

Research Article

Construction and Application Research of Isomap-RVM Credit Assessment Model

Guangrong Tong and Siwei Li

School of Economics and Management, Wuhan University, Wuhan 430072, China

Correspondence should be addressed to Guangrong Tong; tonggr2014@126.com

Received 5 November 2014; Accepted 10 January 2015

Academic Editor: Honglei Xu

Copyright © 2015 G. Tong and S. Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Credit assessment is the basis and premise of credit risk management systems. Accurate and scientific credit assessment is of great significance to the operational decisions of shareholders, corporate creditors, and management. Building a good and reliable credit assessment model is key to credit assessment. Traditional credit assessment models are constructed using the support vector machine (SVM) combined with certain traditional dimensionality reduction algorithms. When constructing such a model, the dimensionality reduction algorithms are first applied to reduce the dimensions of the samples, so as to prevent the correlation of the samples' characteristic index from being too high. Then, machine learning of the samples will be conducted using the SVM, in order to carry out classification assessment. To further improve the accuracy of credit assessment methods, this paper has introduced more cutting-edge algorithms, applied isometric feature mapping (Isomap) for dimensionality reduction, and used the relevance vector machine (RVM) for credit classification. It has constructed an Isomap-RVM model and used it to conduct financial analysis of China's listed companies. The empirical analysis shows that the credit assessment accuracy of the Isomap-RVM model is significantly higher than that of the Isomap-SVM model and slightly higher than that of the PCA-RVM model. It can correctly identify the credit risks of listed companies.

1. Introduction

By constructing an accurate and reliable credit assessment model, we can conduct an in-depth analysis of the financial data of listed companies and identify the financial risks of such companies. This is of great significance to the operational decisions of shareholders, corporate creditors, and management. Along with the development of large data methods, the theory of applying machine learning methods to construct credit assessment models has become increasingly reliable. Machine learning methods are superior to traditional multivariate discriminant analysis methods and logistic discriminant analysis models, when processing any data subject to less stringent hypothetical restrictions and dealing with nonlinear relationships. At present, commonly used machine learning methods with good data classification results include the neural network (NN) [1] and the support vector machine (SVM). Odom and Sharda [2] built an early warning model for financial crises. They did so by

applying the neural network and comparing this model with the Fisher multivariate discriminant analysis model. The research results showed that the artificial neural network has higher prediction accuracy and robustness. However, the neural network has methodological defects, such as slow convergence rate, overfitting, and falling into local minima [3, 4]. Later, Cortes and Vapnik [5] proposed the vector machine method, which is based on the theory of statistical machine learning. This method laid stress on structural risk minimization and is able to effectively overcome the defects of the neural network. Min and Lee [6] applied the SVM to build a financial early warning model for the purpose of corporate bankruptcy prediction. The results showed that the SVM has higher discriminant analysis accuracy than BP neural network, MDA, and logit models. However, SVM has the following main defects: the penalty parameter C must be determined in the model building process, and the selection of kernel function must comply with "Mercer's theorem" [7]. For these reasons, this paper suggests using a relevance vector

machine (RVM) to overcome the defects of SVM. RVM is another efficient supervised learning method proposed by Tipping [8]. By applying this method to conduct machine learning under the Bayesian theory, the model obtained will be sparser than with SVM, and the result probability output can also be obtained. While properly maintaining all the advantages of SVM, this method has reduced the inaccurate assignment of key parameters, broadened the application scope of vector machines, provided a greater degree of freedom, and effectively overcome the defects of SVM [9]. This will help to improve the accuracy of vector machine classification.

In credit risk assessment samples, there tends to be a close correlation between the selected financial risk characteristic indices. High dimensionality and high correlation of the sample characteristic indices may have a strong impact on the accuracy of risk assessment. Therefore, a data dimensionality reduction method is required for the pretreatment of the sample indices, so as to reflect the main features of the data as much as possible and reduce correlation between the characteristic indices. At present, the principal component analysis method (PCA) is one of the most widely used methods [10]. However, as a linear dimensionality reduction method, PCA may not achieve satisfactory dimensionality reduction results when applied to nonlinear data. Therefore, this paper attempts to apply a nonlinear dimensionality reduction method—*isometric mapping (Isomap)*—to conduct dimensionality reduction pretreatment on the sample data. Isomap is a nonlinear manifold learning algorithm proposed by Tenenbaum et al. [11]. By seeking low-dimensional embedding among high-dimensional manifolds, this algorithm has maintained low-dimensional embedding in the neighborhood structure between high-dimensional manifold data points, while producing excellent robustness and global optimality. Lin et al. [12] used the Isomap-SVM, PCA-SVM, and SVM separately to conduct risk assessment classification of more than one hundred listed Taiwan companies. It proved that Isomap-SVM has the highest prediction accuracy. This research showed that, in the process of nonlinear data classification, Isomap can improve accuracy through reasonable dimensionality reduction. Ribeiro et al. [13] constructed a Semi-Supervised Isomap model with Isomap and SVM. They used the Semi-Supervised Isomap model, SVM, RVM, and KNN separately to conduct bankruptcy prediction of more than one thousand industrial French companies. The results showed that the classification accuracy of Semi-Supervised Isomap model is comparable to SVM and RVM. But in the study they did not propose the idea of constructing a new model by combining Isomap with other machine learning methods such as RVM.

By applying the Isomap method to reduce the dimensions and avoid correlation between sample characteristic indices and by using the RVM to conduct the classification, this paper builds an Isomap-RVM model. It conducts a credit assessment classification analysis of China's listed companies. It then compares this model with the Isomap-SVM and PCA-RVM models and examines Isomap-RVM model discriminant analysis accuracy.

2. Isomap-RVM Model Construction

2.1. RVM Classification Method. Assume a training set of n samples, where its characteristic index part constitutes a vector set $\{\mathbf{x}_i\}$, $i = 1, 2, \dots, n$; the corresponding output target value of the training set is $\mathbf{t} = [t_1, t_2, \dots, t_n]^T$, where t_i may only be either 0 or 1, which corresponds to two categories. Like the support vector machine, \mathbf{x} has the following nonlinear function expression:

$$y(\mathbf{x}; \mathbf{w}) = \sum_{i=1}^n w_i K(\mathbf{x}, \mathbf{x}_i) + w_0, \quad (1)$$

where $\mathbf{w} = [w_0, w_1, \dots, w_n]^T$ stands for the weight, while $K(\mathbf{x}, \mathbf{x}_i)$ stands for the kernel function selected. There are four commonly used kernel functions, as follows.

Linear kernel:

$$K(a, b) = a \cdot b. \quad (2)$$

Polynomial kernel:

$$K(a, b) = [(a \cdot b) + 1]^c. \quad (3)$$

Neural kernel:

$$K(a, b) = \tanh(\eta(a \cdot b) - \zeta). \quad (4)$$

RBF kernel:

$$K(a, b) = \exp\left(-\frac{\|a - b\|^2}{2\sigma^2}\right). \quad (5)$$

In this study, radial basis function (RBF) is adopted. The reason is that, compared with other kernel functions, RBF can be applied to conduct nonlinear mapping without substantially increasing the complexity of the model.

After constructing $y(\mathbf{x}; \mathbf{w})$, the correspondence between \mathbf{t} and \mathbf{x} can be expressed by the following sigmoid function:

$$p(t_i = 1 | \mathbf{w}) = \frac{1}{1 + e^{-y(\mathbf{x}_i; \mathbf{w})}}. \quad (6)$$

Then, provided that the weight \mathbf{w} is known, the probability of \mathbf{t} is as follows:

$$p(\mathbf{t} | \mathbf{w}) = \prod_{i=1}^n \left[\frac{1}{1 + e^{-y(\mathbf{x}_i; \mathbf{w})}} \right]^{t_i} \left[1 - \frac{1}{1 + e^{-y(\mathbf{x}_i; \mathbf{w})}} \right]^{1-t_i}. \quad (7)$$

After calculating the best weight \mathbf{w}_M , substitute \mathbf{w}_M and the sample testing set characteristic index \mathbf{x}_* into (6); we will get the probability of the testing set output target value t_* .

To avoid overlearning due to the fact that most weights are not zero, it can be assumed that \mathbf{w} is subject to a normal distribution:

$$\mathbf{w} \sim N\left(0, \frac{1}{\alpha}\right), \quad (8)$$

where $\alpha = [\alpha_0, \alpha_1, \dots, \alpha_n]^T$.

The prior probability of weight is expressed as

$$p(\mathbf{w} | \boldsymbol{\alpha}) = \prod_{i=0}^n (2\pi)^{-1/2} \alpha_i \exp\left(-\frac{\alpha_i w_i^2}{2}\right). \quad (9)$$

The posterior probability $p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha})$ of weight is expressed as

$$p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}) = \frac{p(\mathbf{t} | \mathbf{w}) p(\mathbf{w} | \boldsymbol{\alpha}) p(\boldsymbol{\alpha})}{p(\boldsymbol{\alpha}, \mathbf{t})}. \quad (10)$$

The best weight \mathbf{w}_M is the weight that maximizes the posterior probability $p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha})$. Provided that $\boldsymbol{\alpha}$ is known, then

$$p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}) \propto p(\mathbf{t} | \mathbf{w}) p(\mathbf{w} | \boldsymbol{\alpha}). \quad (11)$$

\mathbf{w}_M can be expressed as follows:

$$\begin{aligned} \mathbf{w}_M &= \arg \max_{\mathbf{w}} p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}) = \arg \max_{\mathbf{w}} p(\mathbf{t} | \mathbf{w}) p(\mathbf{w} | \boldsymbol{\alpha}) \\ &= \arg \max_{\mathbf{w}} \lg [p(\mathbf{t} | \mathbf{w}) p(\mathbf{w} | \boldsymbol{\alpha})]. \end{aligned} \quad (12)$$

Let

$$\begin{aligned} z_i &= \frac{1}{1 + e^{-\gamma(\mathbf{x}_i; \mathbf{w})}}, \\ \mathbf{z} &= [z_1, z_2, \dots, z_n]^T. \end{aligned} \quad (13)$$

Then $\lg[p(\mathbf{t} | \mathbf{w}) p(\mathbf{w} | \boldsymbol{\alpha})]$ is expressed as follows:

$$\begin{aligned} \lg [p(\mathbf{t} | \mathbf{w}) p(\mathbf{w} | \boldsymbol{\alpha})] &= \sum_{i=1}^n [t_i \lg z_i + (1 - t_i) \lg (1 - z_i)] \\ &\quad - \frac{\mathbf{w}^T A \mathbf{w}}{2}, \end{aligned} \quad (14)$$

where $A = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_n)$.

Use the Newton iteration method to seek \mathbf{w}_M in (14):

$$\begin{aligned} \Delta \mathbf{w} &= - \{ \nabla_{\mathbf{w}} \nabla_{\mathbf{w}} \lg [p(\mathbf{t} | \mathbf{w}) p(\mathbf{w} | \boldsymbol{\alpha})] \}^{-1} \\ &\quad \cdot \{ \nabla_{\mathbf{w}} \lg [p(\mathbf{t} | \mathbf{w}) p(\mathbf{w} | \boldsymbol{\alpha})] \} \\ &= (\phi^T B \phi + A) [\phi^T (\mathbf{t} - \mathbf{z}) - A \mathbf{w}], \end{aligned} \quad (15)$$

where $B = \text{diag} [z_1(1 - z_1), z_2(1 - z_2), \dots, z_n(1 - z_n)]$,

$$\phi = \begin{bmatrix} 1 & K(\mathbf{x}_1, \mathbf{x}_1) & K(\mathbf{x}_1, \mathbf{x}_2) & \cdots & K(\mathbf{x}_1, \mathbf{x}_n) \\ 1 & K(\mathbf{x}_2, \mathbf{x}_1) & K(\mathbf{x}_2, \mathbf{x}_2) & \cdots & K(\mathbf{x}_2, \mathbf{x}_n) \\ & & \vdots & & \\ 1 & K(\mathbf{x}_n, \mathbf{x}_1) & K(\mathbf{x}_n, \mathbf{x}_2) & \cdots & K(\mathbf{x}_n, \mathbf{x}_n) \end{bmatrix}. \quad (16)$$

When $\Delta \mathbf{w}$ is obtained, \mathbf{w}_M and $\boldsymbol{\alpha}$ should be updated by formula

$$\begin{aligned} \mathbf{w}_M^{\text{new}} &= \mathbf{w}_M + \Delta \mathbf{w}, \\ \alpha_i^{\text{new}} &= \frac{1 - \alpha_i \left((\phi^T B \phi + A)^{-1} \right)_{ii}}{(\mathbf{w}_M^{\text{new}})_i^2}. \end{aligned} \quad (17)$$

Substitute the new value back into (15) to calculate $\Delta \mathbf{w}$ over and over again until $\Delta \mathbf{w}$ is reduced to a level that is below a certain limit; then we can substitute \mathbf{w}_M into (6) to calculate the accurate classification probability and determine the sample category based on the classification probability $p(t_i = 1 | \mathbf{w})$.

2.2. Dimensionality Reduction Method Based on Isomap Manifold Learning. Isomap is developed on the basis of multi-dimensional scaling (MDS). It aims at seeking out the low-dimensional coordinates embedded in the high-dimensional space and achieves dimensionality reduction by constructing the shortest path distance matrix for high-dimensional sample data, under the assumption that the intrinsic geometric properties of the sample data remain unchanged [14]. The dimensionality reduction steps are shown below.

Step 1. Construct a neighborhood graph G . With respect to the characteristic index vector set $\{\mathbf{x}_i\}$, $i = 1, 2, \dots, m$, for a total number of m samples in the training set and the testing set, each original sample \mathbf{x}_i has N characteristic indices $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iN})$. The Euclidean distance $d(\mathbf{x}_i, \mathbf{x}_j)$ between each \mathbf{x}_i and each of the remaining vectors in the vector set can be calculated by the following formula:

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{k=1}^N (x_{ik} - x_{jk})^2}. \quad (18)$$

When \mathbf{x}_j is one of the q vectors closest to \mathbf{x}_i , they should be deemed as neighboring vectors. Such neighboring $q + 1$ vectors will together constitute an undirected graph G ; the connection between \mathbf{x}_i and its neighboring vector will constitute the boundary of graph G , where the boundary value is the distance between any two vectors constituting the boundary.

Step 2. Construct the shortest path distance matrix. Apply Dijkstra's algorithm to calculate the shortest path distance between any two vectors on the neighborhood graph G .

Above all, for any two vectors \mathbf{x}_i and \mathbf{x}_j , the boundary value should be directly taken as the distance when they constitute the boundary, or the distance may be assumed to be ∞ when constituting no boundary at all; that is,

$$\begin{aligned} d_0(\mathbf{x}_i, \mathbf{x}_j) &= \begin{cases} d(\mathbf{x}_i, \mathbf{x}_j) & \mathbf{x}_i, \mathbf{x}_j \text{ constitute the boundary} \\ \infty & \text{others.} \end{cases} \end{aligned} \quad (19)$$

Then, find out all the indirect paths between \mathbf{x}_i and \mathbf{x}_j along the boundary, where the path distance should be the sum of boundary values (denoted as d_1, d_2, \dots, d_L); next, compare all the path distances and find the shortest distance $d_G(\mathbf{x}_i, \mathbf{x}_j)$:

$$d_G(\mathbf{x}_i, \mathbf{x}_j) = \min \{d_0(\mathbf{x}_i, \mathbf{x}_j), d_1, \dots, d_L\}. \quad (20)$$

Use the squared value of the shortest path distance between each pair of vectors in G to create the shortest path distance matrix $D_G = [d_G^2(\mathbf{x}_i, \mathbf{x}_j)]_{m \times m}$.

Step 3. Apply a classical MDS algorithm to compress the vector to a d -dimensional vector.

Let

$$\begin{aligned} \xi_{ij} &= \begin{cases} 0 & i \neq j \\ 1 & i = j, \end{cases} \\ u_{ij} &= \xi_{ij} - \frac{1}{m}, \\ U &= (u_{ij})_{m \times m}. \end{aligned} \quad (21)$$

Calculate the matrix τ :

$$\tau = -\frac{UD_GU}{2}. \quad (22)$$

Eigenvectors u_1, u_2, \dots, u_d corresponding to a maximum of d eigenvalues ($\lambda_1, \lambda_2, \dots, \lambda_d$) of τ will constitute the eigenvector matrix $U = [u_1, u_2, \dots, u_d]$. Then, the dimensionality reduction result X_d is shown as follows:

$$X_d = \text{diag}(\lambda_1^{1/2}, \lambda_2^{1/2}, \dots, \lambda_d^{1/2})U^T. \quad (23)$$

The dimensionality reduction error is

$$\text{Er} = 1 - R^2(S_G, D_{X_d}), \quad (24)$$

where R^2 is the correlation coefficient; $S_G = [d_G(\mathbf{x}_i, \mathbf{x}_j)]_{m \times m}$; D_{X_d} is the Euclidean distance matrix of the vectors in the d -dimensional space.

The reduced dimensionality d may be determined by observing the Er error curve. When there is an inflection point on the error curve or the Er turns to be stable and sufficiently small, the dimensionality used is the optimal dimensionality d .

2.3. Isomap-RVM Model Classification Steps.

Step 1. Conduct normalized pretreatment of the sample data. Normalize different characteristic indices into a $[0, 1]$ interval, maintain the commensurability of different indices during operation, and improve the operation speed. For m samples, the normalization formula for a sample index is

$$\bar{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \quad (25)$$

where, $i = 1, 2, \dots, m$, \bar{x}_i is the value of the normalized characteristic index of the i th sample; x_i is the value of the original characteristic index of the i th sample; x_{\max} and x_{\min} are the maximum and minimum values, respectively, of the characteristic index in all samples.

Step 2. Apply Isomap to reduce the dimensions of the normalized sample data, reduce the correlation between the sample characteristic indices, and improve the sample quality.

Step 3. Train the training set data and apply the genetic algorithm to optimize the kernel width of RVM.

TABLE 1: Sample size of listed companies.

ST implementation year	2009	2010	2011	2012	2013	Total
ST companies	26	36	13	22	19	116
Non-ST companies	78	108	39	66	57	348

Step 4. Substitute the optimized kernel width, sample training set data, and the characteristic index part of testing set data into RVM for classification and get the final result.

3. Empirical Analysis

3.1. Samples and Indices. Special treatment (ST) is used to signify listed companies that have undergone a financial crisis in China's securities market. In research, ST companies are generally deemed as companies with poor credit standing, while non-ST companies are generally deemed as companies with normal credit standing. A listed company may be declared as an ST company in year t based on the financial report for year $t - 1$; hence using the financial data of year $t - 1$ to predict whether such listed company will be declared as an ST company in year t will overestimate the assessment capability of the model. For this reason, this paper has adopted the financial data of year $t - 2$ to predict whether such listed company will be declared as an ST company in year t . The raw data are derived from the financial index database on CSMAR Chinese listed companies. This paper has selected 116 companies subjected to ST between 2009 and 2013 in this database. Such new ST sample companies exclude those declared as ST companies due to nonfinancial reasons. 348 non-ST sample companies in the same industry for the corresponding year were selected, based on the ratio of 1:3. Altogether 464 sample companies were chosen for the empirical analysis. The sample size is shown in Table 1.

Sample category label is expressed as 0 or 1, where 0 means that the sample is an ST sample and 1 means that it is a non-ST sample. The sample characteristic index is derived from the financial index data two years before the corresponding year (i.e., 2007~2011) of each sample. This paper has selected 25 indices as the sample for estimating the characteristic index of a company's credit standing (as shown in Table 2) based on the following 8 aspects of the listed companies: cash flow capacity, return on equity, earning capacity, operating capacity, growth capacity, risk level, short-term solvency, and long-term solvency [15–18].

The 464 samples were randomly grouped into the training set and testing set. This paper classified 320 samples into the training set (accounting for about 69% of the total sample size), including 80 ST companies and 240 non-ST companies. 144 samples were classified into the testing set (accounting for about 31% of the total sample size), including 36 ST companies and 108 non-ST companies.

3.2. Empirical Results. This paper has established three credit assessment models—Isomap-RVM, Isomap-SVM, and PCA-RVM using MATLAB software. The aim is to examine

TABLE 2: Sample characteristic index.

Type	Variables	Name of index
Cash flow capacity	X1	Cash Interest Cover
	X2	Ratio of Cash to Operating Revenue
Return on equity	X3	Earnings per Share
	X4	<i>P/B</i> Ratio
	X5	Net Assets per Share
Earning capacity	X6	Return on Equity (ROE)
	X7	Return on Assets (ROA)
	X8	Net Profit Margin of Fixed Assets
Operating capacity	X9	Accounts Receivable Turnover
	X10	Inventories Turnover
	X11	Ratio of Current Assets to Income
	X12	Current Assets Turnover
	X13	Total Assets Turnover
Growth capacity	X14	Capital Preservation & Appreciation Rate
	X15	Growth Rate of Fixed Assets
	X16	Growth Rate of Total Assets
	X17	Growth Rate of Operating Revenue
Risk level	X18	Financial Leverage
Short-term solvency	X19	Current Ratio
	X20	Quick Ratio
	X21	Cash Ratio
	X22	Ratio of Working Capital to Net Assets
Long-term solvency	X23	Ratio of Liabilities to Total Assets
	X24	Times Interest Earned
	X25	Ratio of Debt to Equity

the classification accuracy of RVM and SVM by comparing Isomap-RVM with Isomap-SVM and to investigate the effect of different dimensionality reduction methods on the classification result by comparing Isomap-RVM with PCA-RVM. Toolboxes IsomapR1 [19] and SBI_Release_110 [20] were used when writing the PCA-RVM and Isomap-RVM; toolboxes IsomapR1 and libsvm-mat-2.89-3 [21] were used when writing the Isomap-SVM. The kernel width of RVM and parameters C and σ of SVM were optimized using the genetic algorithm.

This paper adopted a default value of 7 for the adjacency parameter q of the Isomap algorithm. Substituting this value into the sample data, we can get the residual error curve as shown in Figure 1. Figure 1 shows that the Er error, starting from the dimensionality of 12, begins to become stable from below 0.05. Thus, the reduced dimensionality of 12 was chosen.

After the data had been normalized, it was input into the SPSS, and PCA analysis was conducted. When the main component is extracted up to 12, the accumulating contribution rate will reach 83.968%. Therefore, this paper has selected information covering 85% of the original indices (i.e., the accumulating contribution rate = 85%) to conduct PCA dimensionality reduction and used the Isomap dimensionality reduction model to compare the results.

During the training process, the vector machine kernel parameters for the three models were optimized and acquired using the genetic algorithm (GA), where the RVM kernel width of the Isomap-RVM model was 1.457; the RVM kernel width of the PCA-RVM model was 0.534; and, for the Isomap-SVM model, $C = 3.9174$, $\sigma = 1.7073$.

The prediction results of the credit assessment are determined by the prediction accuracy of the test sample and are often measured by two types of errors. Type I error occurs when mistakenly classifying any ST company as a non-ST company; Type II error occurs when mistakenly classifying any non-ST company as an ST company. The results of prediction on the testing set by the three models are shown in Table 3.

It turned out that Isomap-RVM has the lowest total false rate (i.e., 9.72%), followed by PCA-RVM (10.41%). Isomap-SVM has the highest total false classification rate (i.e., 13.89%). The probability of a Type II error occurring in any of the three models is minimal, which indicates that these three models can all correctly identify any non-ST company and will not classify it as an ST company. This is because non-ST companies have normal financial data and are more easily distinguished than those classified as ST companies due to significant deterioration of financial data. The false classification gap among these three models

TABLE 3: False results of prediction on the testing set by the three models.

Model	Type I error	Type II error	Total false
Isomap-RVM	12/36 (33.33%)	2/108 (1.85%)	14/144 (9.72%)
Isomap-SVM	17/36 (47.22%)	3/108 (2.78%)	20/144 (13.89%)
PCA-RVM	11/36 (30.56%)	4/108 (3.7%)	15/144 (10.41%)

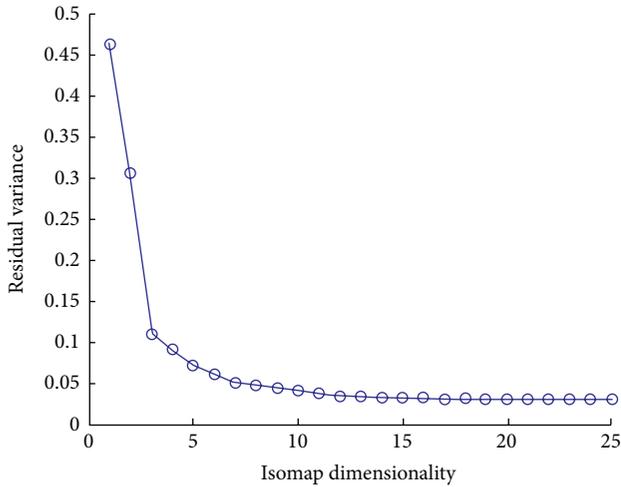


FIGURE 1: Er error curve.

is mainly manifested by the number of occurrences of Type I error.

By comparing the false results of Isomap-RVM and Isomap-SVM after dimensionality reduction, we can discover that, for the same data, the classification accuracy of the RVM model is significantly higher than that of the SVM model. In particular, the occurrence rate of Type I errors using the SVM method is 13.89% higher than when using RVM. Compared with the RVM model, the SVM model is more likely to classify any ST company as a non-ST company. This will underestimate the risk of financial deterioration of listed companies and may not help the companies and investors effectively circumvent and spread their financial risk.

By comparing the classification results of Isomap-RVM and PCA-RVM, we can discover that the classification accuracy of these two models is quite close. The total false classification rate of Isomap-RVM is lower than that of PCA-RVM, but the Isomap-RVM had one more Type I error occurrence than the PCA-RVM. These results may not indicate that using Isomap instead of PCA to reduce dimensions can better help improve the classification accuracy of the model, but it still shows that, for nonlinear data, Isomap can also be deemed as a reliable dimensionality reduction method for reducing data correlation.

4. Summary and Conclusions

This paper has constructed a credit assessment model—Isomap-RVM. Compared with existing credit assessment models, the methods applied by the new model have the

following two major advantages: while inheriting advantages of SVM, RVM has met the Mercer conditions of the kernel function and avoided the penalty parameter selection process; Isomap is suitable for dimensionality reduction of nonlinear samples. Isomap-RVM uses Isomap to reduce the dimensions of the sample so as to reduce the sample correlation and uses RVM to conduct the training and classification prediction so as to create a new machine learning model and achieve better credit assessment results.

The empirical results show that when using the Isomap-RVM model to conduct a credit assessment of listed companies, the accuracy is significantly higher than when using the Isomap-SVM model and slightly higher than when using the PCA-RVM model. In all cases (i.e., total false classification, Type I error, or Type II error), the classification error occurrence rate of the RVM method is lower than that of the SVM method. This demonstrates that the RVM method performs better than the SVM method in terms of credit assessment.

From the perspective of characteristic index dimensionality reduction, the total false classification rate of the RVM method using Isomap to reduce dimensions is lower than that of the RVM method using PCA to reduce dimensions, yet Type I error occurrence rate of the former is higher than that of the latter. The classification accuracy gap between the two methods is minimal. Since the degree of dimensionality reduction effect is related to the sample data structure, the reasons for this result may be that the nonlinear structure of the sample selected in this research is not clear enough. However, the empirical results still show that, just like PCA, Isomap can also reliably achieve data dimensionality reduction and thereby improve the assessment precision of the model.

There are two issues which will require further investigation in subsequent research. Firstly, only China's listed companies were selected as the sample in this research, so the sample data structure lacked diversification. Hence, subsequent research may select company data from other countries for analysis. It may also further analyze the effect on model accuracy when applying the Isomap method and the PCA method to RVM. Secondly, the optimization algorithm for the RVM kernel width is not limited to GA. Other parameter optimization algorithms may be applied to conduct research on how to further improve the assessment accuracy of the model.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] S. Dutta and S. Shekhar, "Bond rating: a nonconservative application of neural networks," in *Proceedings of the IEEE International Conference on Neural Network*, pp. 443–450, San Diego, Calif, USA, July 1988.
- [2] M. D. Odom and R. Sharda, "A neural network model for bankruptcy prediction," in *Proceedings of the International Joint Conference on Neural Prediction Networks (IJCNN '90)*, pp. 163–168, 1990.
- [3] K. G. Coleman, T. J. Graettinger, and W. F. Lawrence, "Neural networks for bankruptcy prediction: the power to solve financial problems," *AI Review*, vol. 4, pp. 48–50, 1991.
- [4] G. Zhang, M. Y. Hu, B. E. Patuwo, and D. C. Indro, "Artificial neural networks in bankruptcy prediction: general framework and cross-validation analysis," *European Journal of Operational Research*, vol. 116, no. 1, pp. 16–32, 1999.
- [5] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [6] J. H. Min and Y.-C. Lee, "Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters," *Expert Systems with Applications*, vol. 28, no. 4, pp. 603–614, 2005.
- [7] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*, Cambridge University Press, Cambridge, UK, 2000.
- [8] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, no. 3, pp. 211–244, 2001.
- [9] N. Hazarika and J. G. Taylor, "Predicting bonds using the linear relevance vector machine," in *Neural Networks and the Financial Markets*, Perspectives in Neural Computing, pp. 145–155, Springer, London, UK, 2002.
- [10] Y. Chen-Guang and D. Xiao-Bo, "Credit risk assessment in commercial banks based on SVM using PCA," in *Proceedings of International Conference on Machine Learning and Cybernetics*, vol. 2, pp. 1207–1211, IEEE, Kunming, China, July 2008.
- [11] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [12] F. Lin, C.-C. Yeh, and M.-Y. Lee, "The use of hybrid manifold learning and support vector machines in the prediction of business failure," *Knowledge-Based Systems*, vol. 24, no. 1, pp. 95–101, 2011.
- [13] B. Ribeiro, A. Vieira, J. Duarte et al., "Learning manifolds for bankruptcy analysis," in *Advances in Neuro-Information Processing*, vol. 5506 of *Lecture Notes in Computer Science*, pp. 723–730, Springer, Berlin, Germany, 2009.
- [14] F. Ratle, A. L. Terrettaz, M. Kanevski, P. Esseiva, and O. Ribaux, "Learning manifolds in forensic data," in *Artificial Neural Networks—ICANN 2006*, vol. 4132 of *Lecture Notes in Computer Science*, pp. 894–903, Springer, Berlin, Germany, 2006.
- [15] W. H. Beaver, "Financial ratios as predictors of failure empirical research in accounting: selected studies," *Journal of Accounting Research*, vol. 4, supplement, pp. 179–199, 1966.
- [16] E. B. Deakin, "A discriminant analysis of predictors of business failure," *Journal of Accounting Research*, vol. 10, no. 1, pp. 167–179, 1972.
- [17] D. Martens, L. Bruynseels, B. Baesens, M. Willekens, and J. Vanthienen, "Predicting going concern opinion with data mining," *Decision Support Systems*, vol. 45, no. 4, pp. 765–777, 2008.
- [18] J. A. Ohlson, "Financial ratios and the probabilistic prediction of bankruptcy," *Journal of Accounting Research*, vol. 18, no. 1, pp. 109–131, 1980.
- [19] <http://isomap.stanford.edu>.
- [20] <http://www.relevancevector.com>.
- [21] <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

