Prediction financial distress of firms based on GA-SVM

Fan tongke
Modern Education Technology Center Xi’an International University, Xi’an, 710077, Shaanxi, (CHINA)

ABSTRACT

Study of financial distress prediction problem. There are many variables affecting the financial dilemma, and choosing reasonable input variables is the key to improve the financial distress prediction accuracy. In order to solve the problem that the input variables selection is not reasonable, causing low precision of financial difficulties prediction, the paper put forward a corporate financial distress prediction method based on genetic algorithm and support vector machine (GA-SVM), and selected the realistic data to do the empirical analysis on the model. The experimental results show that the GA-SVM prediction method of enterprise financial distress can improve the financial distress prediction accuracy.

KEYWORDS

Financial distress; Support vector machine; Genetic algorithm; Prediction accuracy.

INTRODUCTION

In the past 30 years, financial crisis has been exerting great influence on enterprise operation, credit institutions, stock investors and even the entire country. Financial crisis is also called financial distress, the worst of which is bankruptcy. As economic exchange becomes more common and market competition fiercer around the world, it is of great significance for an enterprise to establish a prediction model for predicting its financial distress[1].

A financial prediction model is established based on genetic algorithm and support vector machine and financial target data is employed in categorizing into sorts. The initial result shows that this kind of prediction method can achieve sound effect. Compared with support vector machine method used alone, genetic algorithm-support vector machine (GA-SVM) has better prospect in application.

DEFINITION OF FINANCIAL CRISIS

There’s as yet no standard defining whether or not an enterprise is under financial distress. A majority of researchers, however, hold the opinion that an enterprise can be deemed as under financial distress when its profitability has dropped to certain level and its possibility in paying interest and capital is slim. This is reflected in four manners: insufficient liquidity, inadequate rights and interests and capital, be in debt and arrears. In China bankruptcy is often seen as a symbol of financial distress[2]. With regard to empirical research at home, listed companies[3,4] subject to special treatment (ST) are usually seen as going through a financial crisis due to data collection reasons. Here the same definition is adopted. Financial distress prediction is essentially classification, that is, classifying enterprise’s finance into normal and distress based on its past financial standing. Given input variables of financial distress at the time of
i \{ x_1, x_2, \ldots, x_n \}, its correspondent financial standing \( y_i \) either equals 1 or -1, with 1 standing for normal, -1 distress, then financial distress prediction model of the enterprise will be:
\[
\hat{y}_i = f(x_1, x_2, \ldots, x_m) \quad (1)
\]

**GENETIC ALGORITHM AND SUPPORT VECTOR MACHINE**

**Genetic algorithm**

Genetic algorithm (GA) is a kind of search (optimization) algorithm based on natural selection principle and natural genetic mechanism. It simulates life evolution mechanism in nature and achieves optimization of specific target in manual system. The essence of genetic algorithm is to obtain optimal solution or quasi optimal solution through group search technology and evolves generation by generation on the basis of survival of the fittest. The procedures are: appearance of initial group, calculating individual fitness, select the best individuals according to the principle of survival of the fittest, match the best individuals in pairs, randomly cross-over chromosome genes and randomly mutate genes of some chromosomes to generate the next generation of group. Follow this procedure so group evolves generation by generation till final evolution condition is reached\[^5\]. The procedure is as follows:

1. Determine feasible solution region based on specific problem and choose an encoding method; each solution should be representable with numerical or character string.
2. Each solution should have a standard to measure its superiority which is called fitness function which is usually made up of object function.
3. Determine group scale, crossover probability, mutation probability and evolution termination condition.

To make calculation easier, the number of individuals for group of each generation is equal. The larger the group is, the easier it is to find the optimal solution, and due to limitation of operational capability of computer, longer time is also needed. Evolution termination condition refers to the condition at which evolution terminates. It can be set to terminate in certain generation, or determined by finding whether approximate optimal solution meets accuracy standard.

**Support vector machine**

Support vector machine (SVM) is a new data mining technology. It’s a new tool that solves machine learning problem with optimization method. It’s first put forward by V.Vapnik et al. and based on structural risk minimization principle. It tries to improve learning machine’s generalization ability. It has sound generalization performance and classification accuracy. Besides, support vector machine algorithm is convex optimization. Local optimal solution must be global optimal solution. These characteristics are beyond other algorithms including neural network\[^6\]. It has been adopted in financial field in recent years and yielded respectable results. The principle is as follows:

Suppose training set is \( \{(x_i, y_i), i=1,2,\ldots,n, x_i \in \mathbb{R}^n, y_i \in \mathbb{R} \} \). The principle of support vector machine is to find a nonlinear mapping \( \phi(x) \) through which data \( x \) is mapped to high-dimensional feature space \( F \) in which the following estimating function \( f(x) \) is used for linear regression\[^7\].

\[
f(x) = [\omega \phi(x)] + b, \phi: \mathbb{R}^n \to F, \omega \in F \quad (2)
\]

Where: \( \omega \) stands for weight vector, \( b \) stands for deviation value. Approximation problem equals minimum of the following function:

\[
R_{\text{reg}}[f] = R_{\text{emp}}[f] + \lambda \| \omega \|^2 = \sum_{i=1}^{n} c(e_i) + \lambda \| \omega \|^2 \quad (3)
\]

Where: \( R_{\text{reg}}[f] \) stands for expected risk; \( R_{\text{emp}}[f] \) stands for experience risk, \( \lambda \) is constant.

Support vector machine determines regression function on the basis of minimization objective function and statistical theory as well as loss function and structural risk minimization theory.

\[
\begin{align*}
\min \left\{ \frac{1}{2} \| \omega \|^2 + c \sum_{i=1}^{n} (\xi_i^* + \xi_i) \right\} \\
\leq y_i - \omega \phi(x) - b \leq \varepsilon + \xi_i^* \quad (i) \\
(\omega, \phi(x)) + b - y_i \leq \varepsilon + \xi_i \quad (ii) \\
\xi_i, \xi_i^* \geq 0
\end{align*}
\quad (4)
\]
Where: C stands for weight parameter used for balancing the error between complicated term of model and training error term; $\xi^+$, $\xi^-$ are relaxation factors; $\xi$ is insensitive loss function. This problem can be translated into the following dual problem:

$$\begin{align*}
\max \left\{ -\frac{1}{2} \sum_{i,j} (a_i - a_j^*) k(x_i, x_j) + \sum_{i} a_i^* \xi_i^* \right\} \\
\text{s.t.} \quad \sum_{i} a_i = \sum_{i} a_i^* = 0 \quad (C) \\
\quad 0 \leq a_i \leq C, \quad 0 \leq a_i^* \leq C
\end{align*}$$

Solving the above problem obtains support vector machine regression function:

$$f(x) = \sum_{i} (a_i - a_i^*) k(x_i, x) + b$$

Where $k(x_i, x)$ is called kernel function which must meet Mercer condition. Usually Gauss radial basis function $k(x_i, x) = \exp(-\|x_i - x\|^2 / 2\sigma^2)$ is selected. Therefore, to predict financial distress of enterprise, it is necessary to determine parameters $\epsilon, C$ and $\sigma$, which are generally determined on experience. This is no good thing for support vector machine promotion. Therefore, genetic algorithm is introduced here to achieve automatic selection of these parameters.

### GA-SVM MODEL EMPIRICAL ANALYSIS

#### Sample selection

Here companies listed on Shanghai and Shenzhen stock market are chosen as samples, including 150 ones from industrial index, 60 ones from commercial index and 50 companies under special treatment from comprehensive index. Equal numbers of non-ST companies are randomly chosen respectively from the three indexes to constitute sample set. Data of the year prior to special treatment was chosen.

#### Financial ratio selection

Financial ratio is the ratio between two data in the balance sheet which reflects various aspects of enterprise management, like profitability, debt paying ability, operational capability and cash ability. 16 commonly used financial ratios are chosen here with details as shown in TABLE 1.

<table>
<thead>
<tr>
<th>ID</th>
<th>Ratio name</th>
<th>Formula</th>
<th>ID</th>
<th>Ratio name</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID1</td>
<td>Current ratio</td>
<td>Current assets/current debt</td>
<td>ID9</td>
<td>Sales profit ratio</td>
<td>Total profit/net sales</td>
</tr>
<tr>
<td>ID2</td>
<td>Quick ratio</td>
<td>Current assets-inventory-advance/</td>
<td>ID10</td>
<td>Cost-profit ratio</td>
<td>Total profit/cost</td>
</tr>
<tr>
<td></td>
<td></td>
<td>current debt</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID3</td>
<td>Inventory turnover ratio</td>
<td>Sales cost/average inventory</td>
<td>ID11</td>
<td>Current assets turnover ratio</td>
<td>Net prime operating revenue/average</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>accounts current assets balance</td>
</tr>
<tr>
<td>ID4</td>
<td>Receivable turnover ratio</td>
<td>Net prime operating revenue/accounts</td>
<td>ID12</td>
<td>Total assets turnover ratio</td>
<td>Net prime operating revenue/total</td>
</tr>
<tr>
<td></td>
<td></td>
<td>receivable</td>
<td></td>
<td></td>
<td>average assets</td>
</tr>
<tr>
<td>ID5</td>
<td>Assets liabilities ratio</td>
<td>Total debt/total assets</td>
<td>ID13</td>
<td>Sales increase ratio</td>
<td>Sales increase of this month/total</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>sales revenue of last month</td>
</tr>
<tr>
<td>ID6</td>
<td>Return on equity ratio</td>
<td>Total of prime operating profit</td>
<td>ID14</td>
<td>Fixed assets turnover ratio</td>
<td>Net prim operating revenue/total fixed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>shareholders’ equity</td>
<td></td>
<td></td>
<td>assets</td>
</tr>
<tr>
<td>ID7</td>
<td>Capital adequacy ratio</td>
<td>Total of shareholders’ equity/total</td>
<td>ID15</td>
<td>Debt-assets ratio</td>
<td>Total debt/total assets</td>
</tr>
<tr>
<td></td>
<td></td>
<td>of assets</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID8</td>
<td>Debt-to-equity ratio</td>
<td>Total debt/total of shareholders’</td>
<td>ID16</td>
<td>Return on assets ratio</td>
<td>Net profit/total assets</td>
</tr>
<tr>
<td></td>
<td></td>
<td>equity</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 1 : Financial ratios and respective definition**

#### Empirical analysis

Matlab genetic algorithm and support vector machine algorithm tool box is used here. The 16 financial ratios are encoded into 16-bit binary string (with 0 and 1 in the middle of the string). “1” in the string represents characteristic variable chosen from correspondent po-
sition which is combined with support vector machine algorithm to establish financial distress predicting model. SVM is used as object model and tested with test sample. Prediction results are shown in TABLE 2.

**TABLE 2 : Accuracy comparison of two financial distress prediction algorithms (%)**

<table>
<thead>
<tr>
<th>Data sample set</th>
<th>Industrial index</th>
<th>Commercial index</th>
<th>Comprehensive index</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>77.40</td>
<td>80.01</td>
<td>79.22</td>
</tr>
<tr>
<td>GA-SVM</td>
<td>82.45</td>
<td>86.75</td>
<td>84.34</td>
</tr>
</tbody>
</table>

As is shown in Table 2, GA-SVM is more accurate than SVM model being at 84.51% and its generalization ability is impressive, thus it can efficiently identify financial distress and accurately predict the financial standing of the enterprise for the foreseeable future based on its financial data.

**CONCLUSION**

The application of genetic algorithm and support vector machine in terms of enterprise financial management is studied and on the basis of which a financial distress prediction model is established with SVM as contrast model. Result shows that GA-SVM model is quite accurate in prediction and has impressive generalization ability, thus is of practical significance in enterprise financial distress prediction research.

**REFERENCE**


