African Journal of Business Management Vol. 5(22), pp. 9355-9362, 30 September, 2011
Available online at http://www.academicjournals.org/AJBM
ISSN 1993-8233 ©2011 Academic Journals

Full Length Research Paper

A new corporate credit scoring system using semi-supervised discriminant analysis

Shian-Chang Huang

Department of Business Administration, National Changhua University of Education, No.2, Shi-Da Rd., Changhua, Taiwan. E-mail: shhuang@cc.ncue.edu.tw. Tel: 886-4-7232105-7420.

Accepted 18 July, 2011

Corporate credit scoring is important for investors and banks in risk management. However, the high dimensional data available from public financial statements make credit analysis difficult. To address the problem, dimensionality reduction is a key step to enhance scoring accuracy. By using semi-supervised discriminant analysis (SSDA) and support vector machines (SVMs), this study develops a novel system for credit scoring, where SSDA transforms high dimensional data space (over 50 financial variables) to a perfect low dimensional representative subspace with maximal discriminating power. Constructing SVM classifier in the new space effectively reduces overfitting and enhances classification accuracy. Empirical results indicate that SSDA is better than traditional dimensionality reduction schemes, and it significant improves SVM performance. More importantly, the new classification system substantially outperforms conventional classifiers. The new decision support system can help corporate bond investors make good assessments on their risks and substantially reduce their losses.

Key words: Semi-supervised discriminant analysis, dimensionality reduction, credit scoring, support vector machine, risk management.

INTRODUCTION

Recent financial crisis in 2007-2008 mainly results from credit risk. Corporate credit scoring is important in contemporary risk management. It determines the risk premiums requested by corporate bond investors. The investors with the most accurate estimation of corporate credit quality will be the most profitable. Typically, corporate credit scoring is costly to obtain, since they require agencies such as standard and poors or moody to invest heavily in terms of time and human resources to perform deep analysis of a firm’s risk status based on various aspects ranging from strategic competitiveness to operational details.

Investors and banks usually evaluate the credit quality of a corporation from publicly disclosed financial statements. The major difficulty is that they need to analyze the high dimensional data (over 50 financial variables) from financial statements. After financial crisis in 2007-2008, all banking and investment institutes attempt to search for a decision support system on evaluating the credit quality of their investments in corporate bonds. The objective of this study is thus to develop a reliable and accurate system for corporate credit scoring.

Recently, corporate credit scoring is a popular issue in academic and business community. Numerous classification techniques have been adopted for corporate credit scoring. These techniques include logistic regression (Stepanova and Thomas, 2001), Bayesian network, k-nearest neighbors (Henley and Hand, 1997), decision trees (Yobas et al., 2000), analytic network process (Chen et al., 2011), neural networks (Yobas et al., 2000; Tang and Chi, 2005; Abdou et al., 2008), support vector machines (Huang et al., 2004; Chen and Shih, 2006), and hybrid models (Lin, 2010).

Abbreviatios: SSDA, Semi-supervised discriminant analysis; **SVMs,** support vector machines; **PCA,** principal component analysis; **LDA,** linear discriminant analysis; **LDA,** linear discriminant analysis; **MSVM,** multi-class support vector machines; **LR,** logistic regressions; **BN,** Bayesian networks; **RFE,** recursive feature elimination.

In credit scoring, large scale financial data usually make conventional classifiers infeasible due to the curse of dimensionality (Bellman, 1961). Consequently, one needs to select key features or transform the input space of data for the classifiers first. Among dimensionality reduction, the linear algorithms such as principal component analysis (PCA) (Mardia et al., 1980) and linear discriminant analysis (LDA) (Fukunaga, 1990) have been the two most popular because of their relative simplicity and effectiveness.

When label information available, for example, for classification task, LDA can achieve significant better performance than PCA. However, when there is no sufficient training (labeled) samples, the covariance matrix of each class may not be accurately estimated. In this case, the generalization capability on testing (unlabeled) samples can not be guaranteed. A possible solution to deal with insufficient labeled samples could be learning on both labeled and unlabeled data (semi-supervised learning). It is natural and reasonable since in reality we usually have only part of data labeled, along with a large number of unlabeled (testing) data. In the last decades, semi-supervised learning (or transductive learning) has attracted an increasing amount of attention. Recently, there are considerable interest and success on semi-supervised learning algorithms (Belkin et al., 2006; Sindhwani et al., 2005; Zhou et al., 2003; Zhu et al., 2003).

This research adopted the semi-supervised discriminant analysis (SSDA) of Cai et al. (2007) for dimensionality reduction to enhance the performance of conventional classifiers. Using SSDA, one could find a good low dimensional projection which respects the discriminant structure inferred from the labeled data points, as well as the intrinsic geometrical structure inferred from both labeled and unlabeled data points. Namely, the labeled and unlabeled data points are combined to build a graph incorporating neighborhood information of the data set. The graph provides a discrete approximation to the local geometry of the data space. Constructing classifiers in the SSDA projection space significantly reduces computational loading and simultaneously enhance classifier performance. This study applies six conventional classifiers for credit scoring, including: nearest neighbors, logistic regressions, Bayesian networks, one-vs-one SVM, one-vs-rest SVM, and multi-class SVM. Empirical results indicated that the performance improvement by SSDA is significant. The combination of SSDA and one-vs-rest SVM performs best. The method developed here will help financial institutions make good assessments about their credit risk, and substantially reduce their losses.

SEMI-SUPERVISED DISCRIMINANT ANALYSIS

Linear discriminant analysis

Linear discriminant analysis (LDA) seeks directions on which the data points of different classes are far from each other while requiring data points of the same class to be close to each other. Suppose we have a set of l samples $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_l \in \mathbf{R}^n$, belonging to c classes.

The objective function of LDA is as follows:

$$\mathbf{a}_{opt} = \arg\max_{\mathbf{a}} \frac{\mathbf{a}^T S_b \mathbf{a}}{\mathbf{a}^T S_w \mathbf{a}}, \quad (1)$$

$$S_b = \sum_{k=1}^{c} l_k (\mu^{(k)} - \mu)(\mu^{(k)} - \mu)^T, \quad (2)$$

$$S_w = \sum_{k=1}^{c} \sum_{i=1}^{l_k} (\mathbf{x}_i^{(k)} - \mu^{(k)})(\mathbf{x}_i^{(k)} - \mu^{(k)})^T, \quad (3)$$

where S_w stands for the within-class scatter matrix and S_b the between-class scatter matrix. In Equation 2 and 3, μ is the total sample mean vector, l_k is the number of samples in the $k-th$ class, $\mu^{(k)}$ is the average vector of the $k-th$ class, and $\mathbf{x}_i^{(k)}$ is the $i-th$ sample in the $k-th$ class. Define the total scatter matrix $S_t = \sum_{i=1}^{l} (\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^T$, we have $S_t = S_b + S_w$ (Fukunaga, 1990). The objective function of LDA in Equation 1 is equivalent to:

$$\mathbf{a}_{opt} = \arg\max_{\mathbf{a}} \frac{\mathbf{a}^T S_b \mathbf{a}}{\mathbf{a}^T S_t \mathbf{a}}. \quad (4)$$

The optimal $\mathbf{a}'s$ are the eigenvectors corresponding to the non-zero eigenvalue of eigen-problem:

$$S_b \mathbf{a} = \lambda S_t \mathbf{a}. \quad (5)$$

The implementation of semi-supervised discriminant analysis (SSDA)

LDA considers seeking the optimal projections purely on the training (labeled) set. In reality, it is possible to acquire a large set of unlabeled data. This section is trying to extend LDA model to incorporate the manifold structure illustrated by unlabeled data.

LDA aims to find a projection vector $\mathbf{a}$ such that the ratio between $\mathbf{a}^T S_b \mathbf{a}$ and $\mathbf{a}^T S_t \mathbf{a}$ is maximized. When there is no sufficient training sample, overfitting may happen. A typical way to prevent overfitting is to impose a regularizer (Hastie et al., 2001). The optimization problem of the regularized version of LDA can be written as follows:

$$\max_{\mathbf{a}} \frac{\mathbf{a}^T S_b \mathbf{a}}{\mathbf{a}^T S_t \mathbf{a} + \alpha J(\mathbf{a})} \quad (6)$$

where $J(\mathbf{a})$ controls the learning complexity of the hypothesis family, and the coefficient α controls balance between the model complexity and the empirical loss.

The regularizer term $J(\mathbf{a})$ provides us the flexibility to incorporate our prior knowledge on some particular applications. When a set of unlabeled examples available, we aim to construct a $J(\mathbf{a})$ incorporating the geometric structure of data. For classification, it means nearby points are likely to have the same label. For dimensionality reduction, it can be interpreted as nearby points will have similar embeddings (low-dimensional representations). Following Cai et al. (2007), given a set of examples, one can use a p-nearest neighbor graph G to model the relationship between nearby data points. Specifically, we put an edge between nodes i and j if $\mathbf{x}_i$ and $\mathbf{x}_j$ are close, that is, $\mathbf{x}_i$ and $\mathbf{x}_j$ are among p nearest neighbors of each other. Let the corresponding weight matrix be S, defined by:

$$S_{ij} = \begin{cases} 1 & if\ \mathbf{x}_i \in N_p(\mathbf{x}_j)\ or\ \mathbf{x}_j \in N_p(\mathbf{x}_i); \\ 0 & otherwise \end{cases} \quad (7)$$

where $N_p(\mathbf{x}_i)$ denotes the set of p-nearest neighbors of $\mathbf{x}_i$. In general, the mapping function should be as smooth as possible on the graph.

Specifically, if two data points are linked by an edge, they are likely to be in the same class. Moreover, the data points lying on a densely linked subgraph are likely to have the same label. Thus, a natural regularizer can be defined as follows:

$$J(\mathbf{a}) = \sum_{ij} (\mathbf{a}^T \mathbf{x}_i - \mathbf{a}^T \mathbf{x}_j)^2 S_{ij}. \quad (8)$$

This formulation is motivated from spectral dimensionality reduction techniques (Belkin and Niyogi, 2001; He and Niyogi, 2003). By incorporating the data dependent regularizer, the SSDA algorithm makes effectively use of labeled and unlabeled data points for classification. For more details regarding the algorithm, we refer to Cai et al. (2007).

MULTI-CLASS SUPPORT VECTOR MACHINES

Recently, support vector machines (Cristianini and Shawe-Taylor, 2000; Schoelkopf et al., 1999; Scholkopf et al., 2002), another form of neural networks, has been gaining popularity and has been regarded as the state-of-the-art technique for regression and classification applications. SVM overcomes the following disadvantages of traditional neural networks models: (I) dependency on a large number of parameters, for example, network size, learning parameters and initial weight chosen, (II) possibility of being trapped into local minima resulting in a very slow convergence, and (III) over-fitting on training data resulting in a poor generalization ability.

SVMs are a set of related supervised learning methods used for classification and regression. SVMs were proposed by Vapnik (1999). Viewing input data as two sets of vectors (two classes classification) in an high dimensional transformed space, an SVM seeks to construct a separating hyperplane in that space, one which maximizes the margin between the two data sets. To calculate the margin, two parallel hyperplanes are constructed, one on each side of the separating hyperplane, which are "pushed up against" the two data sets. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the neighboring datapoints of both classes, since in general the larger the margin the better the generalization error of the classifier. That is, based on the structured risk minimization principle, SVMs seek to minimize an upper bound of the generalization error instead of the empirical error as in neural networks. The SVM classification function is formulated as follows:

$$y = sign(\mathbf{w}^T \varphi(\mathbf{x}) + b), y \in \{-1, 1\}, \quad (9)$$

where y is output (1 for type A, -1 for type B), $\mathbf{x}$ the input; $\varphi(\mathbf{x})$ is a nonlinear mapping from the input space to the high dimensional transformed space. SVMs exploit the idea of mapping input data into a high dimensional feature space where classification could be easily performed. Coefficients W and b are estimated by the following optimization problem:

$$\min_{\mathbf{w},b} R(\mathbf{w}, \xi) = \frac{1}{2} \| \mathbf{w} \|^2 + \mathrm{C} \sum_{i=1}^{m} \xi_i \quad (10)$$

with

$$y_i(\mathbf{w}^T \varphi(\mathbf{x}) + b) \geq 1 + \xi_i, \quad i = 1, ..., m.$$
$$\xi_i \geq 0,$$

where C is a prescribed parameter to evaluates the

Table 1. The financial variables.

Profitability index	Per-share rate index (dollar)	Growth rate index	Debt-paying ability index	Management ability index
ROA(A)-EBIT%	BPS(B)	YOY%-sales	C/F adequacy ratio%	Inv. and A-R/equity
ROA(C)-EBID%	BPS(A)	Gross margin growth%	Cash reinvest %	Total asset turnover
ROA(B)-EBITD%	BPS(C)	YOY%-	Current %	A/R and N/R turnover
ROE(A)-NI%	EPS-net income	Real. GM	Acid test%	Days-A/R turnover
ROE(B)-NI% exdisposal%	PS-cashflow	YOY%-	Interest exp. %	Inventory turnover
Real. gross profit%	PS-sales	Oper.Income	D/E ratio %	Days-inventory turn.
Operating income %	PS-operating income	YOY%-pre-tax income	Liabilities %	Fixed asset turnover
Net income% -exc. disp.	PS-pre tax income	YOY%-net income	Equity/TA %	Equity turnover
		YOY%-	(L-T Liab.+SE)/FA %	Days-A/P turnover
		Ordin.Income	Debt/Equity %	Net operating cycle
		YOY%-	Contingent liab %	
		Recurring Inc.	Times interest earned	
		YOY%-total assets	Oper. income/capital	
		YOY%-total equity	Pretax incom/capital	
		YOY%-fixed assets		
		YOY%-return on TA		
		Retention ratio		

trade-off between the empirical risk and the smoothness of the model.

One approach to solving multi-class classification problem by SVM is to consider the problem as a collection of binary classification problems. For c categories, c classifiers can be constructed, one for each class. The k-th classifier constructs a hyperplane between class k and the $c-1$ other classes. A majority vote across the classifiers or some other measure can then be applied to classify a new point. This is one-against-rest SVM.

Alternatively, $C_2^c = \frac{c(c-1)}{2}$ hyperplanes can be constructed, separating each class from each other class, and similarly some voting schemes can be applied. This is one-against-one SVM. The above two methods have been used widely in the support vector literature to solve multi-class classification problems. Still another way to solve multi-class problems is to construct a decision function by considering all classes at once, which is multi-class support vector machines (MSVM) developed by Weston and Watkins (1999).

EMPIRICAL RESULTS AND ANALYSIS

The financial variables of this research are selected from the Taiwan Economic Journal (TEJ) Financial database, which contains the following five financial indexes: profitability index, per share rates index, growth rates index, debt-paying ability index, management ability index. Altogether, there are 53 financial ratios covered by the 5 indexes. These financial ratios are listed in Table 1 (we refer to TEJ for variable details). If some values of a ratio lose on some firms, this ratio is deleted. The deleted ratios are Net Income%-Exc Disp, YOY%-Fixed Assets, C/F Adequacy Ratio%, Cash Reinvest%, Interest Exp.%, (L-T Liab.+SE)/FA%, Contingent Liab%, and Times Interest Earned. As a result, overall 45 financial ratios were obtained for analysis.

These variables are derived from publicly disclosed information that a company is required to file with investors. Besides the financial variables, this study also included the historical rating of each company to improve the rating accuracy.

Credit quality or rating information on target companies was also obtained from the TEJ, which provides the credit quality rating for every publicly

Table 2. Forecasting performance (error rates) of conventional models.

Model	2000	2001	2002	2003	2004
Nearest neighbors(NN1)	0.2661	0.2232	0.2286	0.2446	0.3427
Nearest neighbors(NN3)	0.1768	0.2196	0.2554	0.2446	0.2661
Logistic regression	0.2564	0.2676	0.1944	0.2557	0.2597
Bayesian network	0.1923	0.1690	0.1806	0.2208	0.2468
1-vs-1 pure SVM	0.2161	0.2232	0.1625	0.1929	0.1786
1-vs-rest pure SVM	0.2143	0.1821	0.1750	0.2679	0.2661
Pure MSVM	0.2375	0.2804	0.2768	0.3304	0.2143

traded Taiwanese company. A TEJ rating indicates a company's capacity to meet its financial commitments over a one-year period, and is classified as: low risk, medium risk and high risk. A low risk rating indicates that an organization has an extremely strong capacity to meet its commitments, whereas a high risk rating indicates that an organization is likely to default.

This study tests six conventional classifiers for corporate credit quality forecasting, including: nearest neighbors (with one and three neighbors, NN1, NN3), logistic regressions (LR), Bayesian networks (BN), one-vs-one SVM (1vs1), one-vs-rest SVM (1-rest), and MSVM. This study collects 88 Taiwanese high technology companies that are traded on the Taiwan's security market. There is one rating for each company every year. The data sampling period is from 2000 to 2004. The data set is randomly divided into ten parts, and ten-folds cross validation is applied to evaluate the model performance.

Performance comparison

Comparing with convectional models

Table 2 shows the average error rates of all methods. The logistic regression and nearest neighbors models have similar performance, but their performance are slightly inferior to the Bayesian network model. Comparing performance among SVM classifiers, Table 2 clearly revealed that one-versus-one SVM performs best. On average, the one-vs-one SVM also outperforms other traditional classifiers. Consequently, in the next experiments we take SVM classifiers as base classifiers to develop new classification systems.

Comparing with other dimensionality reduction schemes

Due to the curse of dimensionality, the performance of traditional classifiers may be degraded by irrelevant variables. It is believed that suitable feature selection or dimensionality reduction schemes will improve their performance. Thus this study integrate some feature selection and dimensionality reduction algorithms in these classifiers, and made a further comparison. First, the SSDA is employed to enhance the performance of three SVM classifiers. We then compared the performance improvement of SSDA with other dimensionality reduction algorithms, the principal component analysis (PCA) and Independent Component Analysis (ICA) (Hyvärinen et al., 2001). This study sets the dimension of subspace to five for all algorithms. The results are listed in Table 3.

Wilcoxon tests

The Wilcoxon rank-sum test (Wilcoxon, 1945) is a nonparametric alternative (for sample median) to the two sample t-test which is based solely on the order in which the observations from the two samples fall. For performance comparison, the Wilcoxon rank-sum tests of the best model (one-vs-one SVM + SSDA) with conventional classifiers are reported in Tables 4 and 5, and the Wilcoxon comparison with other dimensionality reduction methods is displayed in Table 6.

Comparing with feature selection scheme

Additionally, we also compare the performance improvement of SSDA with a famous feature selection algorithm, the recursive feature elimination (RFE) method proposed by Guyon et al. (2002). RFE algorithm recursively eliminates input variables to identify the most important five, ten, fifteen, and twenty features for comparison. Table 7 lists the results. Tables 3 and 6 also list pure classifiers without any dimensionality reduction and feature selection schemes for comparison.

Results analysis

Table 3 shows that only the SSDA algorithm can significantly improve the performance of SVM classifiers. PCA and ICA are not very effective. The one-vs-one SVM with SSDA achieves the highest accuracy than other

Table 3. Performance comparison (error rates) of three dimensionality reduction schemes.

Model	2000	2001	2002	2003	2004
1-vs-1 Pure SVM	0.2161	0.2232	0.1625	0.1929	0.1786
1-vs-rest Pure SVM	0.2143	0.1821	0.1750	0.2679	0.2661
Pure MSVM	0.2375	0.2804	0.2768	0.3304	0.2143
1-vs-1 + PCA	0.1661	0.2643	0.2393	0.2339	0.2964
1-vs-rest + PCA	0.2036	0.2500	0.2411	0.2071	0.2821
MSVM + PCA	0.2786	0.2214	0.4375	0.2589	0.3946
1-vs-1 + ICA	0.2750	0.1929	0.2196	0.3071	0.2911
1-vs-rest + ICA	0.3250	0.1929	0.2161	0.3714	0.4268
MSVM + ICA	0.3518	0.2214	0.2054	0.3589	0.3304
1-vs-1 + SSDA	0.0750	0.0982	0.0133	0.0900	0.0250
1-vs-rest + SSDA	0.0875	0.0839	0.0810	0.1925	0.1542
MSVM + SSDA	0.2925	0.3232	0.3429	0.3467	0.2567

Table 4. Wilcoxon tests on the differences between the best hybrid model and conventional classifiers (p-value).

Model	1vs1+SSDA	1vs1	1-rest	MSVM
1vs1+SSDA	1.0000	0.0079	0.0079	0.0079
1vs1	0.0079	1.0000	0.5476	0.0317
1-rest	0.0079	0.5476	1.0000	0.1111
MSVM	0.0079	0.0317	0.1111	1.0000

Table 5. Wilcoxon tests on the differences between the best hybrid model and conventional classifiers (p-value).

Model	1vs1+SSDA	BN	LR	NN3	NN1
1vs1+SSDA	1.0000	0.0079	0.0079	0.0079	0.0079
BN	0.0079	1.0000	0.0317	0.3095	0.0556
LR	0.0079	0.0317	1.0000	0.3095	1.0000
NN3	0.0079	0.3095	0.3095	1.0000	0.5079
NN1	0.0079	0.0556	1.0000	0.5079	1.0000

Table 6. Wilcoxon tests on the differences among SSDA, ICA, and PCA (p-value).

Model	1vs1+SSDA	1vs1+ICA	1vs1+PCA	1vs1	1-rest	MSVM
1vs1+SSDA	1.0000	0.0079	0.0079	0.0079	0.0079	0.0079
1vs1+ICA	0.0079	1.0000	0.6905	0.0635	0.1508	0.8413
1vs1+PCA	0.0079	0.6905	1.0000	0.0952	0.8413	0.5476
1vs1	0.0079	0.0635	0.0952	1.0000	0.5476	0.0317
1-rest	0.0079	0.1508	0.8413	0.5476	1.0000	0.1111
MSVM	0.0079	0.8413	0.5476	0.0317	0.1111	1.0000

multi-class classifiers. Under 10% level of significance, Tables 4 and 5 clearly reveal that one-vs-one SVM with SSDA outperforms other pure classifiers. Table 6 shows that SSDA distinctly outperforms other dimensionality reduction schemes. These results fully demonstrate that in real rating problems training (labeled) and testing unlabeled) data provide different information for classification. The labeled data are used to maximize the

Table 7. Performance comparison (error rates) of SSDA and RFE.

Model	2000	2001	2002	2003	2004
1-vs-1 Pure SVM	0.2161	0.2232	0.1625	0.1929	0.1786
1-vs-rest Pure SVM	0.2143	0.1821	0.1750	0.2679	0.2661
Pure MSVM	0.2375	0.2804	0.2768	0.3304	0.2143
1-vs-1 + RFE 5	0.1250	0.2536	0.0250	0.1393	0.1536
1-vs-1 + RFE 10	0.1375	0.1946	0.0679	0.0893	0.1393
1-vs-1 + RFE 15	0.1250	0.2214	0.1107	0.1018	0.1268
1-vs-1 + RFE 20	0.1750	0.2625	0.1339	0.1554	0.1518
1-vs-rest + RFE 5	0.1268	0.2786	0.0393	0.1536	0.1893
1-vs-rest + RFE 10	0.1625	0.1946	0.0929	0.1661	0.1929
1-vs-rest + RFE 15	0.1500	0.1679	0.1089	0.1643	0.1393
1-vs-rest + RFE 20	0.1893	0.1643	0.1500	0.2286	0.2429
MSVM + RFE 5	0.3071	0.3054	0.3304	0.2536	0.2661
MSVM + RFE 10	0.2518	0.2929	0.3161	0.2661	0.2643
MSVM + RFE 15	0.2375	0.2643	0.3036	0.3179	0.2786
MSVM + RFE 20	0.2625	0.2786	0.2750	0.2679	0.2768
1-vs-1 + SSDA	0.0750	0.0982	0.0133	0.0900	0.0250
1-vs-rest + SSDA	0.0875	0.0839	0.0810	0.1925	0.1542
MSVM + SSDA	0.2925	0.3232	0.3429	0.3467	0.2567

Table 8. Wilcoxon tests on the difference between SSDA and RFE.

Model	1vs1+SSDA	1vs1 +RFE 5	1vs1 +RFE 10	1vs1+RFE 15	1vs1+RFE 20
1vs1+SSDA	1.0000	0.0635	0.1508	0.0079	0.0079
1vs1+RFE 5	0.0635	1.0000	0.7302	0.5952	0.3095
1vs1+RFE 10	0.1508	0.7302	1.0000	1.0000	0.2222
1vs1+RFE 15	0.0079	0.5952	1.0000	1.0000	0.0952
1vs1+RFE 20	0.0079	0.3095	0.2222	0.0952	1.0000

(separability between different classes, while the unlabeled data are used to estimate the intrinsic geometry of the data space. Unsupervised algorithms such as PCA and ICA could not extract information containing in the unlabeled data. Hence, their performance is poor. Considering graph or manifold-based semi-supervised learning for dimensionality reduction are more effective.

Comparing SSDA with RFE, Table 7 reveals the new model, one-vs-one SVM with SSDA is the most cost-efficient model, since it reduces original data to the fewest dimensional subspace and achieves the best accuracy. Under 10% level of significance, Table 8 displays the excellence of SSDA, in which SSDA is only indifferent to RFE with ten features, but SSDA uses just five (a half) dimensions to represent the data. Clearly, the accuracies of the pure one-vs-one, one-vs-rest, and MSVM classifiers, which contain all of the variables which are lower than for the hybrid classifier containing fewer variables. That is, more information does not necessarily improve accuracy. Table 7 also shows that the performance improvement owing to RFE is lower than SSDA no matter 5, 10, 15, 20 key features are selected. The subspace formed by SSDA is a smooth low dimensional approximation of data space which simultaneously maintains the discriminating power, while the subset formed by RFE does not.

Conclusions

Corporate credit scoring provides important information on credit risk for banks and investors in financial markets. This study employed SSDA to enhance conventional classifiers. The performance improvement was examined using a data set comprising a large amount of financial information regarding Taiwanese high technology companies. The empirical results showed that the new hybrid system (one-vs-one SVM with SSDA) is more accurate and robust than traditional classifiers in credit scoring.

SSDA maps high dimensional data space to a good low dimensional representative subspace. In SSDA, the labeled data are used to maximize the discriminating power, while both labeled and unlabeled data are used to estimate the intrinsic geometry of data space. Empirical results showed that one-versus-one SVM with SSDA performs best. The classifiers using all input variables are less accurate than models using fewer but more important latent variables. Compared with traditional dimensionality reduction algorithms, the performance improvement obtained by SSDA is significant and robust; namely, semi-supervised learning is a key technique for improving the performance of multi-class classifiers. More importantly, combining SSDA with conventional classifiers could reduce their computational loadings and simultaneously enhance their performance. The decision support system developed in this study can help corporate bond investors make good assessments on their credit risk and substantially reduce their losses.

Future research may consider non-financial and macroeconomic variables for inputs. But including more information does not guarantee higher accuracy. In this situation, nonlinear dimensionality reductions are important strategies for enhancing classifier performance. What type of supervised or semi-supervised subspace learning algorithm is efficient to incorporate with conventional classifiers need further study.

REFERENCES

Abdou H, Pointon J, El-Masry A (2008). Neural nets versus conventional techniques in credit scoring in Egyptian banking. Expert Syst. Appl., 35(3): 1275-1292.

Belkin M, Niyogi P (2001). Laplacian eigenmaps and spectral techniques for embedding and clustering. In Advances in Neural Information Processing Systems. MIT Press. Cambridge. MA, 14: 585-591.

Belkin M, Niyogi P, Sindhwani V (2006). Manifold regularization: A geometric framework for learning from examples. J. Machine Learning Res., 7: 2399-2434.

Bellman R (1961). Adaptive Control Processes: A Guided Tour. Princeton University Press.

Cai D, He X, Han J (2007). Semi-Supervised Discriminant Analysis. Proceedings of the Eleventh IEEE International Conference on Computer Vision (ICCV2007).

Chen YS, Lin CT, Lu JH (2011). The analytic network process for the banking sector: An approach to evaluate the creditability of emerging industries. Afr. J. Bus. Manage., 5(4): 1343-1352.

Chen WH, Shih JY (2006). A paper of Taiwan's issuer credit rating systems using support vector machines. Expert Syst. Appl., 30(3): 427-435.

Cristianini N, Shawe-Taylor J (2000). An Introduction to Support Vector Machines. Cambridge University Press.

Fukunaga K (1990). Introduction to Statistical Pattern Recognition. Academic Press. 2nd edition.

Guyon I, Weston J, Barnhill S, Vapnik V (2002). Gene selection for cancer classification using support vector machines. Machine Learn., 46: 389-422.

Hastie T, Tibshirani R, Friedman J (2001). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer-Verlag.

He X, Niyogi P (2003). Locality Preserving Projections. Advances in Neural Information Processing Systems. 16 (NIPS 2003). Vancouver. Canada.

Henley WE, Hand DJ (1997). Construction of a k-nearest neighbour credit-scoring system. IMA J. Manage. Math., 8: 305-321.

Huang Z, Chen HC, Hsu CJ, Chen WH, Wu SS (2004). Credit rating analysis with support vector machines and neural networks: A market comparative paper. Decis. Support Syst., 37: 543-558.

Hyvärinen A, Karhunen J, Oja E (2001). Independent Component Analysis. Wiley Interscience.

Lin SL (2010). A two-stage logistic regression-ANN model for the prediction of distress banks: Evidence from 11 emerging countries. Afr. J. Bus. Manage., 4(14): 3149-3168.

Mardia KV, Kent JT, Bibby JM (1980). Multivariate Analysis. Academic Press.

Scholkopf B, Smola AJ (2002). Learning with Kernels Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press. Cambridge. MA.

Schoelkopf B, Burges CJC, Smola AJ (1999). Advances in kernel methods - support vector learning. MIT Press. Cambridge. MA.

Sindhwani V, Niyogi P, Belkin M (2005). Beyond the point cloud: from transductive to semi-supervised learning. In Proc. 2005 Int. Conf. Machine Learning (ICML05).

Stepanova M, Thomas LC (2001). PHAB scores: proportional hazards analysis behavioural scores. J. Oper. Res. Soc., 52: 1007-1016.

Tang TC, Chi LC (2005). Neural networks analysis in business failure prediction of Chinese importers: a between-countries approach. Expert Syst. Appl., 29(2): 244-255.

Vapnik VN (1999). The nature of statistical learning theory. second edition. New York, Springer.

Weston J, Watkins C (1999). Support Vector Machines for Multi-Class Pattern Recognition. Proceedings of the Seventh European Symposium On Artificial Neural Networks 1999 (ESANN'99).

Wilcoxon F (1945). Individual comparisons by ranking methods. Biometrics, 1: 80-83.

Yobas MB, Crook JN, Ross P (2000). Credit scoring using neural and evolutionary techniques. IMA J. Manage. Math., 11: 111-125.

Zhou D, Bousquet O, Lal T, Weston J, Schölkopf B (2003). Learning with local and global consistency. Adv. Neural Inf. Process. Syst., p. 16.

Zhu X, Ghahramani Z, Lafferty J (2003). Semi-supervised learning using gaussian fields and harmonic functions. In Proc. of the twentieth Internation Conference on Machine Learning.