5. A HYBRID BUSINESS FAILURE PREDICTION MODEL USING LOCALLY LINEAR EMBEDDING AND SUPPORT VECTOR MACHINES

Fengyi LIN [1]
Ching-Chiang YEH [2]
Meng-Yuan LEE[3]

Abstract

The purpose of this paper is to propose a hybrid model which combines locally linear embedding (LLE) algorithm and support vector machines (SVM) to predict the failure of firms based on past financial performance data. By making use of the LLE algorithm to perform dimension reduction for feature extraction, is then utilized as a preprocessor to improve business failure prediction capability by SVM. The effectiveness of the methodology was verified by comparing principal component analysis (PCA) and SVM with our proposed hybrid approach. The results show that our hybrid approach not only has the best classification rate, but also produces the lowest incidence of Type I and Type II errors, and is capable to provide on time signals for better investment and government decisions with timely warnings.

Keyword: business failure, manifold learning, locally linear embedding, support vector machines

JEL Classification: G33,C45

I. 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 (Altman, 1968; Beaver, 1966; Chang and Lin, 2001; Li and Sun, 2009; Pettis *et al*.,

[1] Department of Business Management, National Taipei University of Technology, Taipei, Taiwan.
[2] Department of Business Administration, National Taipei College of Business, Taipei, Taiwan. Corresponding author. E-mail: ychinc@webmail.ntcb.edu.tw
[3] Institute of Commerce Automation and Management, National Taipei University of Technology, Taipei, Taiwan.

1979). 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 (Ahn *et al.*, 2000). Recent outbreaks of corporate financial crises worldwide have intensified the need to reform the 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 a company's real situation.

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 (Coleman *et al.*, 1991; Rahimian *et al.*, 1993; Salchengerger *et al.*, 1992; Sharda and Wilson, 1996; Tam and Kiang, 1992; Wilson and Sharda, 1994; Zhang *et al.*, 1999). Support vector machines (SVM), developed by Vapnik (1995), 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, Lee and Han (2006) have demonstrated that SVM outperform NN, MDA and logistic regression in business failure prediction. However, despite the fact that SVM has outstanding performance, its classification performance and classifier's generalization ability is often influenced by its dimension or number of feature variables.

Dimensionality reduction methods aim to discover this underlying low dimensional structure. These methods can be categorized as linear or nonlinear. Linear methods are limited to discovering the structure of data lying on or near a linear subspace of the high dimensional input space. The most widely used linear dimensionality reduction methods include the principal component analysis (PCA) (Jolliffe, 1986). These methods have been successfully applied to feature extraction in many applications in the past.

However, Beaver (1996) used financial ratios as the input variables in conducting business failure predictions and they are commonly used to develop the models or classifiers. One aspect is the complex and nonlinear relationship between financial ratios of business failure such as cash flow and net profit. Another aspect is if these financial ratios lie on a low dimensional manifold (Seung and Lee, 2000) in the high dimensional input space. Therefore, the linear methods will fail to discover the low dimensional structure. Recently, several algorithms for nonlinear dimensionality reduction (NLDR) that overcome this limitation of PCA, including isometric feature mapping (Isomap) (de Silva and Tenenbaum, 2003), or locally linear embedding (LLE) (Roweis and Saul, 2000) were proposed. Most NLDR techniques presume that the data lies on or in the vicinity of a low dimensional manifold and attempt to map the high dimensional data into a single low dimensional, global coordinate system. Like PCA, these algorithms are simple to implement; they can find an intrinsic low dimensional structure embedded in a high dimensional observation space. The utility of NLDR has been illustrated in different applications, such as face pose detection

(Hadid *et al*., 2002; Li *et al*., 2001), face recognition (Yang, 2002; Zhang *et al*., 2004), analysis of facial expressions (Chang *et al*., 2004; Elgammal and Lee, 2004), human motion data interpretation (Jenkins and Mataric, 2004), gait analysis (Elgammal and Lee, 2004a, 2004b), visualization of fiber traces (Brun *et al*., 2003) and wood texture analysis (Niskanen and Silvn, 2003). Van der Maaten, Postma and Van den Herik (2008) provide a detailed review of the algorithms. However, there has been little work on the application of LLE for business failure prediction.

The purpose of this paper is to propose a hybrid model of NLDR approach which combines both LLE algorithm and SVM. The LLE algorithm which is firstly used to the exploratory analysis and visualization of the distribution of high-dimensional data in a 2-dimensional space shows that some of the data, which belong to different classes, are absolutely mixed. Besides, by making use of the LLE algorithm to perform dimension reduction for feature extraction, it 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 literature review such as the prior studies on bankruptcy prediction, LLE algorithm and SVM. In section 3 we outline the research methodology and experiment framework used in this research. Section 4 presents the experiment results of the proposed method. Finally, the conclusion and future research suggestions are presented in Section 5.

2. Literature Review

2.1 Prior Studies on Business Failure Prediction

In the past, Beaver (1966) used financial ratios as the input variables for linear regression models to classify healthy/bankrupt firms. Altman (1968) used the classical multivariate discriminate analysis technique. In addition, research by Ohlson (1980) used LOGIT and PROBIT models to predict bankruptcies. On the other hand, to develop a more accurate and generally applicable prediction approach, data mining and machine learning techniques including decision trees, NN, fuzzy logic, genetic algorithm (GA), SVM, etc., have been successfully applied in corporate financial distress forecasting. Kumar and Ravi (2007) and Verikas *et al*. (2009) investigated a complete review of methods used for the prediction of business failure and introduced the new trends in this area. Related work shows that data mining models (e.g. neural networks) outperform statistical approaches (e.g. logistic regression, linear discriminate analysis, and multiple discriminate analysis) (Huang *et al*., 2004; Kumar and Ravi, 2007; Min and Lee, 2005; Shin *et al*., 2005; Zhang *et al*., 1999).

Recently, several studies investigated the efficacy of applying SVM to business failure prediction. Fan and Palaniswami (2000) showed that SVM outperformed traditional classifiers for bankruptcy prediction such as DA, multi-layer perceptron, and learning vector quantization. Van Gestel *et al*. (2003) also reported on the experiment with least squares SVM, a modified version of SVM, and showed significantly better results in bankruptcy prediction when contrasted with the classical techniques. Shin, Lee and

Kim (2005) used SVM for predicting the corporate bankruptcy and compared the results with NN. They showed that the accuracy and generalization performance of SVM is better than that of NN as the training set size gets smaller. Min and Lee (2005) showed that SVM outperformed LOGIT, DA, and NN for business failure bankruptcy prediction.

Min, Lee, and Han (2006) proposed methods for improving SVM performance in two aspects: parameter optimization and feature selection. Min, Lee, and Han (2006) and Wu, Tzeng, Goo, and Fang (2007) used GA to optimize both a feature subset and parameters of SVM simultaneously for bankruptcy prediction. On the other hand, Yeh, Chi, and Hsu (2010) use rough set theory (RST) to feature selection to improve business failure prediction by SVM. They show that RST-SVM model provides better prediction results. Ribeiro, Silva, Sung, Vieira, Neves, Duarte, and Liu (2009) and Ribeiro, Vieira, and Neves (2008) applied Isomap algorithm to a large database of distressed and healthy companies for dimension reduction. Their results showed that Isomap-SVM approach provides better than SVM. In general, the above studies that used SVM to predict financial failure show that SVM is better than NN and statistical methods in predicting the bankruptcy.

2.2 Locally Linear Embedding

LLE (Roweis and Saul, 2000) is an unsupervised learning algorithm that computes low dimensional embeddings of high dimensional data. The principle of LLE is to compute a low dimensional embedding with the property that nearby points in the high dimensional space remain nearby and similarly co-located with respect to one another in the low dimensional space. In other words, the embedding is optimized to preserve local neighborhoods.

The LLE algorithm can be summarized in three steps:

1. For each data point X_i, compute its k nearest neighbors (based on Euclidean distance or some other appropriate definition of 'nearness').

2. Compute weights W_{ij} that best reconstruct each data point X_i from its neighbors, minimizing the reconstruction error E:

$$E(W) = \sum_i \left| X_i - \sum_j W_{ij} X_j \right|^2 \quad (1)$$

3. Compute the low dimensional embeddings, Y_i, best reconstructed by the weights W_{ij}, minimizing the cost function Ω:

$$\Omega(W) = \sum_i \left| Y_i - \sum_j W_{ij} Y_j \right|^2 \quad (2)$$

In step 2, the reconstruction error is minimized subject to two constraints: first, that each input is reconstructed only from its nearest neighbors, or $W_{ij} = 0$ if X_i is not a neighbor of X_j; second, that the reconstruction weights for each data point sum to one, or $\sum_i W_{ij} = 1 \ \forall \ i$. The optimum weights for each input can be computed efficiently by solving a constrained least squares problem.

The cost function in step 3 is also based on locally linear reconstruction errors, but here the weights W_{ij} are kept fixed while optimizing the outputs Y_i. The embedding cost function in Equation (2) is a quadratic function in Y_i. The minimization is performed subject to constraints that the outputs are centered and have unit covariance. The cost function has a unique global minimum solution for the outputs Y_i. This is the result returned by LLE as the low dimensional embedding of the high dimensional data points X_i.

The LLE algorithm has two free parameters: the number of neighbors, *k,* in the first step, and the target dimensionality, d, in the third step. The number of neighbors should always be higher than the target dimensionality. Moreover, if we represent the inputs by the vertices of an undirected graph with edges connecting neighboring inputs, then LLE should only be applied to inputs that give rise to connected graphs. Naturally, the target dimensionality depends on the intended use of the embedding, with $d = 2$ or $d = 3$ typically chosen for visualization. For other tasks, however, estimating the underlying dimensionality of sampled manifolds is an important issue that LLE does not itself address. Various methods have been proposed to estimate this dimensionality; these can be used in conjunction with the first step of LLE to select the target dimensionality, d.

2.3 Support Vector Machine

A support vector machine (SVM) is a theory based on the 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 that 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.

Consider the problem of separating a set of training data $x_i \in R^n (i-1,2,...,n)$ with the desired output $y_i \in \{+1,-1\}$ corresponding to the two classes. In addition, suppose there is a separating hyperplane with the target functions $w \cdot x_i + b = 0$ (*w* represents the weight vector and *b* the bias). To ensure that all training data can be classified, we must maximize the margin of separation $(2/\|w\|)$. Then, in the case of linear separation, the linear SVM for optimal separating hyperplane has the following optimization problem, Minimize $$\phi(w)=\frac{1}{2}w^T w \tag{3}$$

Subject to $$y_i(x_i\cdot w+b)\geq 1,\quad i=1,2,...,n \tag{4}$$

As to the non-linear 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 $\{\alpha_i\}_{i=1}^{n}$ by solving the following optimization problem (Gol and Sollish, 2005; Sinalingam and Pandia, 2005; Zhu and Zhang, 2003),

Maximize $$Q(\alpha)=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j K(x_i,x_j) \tag{5}$$

Subject to $$\sum_{i=1}^{n}\alpha_i y_i=0\quad 0\leq\alpha_i\leq C,\quad i=1,2,...,n \tag{6}$$

Hence, one constructs the final classification function is

$$f(x)=\mathrm{sgn}\left\{\sum_{i=1}^{n}\alpha_i^{\phi}y_i K(x_i,x_j)+b^{\phi}\right\} \tag{7}$$

The common used kernel function is RBF kernel function.

$$K(x,x^{'})=\exp(-\frac{\left\|x-x^{'}\right\|^2}{2\sigma^2}) \tag{8}$$

2.4 Limitation of Current Prediction Techniques

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*. (2008, 2009) and Lin *et al*. (2011) used the Isomap algorithm. However, they did not use the LLE algorithm and 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, to examine the performance of linear and nonlinear dimensionality reduction methods, especially for LLE and PCA that include average accuracy and Type I and Type II errors by using a small size of datasets is the aim of this paper.

3. Experimental Design

We use LLE as a tool for visualization of financial data and dimensionality reduction. The number of nearest neighbours, k, used in LLE was set equal to 3 and 16, respectively. These values were chosen empirically by varying k and examining the classification performance. The performance of both methods was found to be sensitive to the choice of k. To launch experiments with our proposed model, we first surveyed the literature related to business failure prediction and analyzed distressed firms in Taiwan to identify significant features from LLE as the inputs of SVM models. 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 (Salzberg, 1997). Besides, as cross-validation is the preferred procedure in testing the out-of-sample classification capability when the dataset size is small (Breiman *et al*., 1984; Johnson and Wichern, 2002) and the size of failed firms is only 80, the five-fold cross-validation will be adopted in this study. As a result, each model will be trained and tested by five times.

Moreover, proper parameters setting can improve the SVM classification accuracy. With the RBF kernel, there are two parameters to be determined in the SVM model: C and σ. The grid search approach (Chen and Lin, 2005; Hsu *et al*., 2003) is an alternative to finding the best C and σ when using the RBF kernel functions.

In the grid search approach, pairs of (C,σ) are tried and the one with the best cross-validation accuracy is chosen. After identifying a ‘‘better’’ region on the grid, a finer grid search on that region can be conducted. To get good generalization ability, grid search approach uses a *5*-fold cross- validation process to decide parameters. That is, we split the data set randomly into *5* subsets and then perform *5* test runs where in each run we take one subset as our test set and the remaining *4* subsets as our training set, then run the process as follows (Chen and Lin, 2005; Hsu *et al*., 2003):

1. Consider a grid space of (C,σ) with $\log_2 C \in \{-5,-3,-1,...,13\}$ and $\log_2 \sigma \in \{-13,-11,-9,...,-5\}$.
2. For each hyper parameter pair (C,σ) in the search space, conduct five -fold cross-validation on the training set.
3. Choose parameters (C,σ) that lead to the lowest CV error classification rate to create.
4. Use the best parameter to create a model as the predictor.

Finally, for verifying the applicability of methodology, we also present combined PCA and SVM as a performance comparison benchmark.

3.1 Dataset

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 were identified in Taiwan from 2000 to 2008 according to the Taiwan Economic Journal (TEJ) databank. The data were gathered based on two criteria: (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.

3.2 Variables

Prior researches on business failure prediction have pinpointed a 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.

Table 1

Definition of Variables

Variables	Description	References
x_1	Current assets/Current liabilities	Brand (2003)
x_2	Cash flow/Total debt	Brand (2003); Blum (1974); Deakin (1972)
x_3	Cash flow/Total assets	Deakin (1972); Ohlson (1980)
x_4	Cash flow/Net sales	Deakin (1972); Li *et al*. (2009)
x_5	Total liabilities/Total assets	Brand(2003), Deakin (1972); Ding *et al*. (2008); Ohlson (1980)
x_6	Working capital/Total assets	Altman(1968); Brand (2003)
x_7	Market value equity/Book value of total debt	Altman (1968); Li *et al*. (2009)
x_8	Current assets / Total assets	Deakin (1972)
x_9	Quick assets/Total assets	Deakin (1972)
x_{10}	Net sales/Total assets	Altman (1968); Li *et al*. (2009)
x_{11}	Current liabilities / Net sales	Deakin (1972
x_{12}	Quick assets/Net sales	Deakin (1972)
x_{13}	Working capital/Net sales	Brand (2003); Ohlson (1980)
x_{14}	Net profit/ Total assets	Brand (2003); Deakin (1972); Ohlson (1980)
x_{15}	Retained earnings/Total assets	Altman (1968); Ding *et al*. (2008)
x_{16}	Earnings before interest and taxes/Total assets	Altman (1968); Li *et al*. (2009)

4. Experimental results

4.1 Dimensionality Reduction by LLE and PCA

The quality of the resulting projection is measured by the squared difference between the original point and its projection. The main parameter to tune is the number *k* of neighbors used for the projection. Figure 1 shows the distribution of points in the first two dimensions, with LLE with *k* = 15 nearest neighbors providing the best resulting manifold. There are two classes in the data (healthy or unhealthy), and we can see that most of the classes of the data are separated on the manifold. This could lead to high accuracy when SVM are implemented to classify the processed data.

Figure 1

Two-Dimensional Embedding with LLE

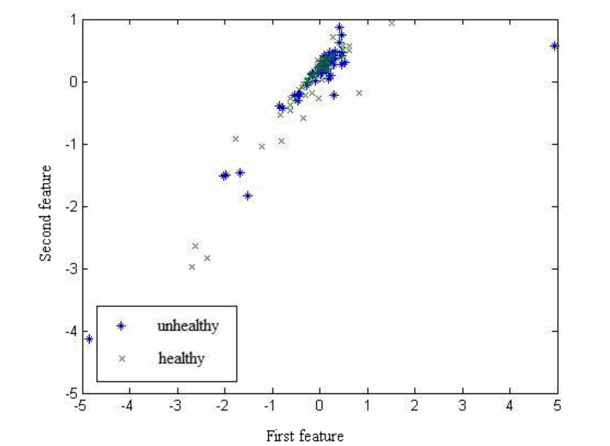
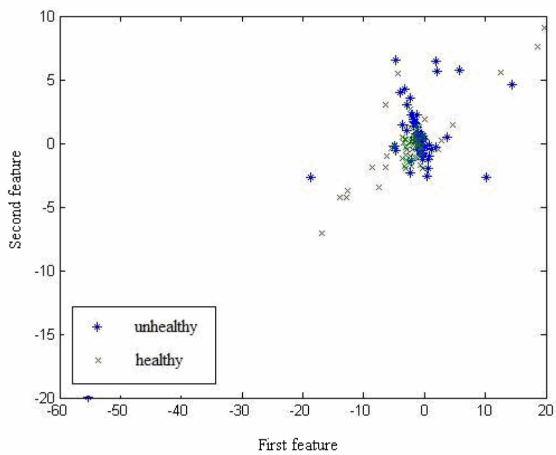


Figure 2

Two-Dimensional Embedding with PCA



For comparing the effectiveness of dimensionality reduction of the LLE approach, experiments are also performed on PCA. Figure 2 presents the two-dimensional embedding obtained. As Figure 2 shows, a lot of the data belonging to different classes is mixed and nearly superposed. Obviously, this superposed data makes the business failure prediction decision too complex, while using SVM-based classification there is much overlap of 'unhealthy' into the 'healthy' classifications than LLE. As for LLE, a structure can be recognized. However, PCA is not as distinct; this could lead to potential misclassification problems later.

4.2 Hybrid Model by LLE+SVM

After the dimensionality of the dataset is reduced, SVM classifiers are implemented. Table 2 shows the prediction performance of the SVM model with different dimension reduction methods based on five-fold cross-validation. In this experiment, SVM achieves accuracy ranging from 65.62% to 75.00% in predicting business failure with an average prediction accuracy of 72.50%. Our proposed hybrid model achieves the best prediction results with an average of 82.50%, outperforming other models. The prediction accuracy falls between 78.12% and 75.00% when using combined SVM and PCA, with the average prediction accuracy of 76.87%. Table 3 shows the t test result based on the average performance of these three models.

Table 2

The Performance of the SVM Model with Different Dimension Reduction Methods

Model	Testing set					
	Subset 1	Subset 2	Subset 3	Subset 4	Subset 5	Average accuracy
LLE+SVM	81.25	87.50	84.37	81.25	78.12	82.50
PCA+SVM	75.00	75.00	78.12	78.12	78.12	76.87
SVM	75.00	75.00	75.00	71.87	65.62	72.50

Table 3 shows the results of the t test to compare statistically the prediction performance of these three models. As shown in Table 3, LLE+SVM outperforms PCA+SVM and SVM at 5% statistical significance level, respectively. However, PCA+SVM does not significantly outperform SVM. Compared to "PCA+SVM", the results reveal that "LLE+SVM" is significantly superior. Apparently, LLE serves as a promising model to improve the classification accuracy of business failure prediction.

Table 3

Paired t Test of Results

	PCA+SVM	SVM
LLE+SVM	3.182 (0.012)*	4.129 (0.003)*
PCA+SVM		2.212 (0.057)
*<0.05		

4.3. Results Compared with Type I and Type II Errors of the Constructed Models

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 three discussed models. The results show that our hybrid techniques have comparatively fewer Type I and II errors, further proof of its all round superiority.

Table 4

Type I and Type II Errors of the Constructed Model

Model	Performance assessment	
	Type I error	Type II error
LLE+SVM	11.28%	23.50%
PCA+SVM	11.36%	35.21%
SVM	23.77%	30.26%

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 NLDR approach, such as LLE, has become very popular among scientists worldwide and is now one of the most developed techniques in dimensionality reduction analyses. Nevertheless, there has been progress in terms of the application of LLE in accounting and finance literature.

The objective of this study is to propose a hybrid model which is composed of the NLDR approach, the LLE algorithm and the SVM. In order to verify the applicability of this methodology, we also present combined PCA and SVM as the benchmark to a dataset on bankruptcies in Taiwan. In this paper, we presented results from visualization of financial data using LLE and we observed that the nonlinear embeddings of LLE separated certain data better than the linear projections of PCA. Therefore, the results show that the proposed hybrid models outperform PCA combined with the SVM model. The afore-mentioned findings justify the presumption that our hybrid approach model is a better alternative when conducting business failure prediction tasks. Furthermore, our hybrid approach model not only has comparatively better classification accuracy, but also produces fewer Type I and II errors. Thus, this forecasting technique can provide a useful indicator for decision makers.

However, there are some directions that will be explored in future work. 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 embeddings from LLE could be incorporated into other artificial intelligence approaches, such as random forest.

References

Ahn, B.S. Cho, S.S. and Kim, C.Y., 2000. The integrated methodology of rough set theory and artificial neural network for business failure prediction. *Expert Systems with Applications,* 18(2), pp. 65-74.

Altman, E.I., 1968. Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. *Journal of Finance,* 23(4), pp. 589-609.

Beaver, W.H., 1966. Financial ratios as predictors of failure, empirical research in accounting: selected studies. *Journal of Accounting Research,* 4: 179-199.

Blum, M., 1974. Failure company discriminant analysis. J*ournal of Accounting Research,* 12(1), pp. 1-25.

Brand, M., 2003. Charting a manifold. In: S. Becker, S. Thrun, and K. Obermayer, ed. 2003. *Advances in Neural Information Processing Systems 15*, Cambridge, MA: MIT Press.

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

Brun, A. Park, H.J. Knutsson, H. and Westin, C.F., 2003. Coloring of DT-MRI fiber traces using Laplacian eigenmaps. In: Proc. *the Ninth International Conference on Computer Aided Systems Theory* (2809).

Bryant, S. M., 1997. A case-based reasoning approach to bankruptcy prediction modelling. *Intelligent System Accounting, Financial and Management* 6: 195-214.

Chang C.C. and Lin, C.J., 2001. LIBSVM: A library for support vector machines, Available at: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

Chang, Y. Hu, C. and Matthew Turk, M., 2004. Probabilistic expression analysis on manifolds ", In: Proc. *the IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* 2: 520-527.

Chen, Y. W. and Lin, C. J., 2005). Combining SVMs with various feature selection strategies, Available at: <http://www.csie.ntu.edu.tw/~cjlin/papers/features.pdf>.

Coleman, K. G. Graettinger, T.J. and Lawrence, W.F., 1991. Neural networks for bankruptcy prediction: the power to solve financial problems. *AI Review,* 4: 48-50.

Deakin, E.B., 1972. A discriminant analysis of predictors of business failure. *Journal of Accounting Research,* 10(1), pp. 167-179.

Dimitras, A. I. Zanakis, S. H. and Zopounidis, C., 1996. A survey of business failures with an emphasis on prediction methods and industrial applications. *European Journal of Operational Research,* 90(3), pp. 487-513.

Ding, Y. S. Song, X.P. and Zen, Y.M., 2008. Forecasting financial condition of chinese listed companies based on support vector machine. *Expert Systems with Applications,* 34(4), pp. 3081-3089.

Elgammal, A. and Lee, C.S., 2004. Separating style and content on a nonlinear manifold. In: Proc. *the IEEE Computer Society Conference on Computer Vision and Pattern Recognition:* 478-489.

Elgammal, A. and Lee, C.S., 2004. Inferring 3D body pose from silhouettes using activity manifold learning. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2: 681-688.

Fan, A. and Palaniswami, M., 2000. Selecting bankruptcy predictors using a support vector machine approach. In: *Proceedings of the International Joint Conference on Neural Networks* 6: 354-359.

Gold, C. and Sollish, P., 2005. Model selection for support vector machine classification. *Neurcomputing,* 55: 221-249.

Hadid, A. Kouropteva, O. and Pietikainen, M., 2002. Unsupervised learning using locally linear embedding: experiments in face pose analysis. In: *Proceedings of the 16th International Conference on Pattern Recognition* I: 111-114.

Hsu, C.W. Chang, C.C. and Lin, C.J., 2003), A practical guide to support vector classification, Available at: <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.

Huang, Z. Chen, H. Hsu, C.J. Chen, W.H. and Wu, S. 2004. Credit rating analysis with support vector machines and neural networks: a market comparative study" *Decision Support Systems,* 37, pp. 543-558.

Jenkins, O. and Mataric, M., 2004. A spatio-temporal extension to isomap nonlinear dimension reduction. In: *Proceedings of the 21st International Conference on Machine Learning*.

Johnson, R. A. and Wichern, D. W. 5th Eds., 2002), *Applied Multivariate Statistical Analysis*, New Jersey: Prentice-Hall.

Jolliffe, I.T. Eds., 1986), *Principal Component Analysis*, New York: Springer-Verlag.

Kegl, B., 2003. Intrinsic dimension estimation using packing numbers. In: S. Becker, S. Thrun, and K. Obermayer, Ed. 2003, *Advances in Neural Information Processing Systems 15*, Cambridge, MA: MIT Press.

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

Li, H. and Sun, J., 2008. Ranking-order case-based reasoning for financial distress prediction. *Knowledge-Based Systems,* 21, pp.868-878.

Li, H. and Sun, J., 2009. Predicting financial distress using multiple case-based reasoning combine with support vector machine. *Expert Systems with Applications,* 36(6), pp. 10085-10096.

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

Lin, F. Yeh, C.C. and Lee, M.Y., 2011. The use of hybrid manifold learning and support vector machines in the prediction of business failure. *Knowledge-Based Systems*, 24(1), pp. 95-101.

Min, J.H. and Lee, Y.C., 2005. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications,* 28(4), pp. 603-614.

Min, S.H. Lee, J. and Han, I., 2006. Hybrid genetic algorithms and support vector machines for bankruptcy prediction. *Expert Systems with Applications,* 31(3), pp. 652–660.

Niskanen, M. and Silvn, O., 2003. Comparison of dimensionality reduction methods for wood surface inspection. In: *Proceedings of the 6th International Conference on Quality Control by Artificial Vision:* 178-188.

Ohlson, J.A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research,* 18(1), pp. 109-131.

Pettis, K. Bailey, T. Jain, A. and Dubes, R., 1979. An intrinsic dimensionality estimator from near-neighbor information. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI* 1(1), pp. 25-37.

Rahimian, E. and Singh, S. Thammachote, T. and Virmani, R., 1993. Bankruptcy prediction by neural networks. In: E. Trippi and E. Turban Ed. 1993, *Neural Networks in Finance and Investing: Using Artificial Intelligence to Improve Real-World Performance*, Chicago: Probus Publishing: 159-176.

Ribeiro, B. Silva, C. Sung, A. Vieira, A. Neves, J. Duarte, J. and Liu, Q., 2009. Learning Manifolds for Bankruptcy Analysis. In: M. Koppen *et al*. Ed. 2009, *Advances in Neuro-Information Processing, Lecture Notes in Computer Science*: 722-729.

Ribeiro, B. Vieira, A. and Neves, J., 2008. Supervised isomap with dissimilarity measures in embedding learning. In: *Proceedings of the Ibero-American Conference on Pattern Recognition*, *Progress in Pattern Recognition, Image Analysis and Applications, Lecture Notes in Computer Science*, Springer Berlin: Heidelberg: 389-396.

Roweis, S.T. and Saul, L.K., 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science,* 290, pp.2323-2326.

Salzberg, S. L., 1997. On comparing classifiers: pitfalls to avoid and a recommended approach. *Data Mining and Knowledge Discovery,* 1,pp.317-327.

Salchengerger, L.M. Cinar, E.M. and Lash, N.A., 1992. Neural networks: a new tool for prediction thrift failures. *Decision Science,s* 23,pp.899-916.

Seung, H.S. and Lee, D.D., 2000. The manifold ways of perception. *Science,* 290,pp. 2268-2269.

Sharda, R. and Wilson, R. L., 1996. Neural networks experiments in business-failure forecasting: predictive performance measurement issues. *International Journal of Computational Intelligence and Organizations* 1(2), pp. 107-117.

Shin, K.S. Lee, T.S. and Kim, H.J., 2005. An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications,* 28(1), pp.127-135.

de Silva, V. and Tenenbaum, J., 2003. Global versus local methods in nonlinear dimensionality reduction. In: *Advances in Neural Information Processing Systems* 15, Cambridge, MA: The MIT Press: 721-728.

Sinalingam, D.M. and Pandia, N., 2005. Minimal classification method with error correlation codes for multiclass recognization. *International Journal of Pattern Recognization and artificial intelligence* 5: 663-680.

Tam, K.Y. and Kiang, M., 1992. Managerial applications of neural networks: the case of bank failure predictions. *Management Science,* 38(7), pp. 926-947.

Tenenbaum, J.B., de Silva, V. and Langford, J.C., 2000. A global geometric framework for nonlinear dimensionality reduction. *Science,* 290, pp.2319-2323.

Van der Maaten, L.J. P. Postma, E.O. and Van den Herik, H.J., 2008. Dimensionality reduction: a comparative review. Submitted to *Neurocomputing*.

Van Gestel, T. Baesens, B. Suykens, J. Espinoza, M. Baestaens, D. E. Vanthienen, J. and De Moor, B., 2003. Bankruptcy prediction with least squares support vector machine classifiers. *IEEE International Conference on Computational Intelligence for Financial Engineering*: 1-8.

Vapnik, V.N., 1995. *The Nature of Statistical Learning Theory*, New York: Springer-Verlag.

Verikas, A. Kalsyte, Z. Bacauskiene, M. and Gelzinis, A., 2009. Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: a survey. *Soft Computing - a Fusion of Foundations, Methodologies and Applications,* 14(9), pp. 995-1010.

Wilson, R.L. and Sharda, R., 1994. Bankruptcy prediction using neural networks. *Decision Support Systems,* 11, pp.545-557.

Wu, C. H. Tzeng, G. H. Goo, Y. J. and Fang, W. C., 2007. A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy. *Expert Systems with Applications,* 32(2), pp. 397-408.

Yang, M.H., 2002. Face recognition using extended isomap. *IEEE International Conference on Image Processing* II, pp. 117-120.

Yeh, C.C. Chi, D.J. and Hsu, M.F., 2010. A hybrid approach of DEA, rough set and support vector machines for business failure prediction. *Expert Systems with Applications,* 37(2), pp. 1535-1541.

Zavgren, C. V., 1983. The prediction of corporate failure: The state of the art. *Journal of Financial Literature,* 2, pp.1-37.

Zhang, G. Hu, M. Y. Patuwo, B.E. and Indro, D. C., 1999. Artificial neural networks in bankruptcy predictions: General framework and cross-validation analysis. *European Journal of Operational Research,* 116, pp.16-32.

Zhang, J. Li, S.Z. and Wang, J., 2004. Nearest manifold approach for face recognition". In: *Proceedings of The 6th International Conference on Automatic Face and Gesture Recognition*.

Zhu, Y.S. and Zhang, Y.Y., 2003. The study on some problems of support vector classifier. *Computer Engineering and Applications,* 13, pp.66-68.