested by 10 different sample sets with 10 different features subsets. Therefore, the number of datasets $N = 10 \times 10 = 100,\ k = 6$. If $\alpha = 0.05, q_\alpha = 2.85$ for two-tailed Nemenyi test and $CD = 0.754$. LOG has the best AUC performance among these six methods, followed by SVM, NN, Ripper, C4.5 and kNN in order. The AUC performance difference between LOG and SVM is almost not significant, but both are significant better than other four methods.

The mean of Acc and AUC on the test sample of 10 logistic regression and SVM models trained by 10 different training samples with features selected by FV-en method with different parameter of m^* is shown as Figs. 4 and 5 respectively. Fig. 4 shows that the Acc performance of LOG is tending toward stability when the number of features selected is equal to or greater than 5, while the Acc performance of SVM fluctuates slightly with the change of the number of features selected. Fig. 5 shows that the AUC performance of both LOG and SVM is tending stability when the number of features selected is equal to or greater than 5. The number of features selected in FV-tt, FV-roc has similar effect on the performance of classification models to that of FV-en. Figs. 6 and 7 shows the mean Acc and AUC of LOG and SVM with features selected by WMBGA with different m^*. Perhaps surprisingly, unlike FV-en, the LOG and SVM achieve the best Acc and AUC performance with one unique features selected by WMBGA and the Acc and AUC performance decrease with the increase of m^*. The LOG and SVM models with features selection guided by WMBGA and achieving the best AUC performance are actually one-variable models, which are consistent with the model proposed by Zmijewski [14]. The possible reason is that it is easy for WMBGA to find the optimal unique feature when the size of searching space is 338 and the performance of the optimal unique feature model is good enough. If the $m^* = 5$, the size of searching space is $3.5686 \times 10^{10}$and WMBGA can only evaluate a small proportion of the whole searching space in a limited time.

3.3.2. Feature subsets obtained by FV-en and WMBGA

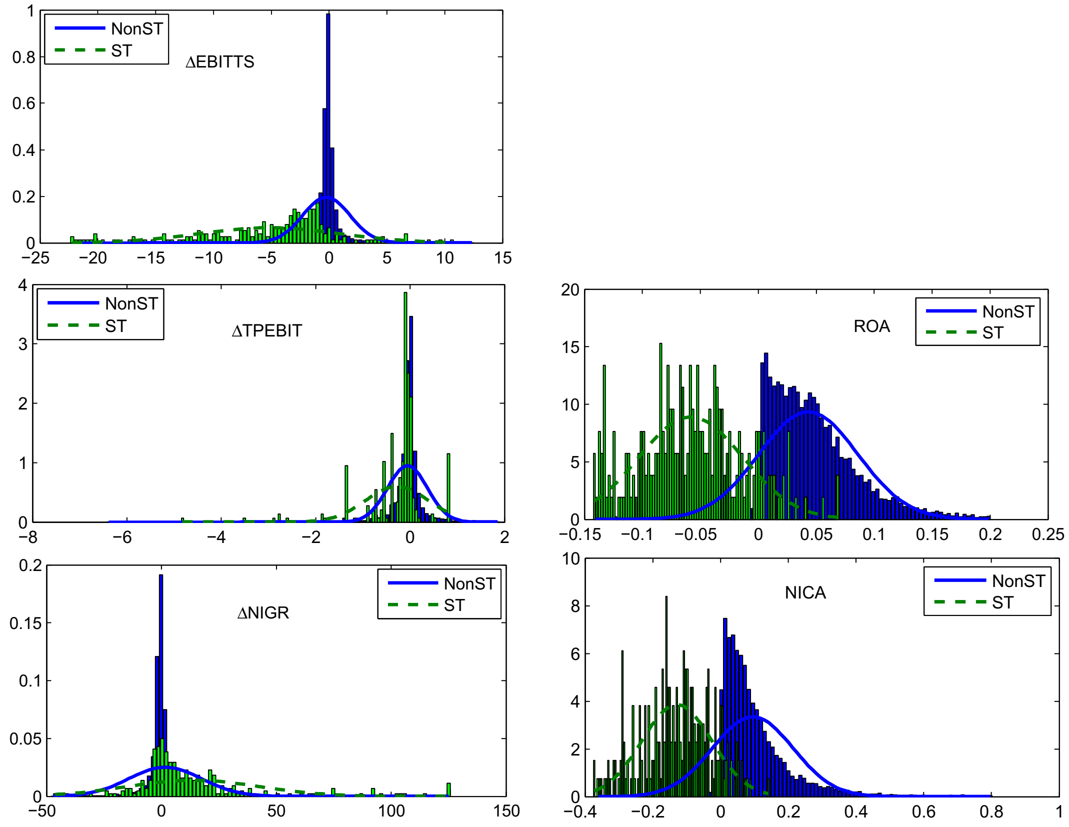
Since LOG models with features selected by FV-en and WMBGA can stably achieve the best AUC, the feature subset that can achieve the best AUC by LOG model in the 10 iterations is studied. When the $m^* = 9$, the mean AUC of LOG with features selected by FV-en in 10 iterations of test obtain the highest value which is greater than 0.97. The FV-en method with $m^* = 9$ is denoted by FV-en-9. It is important to find out what features are mostly selected by FV-en in each of the 10 different groups of training sample and their relationship with the widely used features selected by domain knowledge. Table 4 summarizes the features selected by FV-en-9 no less than 5 times from 10 iterations of tests.

Table 4
The description and frequency of the features selected by FV-en.

ID	Description	Frequency
ΔEBITTS	ΔEBIT/total sales	6
ΔTPEBIT	ΔTotal profit/EBIT	6
ΔNIGR	ΔNet income growth rate	6
ΔROA	ΔROA	5
ΔNITS	ΔNet income/total sales	5
ΔROE	ΔROE	5

Table 5
The description and frequency of the features selected by WMBGA-1.

ID	Description	Frequency
ROA	Net income/total assets (ROA)	4
NICA	Net income/current assets	4
NIACA	Net income/average current assets	1
EPS	Net income/shares of stock	1



(a) FV-en-9

(b) WMBGA-1

Fig. 8. Distribution of features selected by FV-en-9 and WMBGA frequently between ST and Non-ST companies.

Table 6
Comparison of average rank of AUC performance of the top four feature subset selection methods.

Methods	GAFASS	FV-en-9	FA2	WMBGA-1
Average rank	1.00	2.50	4.00	2.50

It is interesting to observe that all these 6 features measure the change of the company's profitability. It shows that change of profitability of a company is a very important factor for predicting a company's financial distress.

Table 5 gives the description and frequency of the features selected by WMBGA with $m^* = 1$ (WMBGA-1). All features selected by WMBGA-1 listed in Table 5 are derived from net income retrieved from a company's financial statements, which is the most important item in measuring a company's profitability. The features found by WMBGA-1 is highly correlated to net income which is consistent with the recent model brought forward by Altman et al. [12] who applied and improved their classical model proposed in 1968 [1] by considering the special characteristics of listed companies in China. The most important variable in the model of Altman et al. [12] is the rate of return on total assets (net income/average total assets).

Fig. 8 shows the distribution of the 5 features having the highest frequency listed in Tables 4 and 5 on all ST and Non-ST sample (both in training sample and test sample). The distribution of each feature fits the normal distribution, although the distribution of some variables does not fit the normal distribution well. From Fig. 8, it can be observed that ROA and NICA found by WMBGA-1 has better discriminative capability between ST and NonST companies than the three features found by FV-en-9. It is because that WMBGA-1 searches the most discriminative feature while FV-en-9 finds a feature subset containing 9 features that can achieve the best discriminative capability and one of them may not have so great discriminative capability as that of features found by WMBGA-1.

3.3.3. Combination of domain knowledge and data mining for feature subset selection

Altman et al. [12] uses net income/total assets to measure the profitability of the company and FV-en and WMBGA-1 also select other ratios listed in Tables 4 and 5 to measure the profitability of a company. To see if the new features found by data mining can help to improve prediction performance for the model from domain knowledge, the NITA ratio in FA2 is substituted by these ratios in Tables 4 and 5 respectively and the new feature subsets are tested in LOG model on 10 different groups of training sample. The Friedman test on the AUC performance of the above 10 feature subsets and feature subset of FA2 shows a statistically significant difference between them; The p-value of the Friedman test is 0.0036. FA2 still has the largest average mean of AUC on 10 iterations of the test.

Both features selected by domain knowledge, such as the feature subset in FA2, and the features selected by WMBGA can contribute to build corporate financial distress prediction model with good performance. Can the combination of domain knowledge and data mining techniques can help to identify the better feature subset for corporate financial distress prediction model? To answer this question, WMBGA features selection algorithm is applied on the union of feature subset of FA1, FA2 and Shumway, i.e. FASS. The WMBGA on FASS obtains a one-variable model which has the largest mean of AUC performance on 10 tests. This one-variable model is denoted by WMBGA-1. The application of GA on FASS is denoted by GAFASS. The Friedman test shows that AUC performance of LOG with features subset obtained by GAFASS, FA2, WMBGA-1 is statistically significantly different. The average rank of 10 tests among these four methods is shown in Table 6. At $\alpha = 0.05$, the Nemenyi test shows that rank of AUC of GAFASS is significantly better than that of FV-en-9, WMBGA-1 and FA2. In addition, the computation time of GAFASS is less than that of WMBGA. On a PC with Intel Core i7-4770R CPU and 8G ram, the computation time of GAFASS and WMBGA with SVM classifier and 5 features is 11.827 and 13.992 seconds respectively. The feature subset selected by GAFASS in each of the 10 iterations of the test includes NITA which is also in FA2.

4. Conclusion

This paper investigates the selection of the feature subset for the corporate financial distress prediction models with domain knowledge and data mining. The empirical results show that the prediction performance of CFDPM with the feature subset selected by data mining can be as good as that by domain knowledge. The features selection guided by both domain knowledge and data mining identify the most important features found by most models for predicting a company's financial distress is the feature ROA measuring the company's profitability. The reason why the feature net ROA dominates other features in the prediction of ST may be that most criteria giving ST to a company are highly related to the profitability of the company, and historically, most ST companies having received ST were due to their poor performance on profitability.

Although the combination of FA2 feature subset and the new features which were obtained by FV-en-9 and WMBGA-1 cannot outperform FA2 significantly, the integration of WMBGA and FAAS, i.e. GAFASS can achieve the best AUC performance among all the feature subset obtained by a unique data mining method or domain knowledge model discussed in this paper, which indicates that the domain knowledge guided data mining can improve the selection of the feature subset for CFDPM.

All features in this study use yearly observation intervals except for the features derived from market information. In practice, different new information about a company is always arrived at any time, such as earning forecast, new product release, and quarterly financial statements. How to use the up-to-date information to improve the accuracy of forecast will be our future research.

References

[1] E.I. Altman, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, J. Finance 23 (1968) 589–609.

[2] C. Charalambous, A. Charitou, F. Kaourou, Comparative analysis of artificial neural network models: application in bankruptcy prediction, Ann. Oper. Res. 99 (2000) 403–425.

[3] R. Wilson, R. Sharda, Bankruptcy prediction using neural networks, Decis. Support Syst. 11 (1994) 545–557.

[4] A. Gepp, K. Kumar, S. Bhattacharya, Business failure prediction using decision trees, J. Forecast. 29 (2009) 536–555.

[5] K.S. Shin, T.S. Lee, H.J. Kim, An application of support vector machines in bankruptcy prediction model, Expert Syst. Appl. 28 (2005) 127–135.

[6] Y.C. Ko, H. Fujita, G.H. Tzeng, An extended fuzzy measure on competitiveness correlation based on WCY 2011, Knowl.-Based Syst. 37 (2013) 86–93.

[7] H.L. Chen, B. Yang, G. Wang, J. Liu, X. Xu, S.J. Wang, D.Y. Liu, A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbor method, Knowl.-Based Syst. 24 (2011) 1348–1359.

[8] M. Divsalar, M.R. Javid, A.H. Gandomi, J.B. Soofi, M.V. Mahmood, Hybrid genetic programming-based search algorithms for enterprise bankruptcy prediction, Appl. Artif. Intell. 25 (2011) 669–692.

[9] A. Verikas, Z. Kalsyte, M. Bacauskiene, A. Gelzinis, Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: a survey, soft computing – a fusion of foundations, Methodol. Appl. 14 (2010) 995–1010.

[10] S. Cho, H. Hong, B.-C. Ha, A hybrid approach based on the combination of variable selection using decision trees and case-based reasoning using the Mahalanobis distance: for bankruptcy prediction, Expert Syst. Appl. 37 (2010) 3482–3488.

[11] F.E. Mays, N. Lynas, Credit Scoring for Risk Managers: The Handbook for Lenders, Thomson, South-Western, 2004.

[12] E. Altman, M. Heine, L. Zhang, J. Yen, Corporate Financial Distress Diagnosis in China, New York University Salomon Center, Working Paper, 2007.

[13] T. Shumway, Forecasting bankruptcy more accurately: a simple hazard model, J. Bus. 74 (2001) 101–124.

[14] M. Zmijewski, Methodological issues related to the estimation of financial distress prediction models, J. Account. Res. 22 (1984) 59–82.

[15] K.P. Ravi, V. Ravi, Bankruptcy prediction in banks and firms via statistical and intelligent techniques – a review, Eur. J. Oper. Res. 180 (2007) 1–28.

[16] C.F. Tsai, Feature selection in bankruptcy prediction, Knowl.-Based Syst. 22 (2009) 120–127.

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

[18] Z. Drezner, G.A. Marcoulides, M.H. Stohs, Financial applications of a tabu search variable selection model, J. Appl. Math. Decis. Sci. 5 (2001) 215–234.

[19] L. Zhou, K.K. Lai, J. Yen, Empirical models based on features ranking techniques for corporate financial distress prediction, Comput. Math. Appl. 64 (2012) 2484–2496.

[20] C. Walsh, Key Management Ratios: Master the Management Metrics that Drive and Control Your Business, Financial Times Management, London, 2003.

[21] S.A. Ross, R. Westerfield, J.F. Jaffe, G.S. Roberts, Corporate Finance, vol. 7, McGraw-Hill/Irwin, Boston, 2002.

[22] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182.

[23] M. Hall, L. Smith, Practical feature subset selection for machine learning, in: Proceedings of the 21st Australian Computer Science Conference, Springer, 1998, pp. 181–191.

[24] I. Guyon, S. Gunn, M. Nikravesh, L.A. Zadeh, Feature Extraction: Foundations and Applications, vol. 207, Springer, 2006.

[25] S. Theodoridis, K. Koutroumbas, Pattern Recognition, second ed., Academic Press, San Diego, USA, 2003.

[26] D.A. Lind, W.G. Marchal, S.A. Wathen, Statistical Techniques in Business & Economics, Mcgraw-Hill, 2012.

[27] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, J. Roy. Stat. Soc. Ser. B 67 (2005) 301–320.

[28] MathWorks, MATLAB: Statistics Toolbox; User's Guide, MathWorks, 2012.

[29] GTA, China Stock Market Financial Database – Financial Indices, Technical Report, 2015.

[30] L. Zhou, Performance of corporate bankruptcy prediction models on imbalanced dataset: the effect of sampling methods, Knowl.-Based Syst. 41 (2013) 16–25.

[31] T. Fawcett, Roc Graphs: Notes and Practical Considerations for Data Mining Researchers, Report, January 2003.

[32] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA data mining software: an update, SIGKDD Explor. 11 (2009) 10–18.

[33] J. Demšar, Statistical comparisons of classifiers over multiple data set, J. Mach. Learn. Res. (2006) 1–30.