Knowledge-Based Systems 24 (2011) 95–101

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

The use of hybrid manifold learning and support vector machines in the prediction of business failure

Fengyi Lin [a], Ching-Chiang Yeh [b,*], Meng-Yuan Lee [c]

[a] Department of Business Management, National Taipei University of Technology, Taipei, Taiwan
[b] Department of Business Administration, National Taipei College of Business, Taipei, Taiwan
[c] Institute of Commerce Automation and Management, National Taipei University of Technology, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 8 September 2009
Received in revised form 26 June 2010
Accepted 18 July 2010
Available online 24 July 2010

Keywords:
Manifold learning
ISOMAP
Support vector machine
Business failure

ABSTRACT

The prediction of business failure is an important and challenging issue that has served as the impetus for many academic studies over the past three decades. This paper proposes a hybrid manifold learning approach model which combines both isometric feature mapping (ISOMAP) algorithm and support vector machines (SVM) to predict the failure of firms based on past financial performance data. By making use of the ISOMAP algorithm to perform dimension reduction, is then utilized as a preprocessor to improve business failure prediction capability by SVM. To create a benchmark, we further compare principal component analysis (PCA) and SVM with our proposed hybrid approach. Analytic results demonstrate that our hybrid approach not only has the best classification rate, but also produces the lowest incidence of Type II errors, and is capable of achieving an improved predictive accuracy and of providing guidance for decision makers to detect and prevent potential financial crises in the early stages.

1. Introduction

The prediction of business failure is an important and challenging issue that has served as the impetus for many academic studies over the past three decades [2,5,8,11,36,45,63]. Business failure is a general term and, according to widespread definition, is the situation in which a firm cannot pay lenders, preferred stock shareholders, suppliers, etc., or where a bill is overdrawn, or the firm is legally bankrupt [1]. Recent outbreaks of corporate financial crises worldwide has intensified the need to reform existing financial structure. It is generally believed that financial failure symptoms can be observed prior to financial difficulty or crises. Thus, accurate business failure prediction models are of critical importance in terms of the decision making of managers, investors, shareholders and other interested parties, as the models provide them with timely warnings of company's real situation.

Many different relevant techniques have been investigated in the course of various comparative studies which are related in several review articles [3,19,31,32,54,70]. Recently, Kumar and Ravi [35] and Verikas et al. [66] investigated a complete review of methods used for the prediction of business failure and introduced the new trends in this area.

The early studies of business failure prediction used statistical techniques such as univariate statistical methods, multiple discriminant analysis (MDA), linear probability models, and logit and probit analysis in order to gather information [2,4]. These conventional statistical methods, however, required to fulfill some restrictive assumptions such as the linearity, normality and independence amongst the predictor or input variables. Considering that the violation of these assumptions for independent variables frequently occurs within financial data [17], the traditional methods contain limitations in terms of effectiveness and validity.

Artificial intelligence approaches such as neural networks (NN) are less vulnerable to these assumptions, and can be used as alternative methods for the solution of prediction problems. NN have shown to have better predictive capability than MDA and logistic regression in business failure prediction [15,47,52,55,59,67,71]. Support vector machines (SVM), developed by Vapnik [64], have gained popularity due to many attractive features and excellent general performance on a wide range of problems. Moreover, SVM embody the structural risk minimization principle (SRM), which has been shown to be superior to the traditional empirical risk minimization principle (ERM) employed by conventional neural networks. Min and Lee [42] have demonstrated that SVM outperform NN, MDA and logistic regression in business failure prediction.

In order to achieve a better classification performance in SVM for business failure prediction, the data inputs for the classifier need special treatment during preparation to guarantee a good performance in the classifier. In this paper, we propose a hybrid

* Corresponding author. Address: Department of Business Administration, National Taipei College of Business, No. 321, Sec. 1, Jinan Rd., Zhongzheng District, Taipei City 100, Taiwan. Tel.: +886 2 2322 6161; fax: +886 2 2322 6404.
E-mail address: ychinc@webmail.ntcb.edu.tw (C.-C. Yeh).

0950-7051/$ - see front matter Crown Copyright © 2010 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.knosys.2010.07.009

model of manifold learning approach which combines both isometric feature mapping (ISOMAP) algorithm and SVM. By making use of the ISOMAP algorithm to perform dimension reduction, is then utilized as a preprocessor in order to improve business failure prediction capability by SVM. The effectiveness of our proposed hybrid approach was verified by experiments that compared PCA combined with SVM.

The remainder of this paper is structured as follows: Section 2, introduces the classification techniques used in previous research such as the manifold learning approach, ISOMAP algorithm and SVM. In Section 3 we outline the hybrid SVM strategies and experiment framework used in this research. Section 4 presents the experimental results of the proposed method. Finally, the conclusion and future research suggestions are contained in Section 5.

2. Literature review

2.1. Isometric feature mapping

Manifold learning for dimensionality reduction is designed to discover significant features and to compute an accurate low-dimensional embedding of high dimensional inputs. Principal component analysis (PCA) is a linear method for dimensionality reduction that projects the data into the subspace with minimum reconstruction error. Though widely used due to its simplicity, PCA is limited by its underlying assumption that the data lies in a linear subspace. However, several algorithms for nonlinear dimensionality reduction (NLDR) have been proposed to overcome the limitation of linear methods.

Most NLDR techniques presume that the data lies on or in the vicinity of a low-dimensional manifold and therefore attempt to map the high dimensional data into a single low-dimensional, global coordinate system. As a relatively new algorithm, the NLDR technique is a new and promising nonparametric dimension reduction approach and can therefore find an intrinsic low-dimensional structure embedded in a high dimensional observation space. Some of the most frequently used manifold learning techniques include isometric feature mapping (ISOMAP) [62], locally linear embedding (LLE) [51], and Laplacian eigenmaps (LE) [6] amongst others. Similar to PCA, these algorithms are easy to implement, but they compute nonlinear embeddings of high dimensional data. Van der Maaten et al. [40] provide a detailed review of the above methods. The utility of manifold learning has been illustrated in different applications, such as face pose detection [25,38], face recognition [69,72], analysis of facial expressions [14,22], human motion data interpretation [28], gait analysis [22,23], visualization of fiber traces [10] and wood texture analysis [44]. Despite very few manifold learning applications have been conducted in accounting and finance fields, it is worth of mentioning the work in [49,50].

The ISOMAP algorithm is adopted in this study to provide a low-dimensional description of high dimensional features. ISOMAP tries to solve the problem of flattening a curved manifold by building on classical multidimensional scaling but aims to recover the intrinsic geometry of the data by preserving geodesic distance within the manifold. That is, Isomap is classical scaling in which the distances have been replaced by estimates of intrinsic geodesic distance. In practice, the basic procedure for implementing manifold to the ISOMAP algorithm can be summarized as follows: for neighboring points the geodesic distances are approximated by Euclidean distances; for distant points the geodesic distances are approximated by the length of the shortest path in a graph with edges connecting nearby points. Finally, ISOMAP applies classical scaling to the matrix of graph distances to obtain a representation of the data in low dimensions [62]:

(1) Constructing neighbor relations on the manifold M based on the pairwise Euclidean distance $d_X(i,j)$ in the input space X. This can be realized either by a fixed neighbor size k or by a fixed distance range. The identified neighbor relations are further represented by a weighted graph G over the data points, with weight $d_X(i,j)$ assigned to the corresponding edges.
(2) Computing the pairwise geodesic distance $d_M(i,j)$ on the manifold M. The geodesic distance is defined as the distance of the shortest path $d_G(i,j)$ in the graph G. For neighboring points, input distance is a good approximation to geodesic distance; for far away points, geodesic distance adding up a sequence of “short hops” between neighboring points which is actually the shortest path in a graph with edges connecting neighboring data points.
(3) Using MDS to the matrix of geodesic distance D_G , where $D_G = \{d_G(i,j)\}$, to construct an embedding of the data in a d-dimensional Euclidean space Y that can best preserve the estimated intrinsic geometry of the manifold.

The key idea here is to use ISOMAP as a means of representing the actual intrinsic dimensionality of the analyzed data. Moreover, ISOMAP manifolds have been reported to retain more global relationships than their LLE counterpart [18] and ISOMAP perform better than the other methodologies [48]. The objective of this research is to investigate if ISOMAP can be implemented to enhance the quality of financial distress prediction.

2.2. Support vector machines

A support vector machine (SVM) is a theory based on statistical learning theory. It incorporates the theory of VC dimension (for Vapnik–Chervonenkis dimension) and the principle of structural risk minimization (SRM). The whole theory can be simply described as follows: searching an optimal hyperplane satisfies the request of classification, then using a certain algorithm to maximize the margin of the separation besides the optimal hyperplane while ensuring the accuracy of correct classification. According to the theory, we can effectively classify the separable data into classes.

Let us define labeled training examples $[x_i, y_i]$, consisting of an input vector $x_i \in R^n$, and a class value $y_i \in -1, 1$, $i = 1, \ldots, I$. For the linearly separable case, the decision rules defined by an optimal hyperplane separating the binary decision classes are given in the following equation in terms of the support vectors

$$Y = sign\left\{\sum_{i=1}^{N} \alpha_i y_i (x \cdot x_i) + b\right\}, \tag{1}$$

where Y is the outcome, y_i is the class value of the training example x_i, and represents the inner product. The vector corresponds to an input and the vectors x_i, $i = 1, \ldots, N$, are the support vectors. In Eq. (1), b and α_i are parameters that determine the hyperplane.

As to the nonlinear separable data, it can be mapped into a high dimensional feature space with a nonlinear mapping in which we can search the optimal hyperplane. Then the problem is converted into searching the nonnegative Lagrange multipliers α_i by solving the following optimization problem [24,57,73],

$$\text{Maximize } Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{2}$$

$$\text{Subject to } \sum_{i=1}^{n} a_i y_i = 0$$
$$0 \le \alpha_i \leqslant C, \quad i = 1, 2, \ldots, n$$

where C is a penalty parameter on the training error, which is now the upper bound on α_i is determined by the user. Hence, the final classification function is given as follows:

$$Y = sign\left\{\sum_{i=1}^{N} \alpha_i y_i K(x, x_i) + b\right\}. \quad (3)$$

The function K is defined as the kernel function for generating the inner products to construct machines with different types of nonlinear decision surfaces in the input space. The common used kernel function is RBF kernel function

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\gamma^2}\right) \quad (4)$$

The SVM classification exercise is implemented in solving a linearly constrained quadratic programming (QP) for finding the support vectors and determining the parameters b and α. For the separable case, there is a lower bound 0 on the coefficient α_i in Eq. (1). For the non-separable case, SVM can be generalized by placing an upper bound C on the coefficients α_i in addition to the lower bound [68].

In brief, the learning process to construct decision functions of SVM is completely represented by the structure of two layers, which seems to be similar with back-propagation neural network (BPNN). However, learning algorithm is different in that SVM is trained with optimization theory that minimizes misclassification based on statistical learning theory. The first layer selects the basis $K(x, x_i)$, $i = 1, \ldots, N$ and the number of support vectors from given set of bases defined be the kernel. The second layer constructs the optimal hyperplane in the corresponding feature space [56,68].

Compared with most other artificial intelligence approaches, SVM leads to improve performance in pattern recognition, regression estimation, financial time-series forecasting [20,34,39,43,61], marketing [7], estimating manufacturing yields [58], text categorization [29], face detection using image [46], hand written digit recognition [12,16] and medical diagnosis [60].

2.3. Discussion

Regarding the above review, some issues as the limitations of literature, are listed below.

1. Most studies only use linear dimensionality reduction techniques for feature selection. Only Ribeiro et al. [49,50] use ISOMAP algorithm. The "ISOMAP + SVM" seems the best model. They did not show whether nonlinear dimensionality reduction techniques outperform the linear dimensionality reduction techniques in terms of predication accuracy.
2. Much related work uses large numbers of training examples and small numbers of testing examples for training and testing the model. However, when the dataset size is small, it may be difficult to make a reliable conclusion based only on a fixed training and testing proportion.
3. Most studies only examine average prediction performance of their models without considering the Type I and Type II errors.

As a result, examining the performance of linear and nonlinear dimensionality reduction methods, especially for ISOMAP and PCA that including average accuracy and Type I and Type II errors by using a small size of datasets is the aim of this paper.

3. Strategies for proposed the hybrid SVM technique

For the proposed business failure prediction method, we employ ISOMAP as a dimensionality reduction techniques. Initially, ISOMAP is applied to select the input features because of its reliability in extracting the significant features. The software ISOMAP was used in Matlab with the algorithm provided by Tenenbaum et al. [62]. We adopted the nearest neighbor method with k from 3 to 50, dependent on which value of k generated the minimal residual variance. Other data processing and visualization steps with the ISOMAP models were carried out with custom written Matlab functions.

To launch experiments with our proposed model, we first survey the solutions that have been proposed in published literature related to business failure prediction and analyze distressed firms in Taiwan to identify significant features from ISOMAP as the inputs of SVM models. Model selection and parameter search play a crucial role in the performance of SVM. However, there is no general guidance for selection of SVM kernel function and parameters so far. One of the advantages of the linear kernel SVM is that there are no parameters to tune except for constant C. But the upper bound C on coefficient α_i affects the prediction performance for the cases where the training data is not separable by a linear SVM [21]. For the nonlinear SVM, there is an additional parameter, the kernel parameter, to tune. The most frequently used such functions are the polynomial kernel, sigmoid kernel and radial basis kernel function (RBF).

The RBF kernel nonlinearly maps the samples into a higher dimensional space unlike the linear kernel, so it can handle the case when the relation between class labels and attributes is nonlinear. The sigmoid kernel behaves like the RBF for certain parameters; however, it is not valid under some parameters [65]. Furthermore, The polynomial function takes a longer time in the training stage of SVM, and it is reported to provide worse results than the RBF function in the previous studies [27,34,61]. Consequently, the RBF is an effective choice for the kernel function. Therefore, this study uses the RBF kernel function in the SVM to discover the optimal solution.

Two major parameters of the RBF kernel, C and γ, have to be set appropriately. To find optimal parameters, the grid algorithm is used for each pair of parameters; a 5-fold cross-validation is conducted on the training set. We implemented the SVM by using LIBSVM – a library for support vector machines [13].

Finally, for verifying the applicability of methodology, we also present combined PCA and SVM as a performance comparison benchmark. Fig. 1 illustrates the overall procedure of this business failure model. Each of the steps is summarized as follows.

Step 1. Use ISOMAP (PCA) algorithm to reduce dimensionality.
Step 2. Use the nearest neighbor method with k that generates minimal residual variance.
Step 3. Choose a kernel function (RBF).
Step 4. Consider a grid space of (C, γ) with $\log_2 C \in \{-5, -3, -1, \ldots, 13\}$ and $\log_2 \gamma \in \{-13, -11, -9, \ldots, 5\}$, for each pair of parameters, conduct 5-fold cross-validation on the training set, in the neighborhood of the parameters (C, γ) that leads to the lowest CV error classification rate, choose a fine grid, and repeat this step.
Step 5. Choose parameters (C, γ) that leads to the lowest CV error classification rate to create a model as the classifier.

4. Empirical analysis

4.1. Data set and variables

A publicly listed firm encounters a business crisis and turns into a distressed company when it declares full-value delivery, stock transaction suspension, re-construction, bankruptcy or withdrawal from the stock market. Based on the above criteria, 80 failed firms and 80 healthy firms are identified in Taiwan from year 2000 to

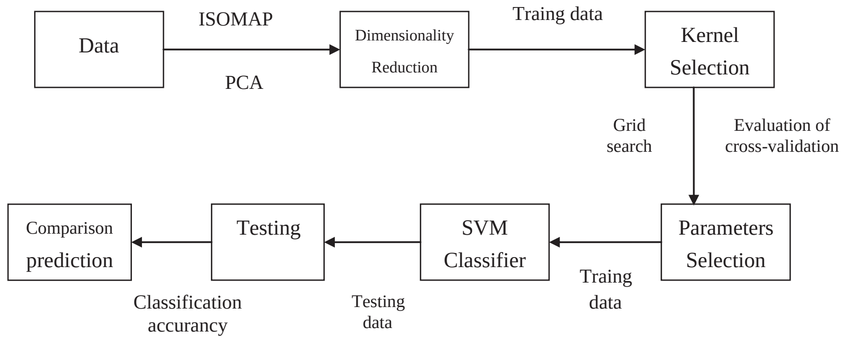

Fig. 1. The procedure of hybrid SVM technique.

2008 according to the Taiwan Economic Journal (TEJ) databank. The data were gathered based on two criteria below: (1) The sample firms need to have at least four quarters of complete public information before the business crisis happens; (2) There should be sufficient comparable companies with similar size and in the same industry to serve as control samples.

Prior researches on bankruptcy prediction have pinpointed a considerable number of significant predictors of business failure as shown in Table 1. We select 16 financial ratios, which have been proved through prior research to be efficient in financial failure prediction, as the potential predictor variables.

After the data including variables are collected, to guarantee that the present results are valid and can be generalized for making predictions regarding new data, the data set is further randomly partitioned into training and independent testing sets via an n-fold cross-validation. Each of the n-subsets acts as an independent holdout test set for the model trained with the rest of $n-1$ subsets. The overall classification accuracy of the built model is then just the simple average of the n individual accuracy measures. The advantages of cross-validation are that the impact of data dependency is minimized and the reliability of the results can be improved [53].

Besides, as cross-validation is the preferred procedure in testing the out-of-sample classification capability when the dataset size is small [9,30] and the size of failed firms is only 80, the 5-fold cross-validation will be adopted in this study. Therefore there are 32 firms in each subset of the dataset, in which 80% of the dataset performs model training, and the other for model testing. Besides, every subset will be trained and tested 5 times, and the average prediction performance can be obtained consequently.

Table 1
Definition of variables.

Variables	Description	References
x_1	Current assets/current liabilities	[5,41]
x_2	Cash flow/total debt	[5,8,17,41]
x_3	Cash flow/total assets	[17,45]
x_4	Cash flow/net sales	[17,45]
x_5	Total liabilities/total assets	[5,17,20,41,45]
x_6	Working capital/total assets	[2,5]
x_7	Market value equity/book value of total debt	[2,37,41]
x_8	Current assets/total assets	[17]
x_9	Quick assets/total assets	[17]
x_{10}	Net sales/total assets	[2,37]
x_{11}	Current liabilities/net sales	[17]
x_{12}	Quick assets/net sales	[17]
x_{13}	Working capital/net sales	[5,41,45]
x_{14}	Net profit/total assets	[5,17,45]
x_{15}	Retained earnings/total assets	[2,20]
x_{16}	Earnings before interest and taxes/total assets	[2,37]

4.2. Dimensionality reduction by ISOMAP and PCA

Because the elements of input variables vary greatly, the dataset is standardized beforehand. Then the Isomap algorithm is executed to reduce the dimension of the dataset with the neighborhood. To evaluate the fits of ISOMAP, and to determine the intrinsic low-dimensional embedding, we find the residual variance defined by $1-R^2(\widehat{D}_M, D_Y)$ [62]. D_Y is the matrix of Euclidean distances in the low-dimensional embedding, and $\widehat{D}_M$ is the best estimate of the intrinsic manifold distances, which is the graph distance matrix D_G. R is the standard linear correlation coefficient, taken over all entries of $\widehat{D}_M$ and D_Y.

The key parameter is the number k of neighbors used for the projection. Fig. 2 illustrates the residual variances of ISOMAP results with the neighborhood, the parameter $k = 20$. From this plot, we can deduce that the first five dimensions of manifold from ISOMAP would be a good description the hyper spectral data. We found that at this value, the resulting residual variance at each lower dimension remains approximately the same for all $d > 5$. As a whole, we can draw a conclusion that the intrinsic dimensionality is 5.

Fig. 3 shows the distribution of points in the first two dimensions of the manifold. There are two classes in the data (healthy or unhealthy), and we can see that most of the classes of the training data are separated on the manifold. This could lead to high test accuracy when SVM are implemented to classify the processed data.

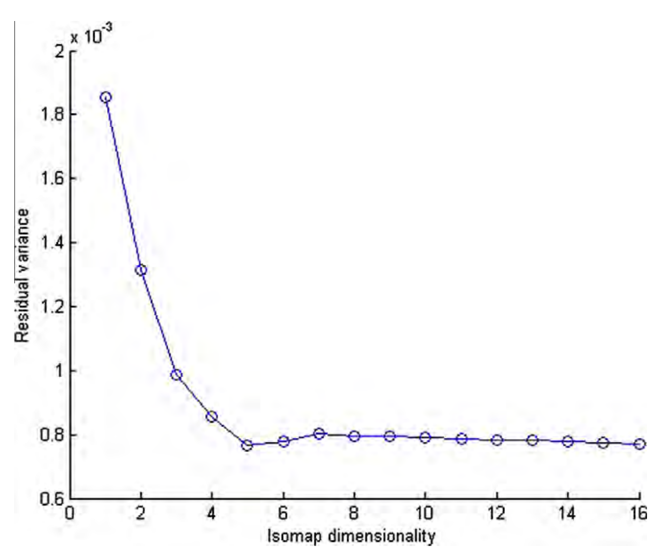

Fig. 2. Residual variance vs. ISOMAP dimensionality.

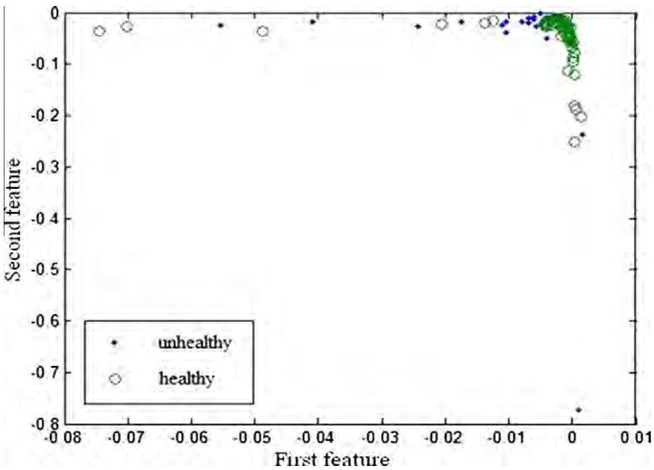

Fig. 3. Two-dimensional embedding with ISOMAP.

Table 3
Paired *t* test of results.

	ISOMAP + SVM	PCA + SVM	SVM
ISOMAP + SVM		0.199 (0.423)	1.421 (0.096)*
PCA + SVM			2.178 (0.030)**

* <0.1.
** <0.05.

Table 4
Type I and Type II errors of the constructed model.

Model	Performance assessment (%)	
	Type I error	Type II error
ISOMAP + SVM	20.82	23.68
PCA + SVM	11.36	35.21
SVM	23.77	30.26

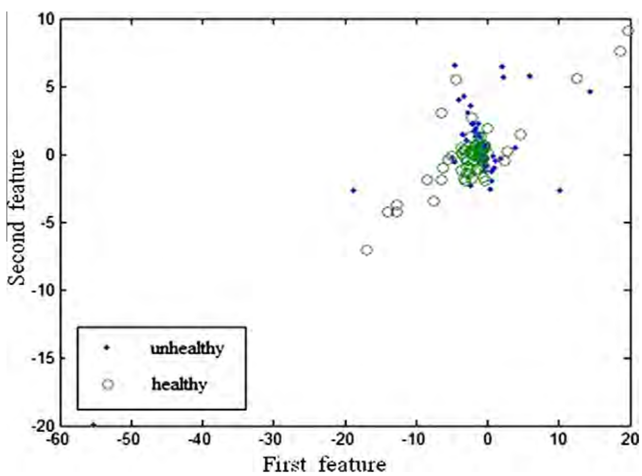

Fig. 4. Two-dimensional embedding with PCA.

In order to evaluate the classification capabilities of the proposed business failure prediction models, we also apply PCA to the same data. The eigenvalues seem to increase linearly in absolute value, and a subset of at least three variables is necessary in order to explain 80% of the data variability. Fig. 4 presents the two-dimensional embedding obtained. As indicated in Fig. 4, a lot of the data belonging to different classes is mixed and nearly superposed. Obviously, this superposed training data makes the business failure prediction decision too complex while using SVM-based classification – there is a small overlap of ‘unhealthy’ into the ‘healthy’ classifications. As for ISOMAP, a structure can be recognized. However, PCA is not as distinct; this could lead to potential misclassification problems later.

4.3. Comparison with other models

After the dimensionality of the dataset is reduced, SVM classifiers are implemented. In this experiment, SVM achieves accuracy ranging from 65.23% to 75.00% in predicting business failure with an average prediction accuracy of 72.50%. Along with two different dimensionality reduction methods, their prediction accuracies are shown in Table 2. In particular, the “ISOMAP + SVM” is the ‘best’ classifier. However, on average there is no exact ‘winner’ as the best model for business failure prediction. Table 3 shows the *t* test result based on the average performance of these three models.

Although there is no significant difference between “ISOMAP + SVM” and “PCA + SVM” in terms of prediction accuracy, on average the “ISOMAP + SVM” seems the best model. If we consider the average prediction result above, the “ISOMAP + SVM”, “PCA + SVM” and SVM produce a rate of 0.7750, 0.7688 and 0.7250 classification accuracy respectively. Above all, the “ISOMAP + SVM” could be used as the baseline a hybrid model of manifold learning algorithm with SVM classifier for future related research.

During these experiments, the incidences of Type I and Type II errors were also noted. A Type I error is misclassification of a healthy company as an unhealthy one; a Type II error is exactly the opposite. Clearly, a good business failure prediction model should have lower Type II errors. Table 4 summarizes the occurrences of Type I and Type II errors in the five models being discussed. The results show our hybrid techniques have comparatively fewer Type II errors, further proof of its all round superiority.

5. Conclusions and future research

The prediction of business failure is an important and challenging issue that has served as the impetus for many academic studies over the past three decades. By minimizing the sum of the empirical risk and the complexity of the hypothesis space, SVM gives good general performance on many business failure prediction problems. In order to ensure an accurate classification process in SVM, the preparation of data inputs for the classifier needs special treatment in order to guarantee a good performance in the classifier. Moreover, the manifold learning approach, such as ISOMAP,

Table 2
The performance of SVM model with two different manifold learning methods using 5-fold cross-validation.

SVM + manifold learning methods	Accuracy of 5-fold cross-validation					
	Set 1	Set 2	Set 3	Set 4	Set 5	Accuracy (%)
ISOMAP + SVM	75.00	78.13	87.50	68.75	78.13	77.50
PCA + SVM	75.00	75.00	78.13	78.13	78.13	76.88
SVM	75.00	75.00	75.00	71.88	65.23	72.50

has become very popular amongst scientists worldwide and is now one of the most developed techniques in dimensionality reduction analyses. However very few manifold learning applications have been conducted in accounting and finance fields [49,50].

This paper focuses on examining the performance of linear and nonlinear dimensionality reduction methods, especially for ISOMAP and PCA that including average accuracy and Type I and Type II errors by using a small size of datasets is the aim of this paper. We propose a hybrid model which is composed of manifold learning approach, "ISOMAP + SVM". In order to verify the applicability of this methodology, we also present "PCA + SVM" as the benchmark to a dataset on bankruptcies in Taiwan. The results show that there is no significant difference between "ISOMAP + SVM" and "PCA + SVM" in terms of prediction accuracy. However, if we consider the average prediction result, the "ISOMAP + SVM" seems the best model. On the other hand, by examining the Type I and Type II errors of these models, our hybrid approach model produces fewer Type II errors. The afore-mentioned findings justify the presumption that our hybrid approach model is a better alternative which conducting business failure prediction tasks.

However, there are some limitations in this article that call for further research. First, we can observe that the selection of neighbor factor k is related to the classification accuracy of the business failure prediction. We will study the feasibility of using other approaches to solve this problem. Finally, the results show that there is no significant difference between "ISOMAP + SVM" and "PCA + SVM" in terms of prediction accuracy. Above all, the "ISOMAP + SVM" could be used as the baseline a hybrid model of manifold learning algorithm with SVM classifier for future related research. Meanwhile, future researches may use other dimensionality reduction approaches, such as LLE, Semi-Supervised ISOMAP and LE amongst others with business failure prediction.

References

[1] B.S. Ahn, S.S. Cho, C.Y. Kim, The integrated methodology of rough set theory and artificial neural network for business failure prediction, Expert Systems with Applications 18 (2000) 65–74.

[2] E.I. Altman, Financial ratios discriminant analysis and the prediction of corporate bankruptcy, Journal of Finance 23 (4) (1968) 589–609.

[3] E.I. Altman, The success of business failure prediction models: an international survey, Journal of Banking and Finance 8 (2) (1984) 171–198.

[4] E.I. Altman, R.G. Haldeman, P. Narayanan, Zeta analysis, Journal of Banking and Finance June (1977) 29–51.

[5] W.H. Beaver, Financial ratios as predictors of failure empirical research in accounting: selected studies, Supplement to the Journal of Accounting Research 4 (1966) 179–199.

[6] M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, Advances in Neural Information Processing Systems 14 (2002) 585–591.

[7] S. Ben-David, M. Lindenbaum, Learning distributions by their density levels: a paradigm for learning without a teacher, Journal of Computer and System Sciences 55 (1997) 171–182.

[8] M. Blum, Failure company discriminant analysis, Journal of Accounting Research (1974) 1–25.

[9] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, Wadsworth, Pacific Grove, CA, 1984.

[10] A. Brun, H.J. Park, H. Knutsson, C.F. Westin, Coloring of DT-MRI fiber traces using Laplacian eigenmaps, in: Proc. Ninth International Conference on Computer Aided Systems Theory, vol. 2809, 2003.

[11] S.M. Bryant, A case-based reasoning approach to bankruptcy prediction modeling, Intelligent System Accounting, Financial and Management 6 (1997) 195–214.

[12] C.J.C. Burges, B. Scholkopf, Improving the accuracy and speed of support vector machines, in: M. Mozer, M. Jordan, T. Petsche (Eds.), Advances in Neural Information Processing Systems, MIT Press, Cambridge, MA, 1997, pp. 475–481.

[13] C.C. Chang, C.J. Lin, LIBSVM: A Library for Support Vector Machines, 2001, <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

[14] Y. Chang, C. Hu, M. Matthew Turk, Probabilistic expression analysis on manifolds, in: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, 2004, pp. 520–527.

[15] K.G. Coleman, T.J. Graettinger, W.F. Lawrence, Neural networks for bankruptcy prediction: the power to solve financial problems, AI Review (1991) 48–50.

[16] C. Cortes, V.N. Vapnik, Support vector networks, Machine Learning 20 (1995) 273–297.

[17] E.B. Deakin, A discriminant analysis of predictors of business failure, Journal of Accounting Research 10 (1972) 167–179.

[18] V. De Silva, J.B. Tenenbaum, Global Versus Local Methods in Nonlinear Dimensionality Reduction, MIT Press, 2003. pp. 705–712.

[19] A.I. Dimitras, S.H. Zanakis, C. Zopounidis, A survey of business failures with an emphasis on prediction methods and industrial applications, European Journal of Operational Research 90 (1996) 487–513.

[20] Y.S. Ding, X.P. Song, Y.M. Zen, Forecasting financial condition of Chinese listed companies based on support vector machine, Expert Systems with Applications 34 (2008) 3081–3089.

[21] H. Drucker, D. Wu, V.N. Vapnik, Support vector machines for spam categorization, IEEE Transactions Neural Networks 10 (5) (1999) 1048–1054.

[22] A. Elgammal, C.S. Lee, Separating style and content on a nonlinear manifold, in: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004, pp. 478–489.

[23] A. Elgammal, C.S. Lee, Inferring 3D body pose from silhouettes using activity manifold learning, in: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, 2004, pp. 681–688.

[24] C. Gold, P. Sollish, Model selection for support vector machine classification, Neurcomputing 55 (2005) 221–249.

[25] A. Hadid, O. Kouropteva, M. Pietikainen, Unsupervised learning using locally linear embedding: experiments in face pose analysis, in: Proc. 16th International Conference on Pattern Recognition, vol. I, 2002, pp. 111–114.

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

[28] O. Jenkins, M. Mataric, A spatio-temporal extension to Isomap nonlinear dimension reduction, in: Proc. 21st International Conference on Machine Learning, 2004.

[29] T. Joachims, Learning to Classify Text Using Support Vector Machines, Kluwer Academic Publishers, London, 2002.

[30] R.A. Johnson, D.W. Wichern, Applied Multivariate Statistical Analysis, fifth ed., Prentice-Hall, Upper Saddle River, NJ, 2002.

[31] F.L. Jones, Current techniques in bankruptcy prediction, Journal of Accounting Literature 6 (1987) 131–164.

[32] K. Keasey, R. Watson, Financial distress prediction models: a review of their usefulness, British Journal of Management 2 (1991) 89–102.

[34] K.J. Kim, Financial time series forecasting using support vector machines, Neurocomputing 55 (1/2) (2003) 307–319.

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

[36] H. Li, J. Sun, Ranking-order case-based reasoning for financial distress prediction, Knowledge-based Systems 21 (2008) 868–878.

[37] H. Li, J. Sun, Predicting financial distress using multiple case-based reasoning combine with support vector machine, Expert Systems with Applications (2009).

[38] S.Z. Li, R.Z. Xiao, Y. Li, H.J. Zhang, Nonlinear mapping from multi-view face patterns to a gaussian distribution in a low dimensional space, in: Proc. IEEE ICCV Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-time Systems, 2001.

[39] C.J. Lu, T.S. Lee, C.C. Chiu, Financial time series forecasting using independent component analysis and support vector regression, Decision Support Systems (2009).

[40] L.J.P. Van der Maaten, E.O. Postma, H.J. van den Herik, Dimensionality reduction: a comparative review, 2007, <http://www.cs.unimaas.nl/l.vandermaaten/dr/DR_draft.pdf>.

[41] D. Martens, L. Bruynseels, B. Baesens, M. Willekens, J. Vanthienen, Predicting going concern opinion with data mining, Decision Support Systems 45 (2008) 765–777.

[42] J.H. Min, Y.C. Lee, Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters, Expert Systems with Applications 28 (2005) 603–614.

[43] S. Mukherjee, E. Osuna, F. Girosi, Nonlinear prediction of chaotic time series using support vector, in: Proc. IEEE Workshop on Neural Networks for Signal Processing, Amelia Island, FL, 1997, pp. 511–520.

[44] M. Niskanen, O. Silvn, Comparison of dimensionality reduction methods for wood surface inspection, in: Proc. Sixth International Conference on Quality Control by Artificial Vision, 2003, pp. 178–188.

[45] J.A. Ohlson, Financial ratios and the probabilistic prediction of bankruptcy, Journal of Accounting Research (1980) 109–131.

[46] E. Osuna, R. Freund, F. Girosi, Training support vector machines: an application to face detection, in: Proc. Computer Vision and Pattern Recognition, 1997, pp. 130–136.

[47] E. Rahimian, S. Singh, T. Thammachote, R. Virmani, Bankruptcy prediction by neural networks, in: E. Trippi, E. Turban (Eds.), Neural Networks in Finance and Investing: Using Artificial Intelligence to Improve Real-World Performance, Probus Publishing, Chicago, 1993, pp. 159–176.

[48] F. Ratle, A.L. Terrettaz, M. Kanevski, P. Esseiva, O. Ribaux, in: S. Kollias et al. (Eds.), Learning Manifolds in Forensic Data, Springer-Verlag, Berlin, Heidelberg, 2006, pp. 894–903.

[49] B. Ribeiro, C. Silva, A. Sung, A.Vieira, J. Neves, J. Duarte, Q. Liu, Learning manifolds for bankruptcy analysis, in: M. Koppen et al. (Eds.), Advances in Neuro-information Processing, Lecture Notes in Computer Science vol. 5506:1, 2009, pp. 722–729.

[50] B. Ribeiro, A. Vieira, J. Neves, Supervised isomap with dissimilarity measures in embedding learning, in: Proc. Ibero-American Conference on Pattern Recognition, Progress in Pattern Recognition, Image Analysis and Applications, Lecture Notes in Computer Science, vol. 5197, Springer, Berlin, Heidelberg, 2008, pp. 389–396.
[51] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[52] L.M. Salchengerger, E.M. Cinar, N.A. Lash, Neural networks: a new tool for prediction thrift failures, Decision Sciences 23 (1992) 899–916.
[53] S.L. Salzberg, On comparing classifiers: pitfalls to avoid and a recommended approach, Data Mining and Knowledge Discovery 1 (1997) 317–327.
[54] J. Scott, The probability of bankruptcy: a comparison of empirical predictions and theoretical models, Journal of Banking and Finance 5 (1981) 317–344.
[55] R. Sharda, R.L. Wilson, Neural networks experiments in business-failure forecasting: predictive performance measurement issues, International Journal of Computational Intelligence and Organizations 1 (2) (1996) 107–117.
[56] K.S. Shin, T.S. Lee, H.J. Kim, An application of support vector machines in bankruptcy prediction model, Expert Systems with Applications 28 (2005) 127–135.
[57] D.M. Sinalingam, N. Pandia, Minimal classification method with error correlation codes for multiclass recognition, International Journal of Pattern Recognition and Artificial Intelligence 5 (2005) 663–680.
[58] D. Stoneking, Improving the manufacturability of electronic designs, IEEE Spectrum 36 (6) (1999) 70–76.
[59] K.Y. Tam, M. Kiang, Managerial applications of neural networks: the case of bank failure predictions, Management Science 38 (7) (1992) 926–947.
[60] L. Tarassenko, P. Hayton, N. Cerneaz, M. Brady, Novelty detection for the identification of masses in mammograms, in Proc. Fourth IEE International Conference on Artificial Neural Networks, Cambridge, 1995, pp. 442–447.
[61] F.E.H. Tay, L. Cao, Application of support vector machines in financial time series forecasting, Omega 29 (2001) 309–317.
[62] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (2000) 2319–2323.
[63] C.F. Tsai, Feature selection in bankruptcy prediction, Knowledge-based Systems 22 (2) (2009) 120–127.
[64] V.N. Vapnik, The Nature of Statistical Learning Theory, second ed., Springer-Verlag, New York, 1995.
[65] V.N. Vapnik, Statistical Learning Theory, Springer, New York, 1998.
[66] 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, Methodologies and Applications, September 2009 (online).
[67] R.L. Wilson, R. Sharda, Bankruptcy prediction using neural networks, Decision Support Systems 11 (1994) 545–557.
[68] I.H. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufman, San Francisco, CA, 2000.
[69] M.H. Yang, Face recognition using extended Isomap, in: IEEE International Conference on Image Processing, vol. II, 2002, pp. 117–120.
[70] C.V. Zavgren, The prediction of corporate failure: the state of the art, Journal of Financial Literature 2 (1983) 1–37.
[71] G. Zhang, M.Y. Hu, B.E. Patuwo, D.C. Indro, Artificial neural networks in bankruptcy predictions: general framework and cross-validation analysis, European Journal of Operational Research 116 (1999) 16–32.
[72] J. Zhang, S.Z. Li, J. Wang, Nearest manifold approach for face recognition, in: Proc. Sixth International Conference on Automatic Face and Gesture Recognition, Seoul, Korea, 2004.
[73] Y.S. Zhu, Y.Y. Zhang, The study on some problems of support vector classifier, Computer Engineering and Applications 13 (2003) 38–66.