

Exploiting Corporate Governance and Common-Size Analysis for Financial Distress Detecting Models

Pei-Wen Huang and Chao-Lin Liu

Department of Computer Science, National Chengchi University, Wen-Shan, Taipei 11605, Taiwan
{g9313, chaolin}@cs.nccu.edu.tw

Abstract

Traditionally, statistical techniques such as multivariate discriminant analysis and logistic regression analysis have been applied for predicting financial distresses by analyzing financial ratios. In addition to statistical methods, recent studies suggest that backpropagation neural networks (BPNs) and support vector machines (SVMs) can be alternative approaches for classification tasks. Hence, we construct two software classifiers, BPNs and SVMs, and then investigate the effects of employing features related to corporate governance and common-size analysis in financial distress model. Experimental results indicate that the proposed features may help SVMs achieve better predication quality when we try to predict financial distresses with more temporally distant data and smaller data set.

Keywords: backpropagation neural networks, support vector machines, corporate governance, common-size analysis, financial distresses

1. Introduction

The problem of early detecting financially distressed companies attracts the public's great attention after the notorious case about Enron. This expensive lesson cost the public around 40 billion as well as the faith in CPA (Certified Public Accountant) firms and the entire investing market. Therefore, we need a more powerful mechanism to assist the public for early warning of landmine stocks. Traditionally, researchers apply such techniques as multivariate discriminant analysis [1] or artificial neural networks [2]–[5] for predicting financial distresses. However, the previous works were founded on the analysis of financial ratios that managers could window-dress through delicate earning reports. In summary, to uncover financially distressed companies with management frauds is a difficult task by only using financial ratios as the feature variables. First, if the earning numbers of financial statements are deliberately fabricated to deceive the public, then financial ratios turn out to be misleading information [6]. Secondly, given new tricks emerged from endless financial shenanigans, financial ratios do not provide

sufficient insight. Furthermore, the study conducted by Fanning et al. [7] suggest that non-financial factors available in the published financial statements can improve the accuracy in early detection, especially those cases that resulted from management frauds. Hence, we introduce two new factors, *common-size analysis* [6] and *corporate governance* [8]–[11] for achieving better prediction accuracy, which is measured by the F measure [12]. Moreover, we employed the backpropagation neural networks (BPNs) and support vector machines (SVMs) as the classifiers in our experiments and compared prediction qualities achieved by using different sets of features that were obtained based on contents of the published financial statements.

Common-size analysis (CSA) converts selected items in financial statements to percentages of size-related measures. When we utilize this technique to analyze balance sheets, items of balance sheets are expressed in proportion to total assets. This is the same with income statements where items are expressed in proportion to net sales or net income. Using CSA, we can assess the financial positions of different-sized companies and of the same company over different periods. CSA is also useful when we compare companies in the same industry to see if they have similar financial structures. For instance, the variations of sales in income statements should correspond to the variations of inventories, account receivables and cash and cash equivalents in balance sheets. Hence, an unreasonable proportion may indicate a red flag of cooking books. Table 1, on the next page, snapshots some red flags for financial distresses, each pointing to a trick or a symptom.

In addition, we need non-financial features to better estimate the risk of financial distresses, especially for companies with management frauds. Hence, we exploited the concept of corporate governance in our research. By the definition provided by OECD (Organization for Economic Co-operation and Development), corporate governance is the system by which business corporations are directed and controlled. The corporate governance structure specifies the distribution of rights and responsibilities among different participants in the corporation, such as the board, managers, shareholders and other stakeholders, and spells out the rules and pro-

Table 1. Some red flags and related tricks

red flags	tricks or symptoms
1. Cash and cash equivalents sharply drops in proportion to total assets.	Poor liquidity might increase loan in the very near future.
2. The growth of account receivables highly increases than COGS.	The principal of recognizing revenue is too aggressive or the management might cook the book.
3. The amount of fixed assets widely increases in proportion to total assets.	The management might capitalize current expenses as assets.
4. The growth of operating expenses obviously descends in proportion to net sales.	The management might capitalize current expenses as assets, or illegally cook the book.
5. Net income before tax profits mainly from side jobs.	Decline of major business
6. Cash flow comes mainly from financing activity like disposal of assets or issuing convertible bonds.	The management might expropriate the wealth of the company or signify decline of major business.

※ This table is extracted from H. Schilit [6].

cedures for making decisions on corporate affairs. By doing this, it also provides both the structure through which the company objectives are set and the means of attaining those objectives and monitoring performances.

Corporate governance is regarded as one of the key factors that caused the Asian financial crisis in 1997. Corporate governance variables provide better explanatory power for the crisis than macroeconomics-related factors [8]. La Porta et al. [9] suggest that when large shareholders effectively control a firm, they might try to expropriate wealth by seeking personal benefit at the expense of minority shareholders, especially in countries without strong investor protection. This is consistent with the conflict of interest between majority and minority shareholders. Lee et al. [10] adopt three variables to represent the corporate governance risk, namely, the percentage of directors occupied by the controlling shareholders, the percentage of the controlling shareholders' shareholding that has been pledged for bank loan, and the deviation between actually controlling power and shareholding. The evidence supports that these three variables of corporate governance risk mentioned above are positively related to the risk for financial distresses in the following year [10]. In summary, weak corporate governance is closely related to the risk of financial distresses.

Given the training data with known labels, we can construct classification models and input test data to verify the accuracy of models. This is the fundamental idea of machine learning. Presently, there are lots of known approaches, such as decision trees, Bayesian networks, artificial neural networks etc., applicable to the classification tasks in many fields including finance and accounting. Support Vector Machines (SVMs), have become a very popular classification tool in recent years. The main idea of SVMs is to find several support vectors to form decision boundaries which separate data clusters. Since SVMs capture geometric character-

istics of the feature space without deriving weights from the subset data, they have the potential to extract the optimal solution with smaller data sets [13]. Fan et al. [14] adopt utilize SVMs to select bankruptcy predictors. Other researchers propose a hybrid model that use SVMs as an underlying classifier [15].

Artificial neural networks (ANNs) are widely applied to prediction task of financial and accounting problems because they are capable of identifying and representing non-linear relationship in the data set [3]–[5],[7]. In general, the most common supervised learning algorithm is *backpropagation algorithms*. The training processes of backpropagation neural networks include two stages, *feedforward* and *backpropagation*. In the feedforward stage, BPNs import input vectors into the input layer and forward them to hidden layer and then to the output layer. The synaptic weights are fixed at this stage. In the backpropagation stage, the error signal from the error function is propagated back through the network from the output layer and BPNs make adjustments to the synaptic weights. However, BPNs employ the gradient descent algorithm to optimize the weights in a way that the sum of square error is minimized along the steepest slope of the error surface. Therefore, the result from training data may be massively multimodal and then encounter the danger of local optimal. To compensate for the possible problems of local optimal and error convergence, we can adjust the learning rate, the momentum factor, and other parameters [2].

We elaborate the selection of features for the prediction of financial distresses in Section 2, explain the design of our experiments in Section 3, discuss the experimental results in Section 4, and conclude this paper in Section 5.

2. The Feature Spaces

Altman [1] is the first to apply financial ratios in discriminant analysis to predict financial distresses. Therefore, like many researches, we take Altman's results as the baseline. We employ financial ratios that Altman chose and compare the resulting performance with other feature spaces that we proposed. (For convenience, we refer to a set of feature as a feature space.)

In addition to using features proposed by Altman, we propose five new feature spaces. Our research adopts three variables to represent the *corporate governance indexes* (CGI) for verifying the positive relation between weak corporate governance and financial distresses [10]. We extract 37 variables from Schilit's research [6] to represent *common-size analysis indexes* (CSAI) for assessing financial positions of sampled companies.

Furthermore, we integrate CGI with CSAI into the *combined indexes* (CI) and *cross time combined indexes* (CTCI). We utilize the CI and CTCI to test whether financially distressed companies simultaneously have both unreasonable financial structure and weak corporate governance. CI and CTCI differ in how

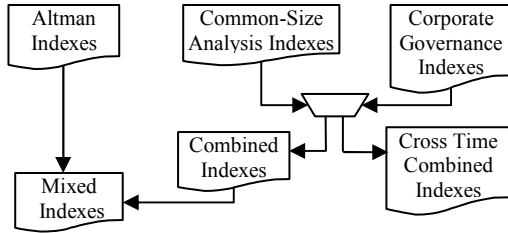


Figure 1. The relationship between six feature spaces

we integrate CGI with CSAI. CTCI combine CGI for a year, denoted y , with CSAI for the following year, i.e., year $(y+1)$. At last, we integrate the traditional *Altman indexes* (AI) with *combined indexes* into *mixed indexes* (MI).

Figure 1 depicts the relationship between the six feature spaces used in our experiments. We list and describe these feature spaces in more detail later. Notice that we use ‘/’ as the sign for arithmetic division.

✧ **Altman Indexes (AI)**

- A. working capital / total assets
- B. retained earnings / total assets
- C. earnings before interest and tax / total assets
- D. market value equity / total liabilities
- E. sales revenue / total assets

✧ **Corporate Governance Indexes (CGI)**

- A. The percentage of the directors’ and supervisors’ shareholdings that have been pledged for bank loans
- B. The percentage of a company’s outstanding shares owned by large shareholders (who own 10 % or more)
- C. The actual percentage of a company’s outstanding shares owned by the directors and supervisors

✧ **Common-Size Analysis Indexes (CSAI)**

We list the factors included in CSAI in Table 2. Notice that, when using common-size analysis, we first convert selected items in financial statements to percentage of size-related measures and then compare the growth rate of converted items. Hence, the contents of Table 2 contain selected items and growth rate of selected items.

✧ **Combined Indexes (CI)**

We integrate CGI with CSAI into combined indexes. However, to reduce the interference of macroeconomics environment, we add three more factors: the growth rate of export, the growth rate of export orders, and the growth rate of Information Electronic Industry’s export orders. With the above three factors, the classifier might slightly adjust the weight when the positions of the healthy companies are bad due to global depression.

✧ **Cross Time Combined Indexes (CTCI)**

Cross Time Combined Indexes integrate CGI with CSAI in the following year. For example, we inte-

Table 2. The components of CSAI

1. The growth rate of inventories / The growth rate of COGS	2. The growth rate of other assets and fixed assets / The growth rate of net sales
3. The growth rate of operating expenses / The growth rate of net sales	4. The growth rate of gross profit / The growth rate of net sales
5. The growth rate of COGS / The growth rate of net sales	6. The growth rate of account payables / The growth rate of net sales
7. The growth rate of inventories / The growth rate of net sales	8. The growth rate of AR / The growth rate of net sales
9. The growth rate of cash and cash equivalents / The growth rate of net sales	10. The growth rate of other assets and fixed assets
11. The growth rate of operating expenses	12. The growth rate of gross profit
13. The growth rate of COGS	14. The growth rate of net sales
15. The growth rate of account payables	16. The growth rate of inventories
17. The growth rate of account receivables	18. The growth rate of cash and cash equivalents
19. Cash flow from financing activity / changes in cash flow	20. Cash flow from investing activity / changes in cash flow
21. Cash flow from operating activity / changes in cash flow	22. Total non-operating expenses / pre-tax income
23. Total non-operating income / pre-tax income	24. Operating income / net sales
25. Operating expenses / net sales	26. Earned (Unearned)-related sales / net sales
27. Gross profit / net sales	28. Operating expenses / net sales
29. Valuation loss of long-term investment / total equity	30. Long-term liabilities / total liabilities
31. Current Of Long-term liabilities / total liabilities	32. Fixed assets / total assets
33. Long-term investment / total assets	34. Inventories / total assets
35. AR & NR-related party / total assets	36. AR / total assets
37. Cash and cash equivalents / total assets	

※COGS: cost of goods sold

※AR: account receivables

※NR: note receivables

grate CGI of T2 (two years before financial distresses takes place) with CSAI of T1 (one year before financial distresses takes place).

✧ **Mixed Indexes (MI)**

We combine Altman indexes and combined indexes.

3. Design of the Experiments

3.1. Data Source

Based on the trading statistics issued by the Taiwan Stock Exchange Corporation, we observed that the Information Electronic Industrial (IEI) group was the biggest trading group which covered 75% of total trading value on the record of September 2005. There were only 48 distressed companies of IEI group with sufficient financial data in the Taiwan Economic Journal (TEJ) database [16]. The definition of financial distresses referred here was defined by TEJ, such as negative net worth, bankruptcy, etc. We referred to this set of financially distressed companies as D1. Our research matched a distressed company with a healthy company in the same industry that was defined by TEJ in an attempt to mitigate the effects of unique industry charac-

teristics and concurrent economic conditions. Notice that, we acquired data for two sets of healthy companies: H1 and H2. H2 was comprised of matching companies that were arbitrarily chosen from all companies in the IEI group, and H1 was comprised of companies that had *positive pre-tax income in 2004*. We considered H1 companies for verifying whether or not carefully matched companies could improve the prediction accuracy of our classifier. Hence, we sampled 144 companies ($=3 \times 48$) from the listed stocks, over the counter stocks, and stocks of some emerging companies in the IEI group.

The declared timing of financially distressed companies was different from each other, so we collected data which was one to three years before the year when the financial distresses took place. Modern Taiwanese laws require public companies to issue financial reports for each quarter every year. We treated the report for each quarter as an individual sample. For instance, if a company encountered financial distresses in August 2004, then we would collect the 12 reports for June 2003, March 2003, December 2002, September 2002, June 2002, and so on. Data for the healthy companies were chosen from 2002 to 2004 (twelve quarters). However, our data collection of financially distressed companies was limited by the law of early days when companies reported annually not seasonally. Hence, we may not collect all the information that we wish to acquire, especially non-financial variables. This problem made our matched-pair companies less than we actually collected. We collected 1728 instances (144 companies \times 3 years \times 4 quarters), and only 1620 instances were used for experiments because of law limitation that mentioned above.

Figure 2 shows how we used data in the experiments for predicting financial distresses. There were two groups of experiments. In the first group, we used data that matched D1 and H1 companies, and, in the second group, we used data that matched D1 and H2. These two groups are shown by solid and dashed lines in Figure 2. In each experiment of these two groups, we further categorized the experiments according to different

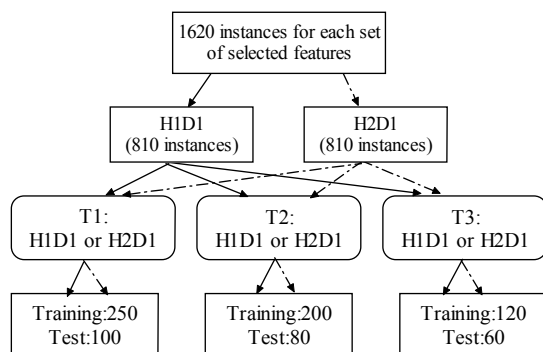


Figure 2. The design of sample data

time stamps for data collected x year(s) before financial distresses took place in the experiments. For instance, $T2$ signifies that we used data that were collected two years before the financial distresses occurred. With this design, we can compare the effects of data collected with different standards and different temporal distance. For each experiment of $T1$ or $T2$, the ratio between amounts of test data and training data was about 2:5. The test data of last year ($T3$) was about half of the training instances. The numbers of the training data and test data in each experiment are depicted in the Figure 2.

3.2. Classifiers

We employed backpropagation neural networks (BPNs) and support vector machines (SVMs) in our classifiers. The functions for manipulating BPNs were implemented in MATLAB [17]. We employed the *trainrp* function for training the BPNs that used the *tansig* transfer function. The number of neurons in the input was decided by the number of variables of individual feature space, and the number of neurons in the hidden layer was half of the sum of input-layer neurons and output-layer neurons. There was only one neuron at the output layer. We classified the test instance as distressed when the output of the output-layer neuron was larger than 0.5. Otherwise, the test instance would be labeled as healthy.

SVMs have provided a relatively new approach to the task of classification [14], [15], so were utilized in our second classifier. We trained SVMs with known labels for pattern recognition, and predicted the classes of test data. We employed the LIBSVM packages provided by Chang and Lin [18], and chose the C-SVC SVMs and the RBF kernel function. In addition, we tried different combinations of *gamma* and *cost*, which were varied from 0 to 2, to search the best resulting F measure.

3.3. Measurement of Prediction Quality

Many researchers who applied artificial neural networks for financial predictions regarded precision as the only criterion of measuring classifying ability [3]–[5], [7]. The definition of precision is the *proportion of companies that the classifier predicts to be distressed are actually distressed*. The denominator of precision is the companies that the classifier classified as distressed, not the total number of financially distressed companies. Hence, using precision as the quality indicator underestimates the impact of false dismissal (or Type I error) when the classifier regards financially distressed companies as healthy companies. Because the financially distressed companies with management fraud tend to have normal financial positions, classifiers may not be able to classify such companies correctly. The problem of false dismissal may become worse when we use data

that were collected two or three years before financial distresses actually took place. Unfortunately, the failure to warm case like Enron will not affect the classifier’s performance measure if we continue to use precision as the measure. For measuring the quality of prediction in a fair manner, we propose the F measure, which simultaneously considers recall and precision into the formula. The F measure, precision and recall are defined as follows [12].

P = the proportion of selected financially distressed companies the classifier correctly predicted

R = the proportion of financially distressed companies the classifier correctly predicted

$$F = \frac{1}{\alpha \frac{1}{P} + (1-\alpha) \frac{1}{R}} \quad (1)$$

We set α to 0.5 for equal weighting of P and R. With this α , the F measure simplifies to $2PR/(P+R)$.

4. Experimental Results and Discussion

As we explained in Section 3.1, the experiments were designed to evaluate the influence on prediction accuracy of different factors. We would like to compare the effects of using the six sets of features, the effects of using T_i data in training the classifiers, and the effect of using H1 and H2 with D1. Hence there should be 36 (= 6 × 3 × 2) experiments for each classifier. However, we could not conduct two experiments which we could not collect required data due to old laws, which we explained in Section 3.1.

The legends in figure3 indicate the setup of the experiments. The first part of each legend symbol shows the name of the feature space, and the second part whether we used H1 or H2 with D1. The vertical axis shows the F measure achieved by the classifier when it predicted the financial status of the test data. The horizontal axis shows the time stamps of the data that we used in training the classifiers and predicting the financial statuses of the companies.

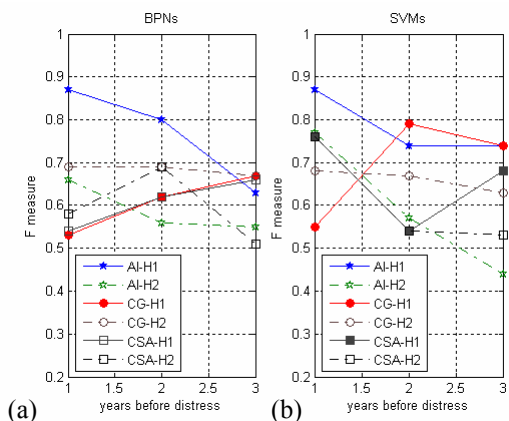


Figure 3. Effects of the three basic feature spaces under different classifiers

Charts in Figures 3 (a) and 3 (b) show the effects of using the three basic feature spaces (AI, CGI, and CSAI) when we used BPNs and SVMs, respectively. Generally speaking, AI performed better than other basic feature spaces when we used T₁ data. As we used more temporally distant data in the experiments, we got lower F measure when we used AI. In contrast, the effects of CGI improved or remained similar from T₁ to T₃ of the charts. The trends for CSAI were not stable, unfortunately. As a consequence, we did not observe a decisive relationship between the performances of these basic feature spaces.

Applying the basic feature spaces with H1 and H2 showed different trends. Using H1 with AI (AI-H1) led to much better performance than using H2 with AI (AI-H2). Hence, carefully choosing healthy companies may help us build a better classifier when we employed AI. In contrast, using H1 with CGI (CGI-H1) led to worse performance than using H2 with CGI (CGI-H2). Again, we did not observe stable trends when we used CSAI.

Since AI appeared to achieve the best performance in the three basic feature spaces, we continued to compare the performances of CI, CTCI, and MI with AI. Charts in Figure 4 (a) and 4 (b) show the experimental results when we used BPNs and SVMs, respectively. In both charts of Figures 4, CTCI-H2 achieved better performance than AI-H2. This phenomenon agrees with the results reported by Lee et al. [10] which suggest that corporate governance is positively related to the risk for financial distresses in the following year. It also indicates that integrating CGI with CSAI in the following year helped us collect data for healthy companies relatively more easily and reach higher prediction of financial distresses.

In addition, MI-H1 in Figure 4(b) provided the best performance in the three complex feature spaces and the performance was also better than AI-H1. The result indicates that combining corporate governance and common-size analysis features with Altman features helped us achieve better prediction accuracy when we used SVMs. The experimental results indicate that MI is a good feature space for predicting financial dis-

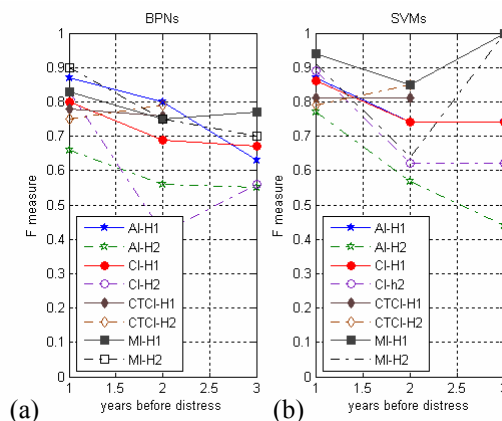


Figure 4. Effects of the three complex feature spaces with AI under different classifiers

tresses.

We show but do not discuss the experimental results for T3 because we are still investigating the trustworthiness of the data for T3, particularly the results at T3 when we used SVMs. The classifiers achieved relatively high prediction quality at T3. There are two possible explanations: the interference of small samples or SVMs is really good for dealing with multiple dimensions problems. The total number of our experiments for each classifier is 34. Comparing all the charts, SVMs helped us achieve better performance in 76% of the experiments (26/34) than BPNs. This phenomenon corresponds to the observation of Shin et al. [13] that SVMs are possibly capable of extracting the optimal solution with smaller data set. Our research has the same observation.

5. Conclusion

With the help of corporate governance and common-size analysis, our classifiers were able to perform well in two or three years before companies encountering financial distresses. The above two methodologies helped us to achieve better accuracy than traditional Altman indexes when we use SVMs. In addition, SVMs provided better performance in 79% of the experiments than BPNs. This phenomenon corresponds to the observation of Shin et al. [13] that SVMs are possibly capable of extracting the optimal solution with smaller data set. Our experience also suggests that we should prefer SVMs as classifier than BPNs when the sample data is smaller. In an extension of this research, we have applied genetic algorithms for selecting features from components in MI to achieve better prediction quality, and the results were reported in [19].

Acknowledgements

This research was supported in part by contract NSC-94-2213-E-004-008 of the National Science Council in Taiwan.

References

- [1] E. I. Altman. Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. *Journal of Finance*, Vol. 23, pp. 589-609, 1968.
- [2] J. R. Coakley and C. E. Brown. Artificial neural networks in accounting and finance: modeling issues. *International Journal of Intelligent Systems in Accounting, Finance & Management*, Vol. 9, pp. 119-144, 2000.
- [3] K. Fanning, K. Cogger, and R. Srivastava. Detection of management fraud: a neural network approach. *International Journal of Intelligent Systems in Accounting, Finance & Management*, Vol. 4, pp. 113-26, 1995.
- [4] M. L. Nasir, R. I. John, S. C. Bennett, D. M. Russell, and A. Patel. Predicting corporate bankruptcy using artificial neural networks. *Journal of Applied Accounting Research*, Vol. 5, pp. 30-52, 2000.
- [5] M. D. Odom and R. Sharda. A neural network model for bankruptcy prediction. *International Joint Conference on Neural Networks*. Vol. 2, pp. 163-168, 1990.
- [6] H. Schilit. *Financial shenanigans: how to detect accounting gimmicks & fraud in financial reports*. New York: McGraw-Hill, 2002.
- [7] K. Fanning, K. Cogger, and R. Srivastava. Detection of management fraud: a neural network approach. *International Journal of Intelligent Systems in Accounting, Finance & Management*, Vol. 4, pp. 113-26, 1995.
- [8] S. Johnson, P. Boone, A. Breach, and E. Friedman. Corporate governance in the Asian financial crisis. *Journal of Financial Economics*, Vol. 58, pp. 141-186, 2000.
- [9] R. La Porta, F. Lopez-de-Silanes, A. Shleifer, and R. Vishny. Agency problems and dividend policies around the world. *Journal of Finance*, Vol. 55, pp. 1-33, 2000.
- [10] T. S. Lee and Y. H. Yeh. Corporate governance and financial distress: evidence from Taiwan. *Corporate Governance*, Vol. 12, pp. 378-388, 2004.
- [11] Y. H. Yeh, T. S. Lee and T. Woidtke. Family control and corporate governance: evidence for Taiwan. *International Review of Finance*, Vol. 2, pp. 21-48, 2001.
- [12] Van Rijsbergen. *Information Retrieval*, 2nd edition, Butterworth-Heinemann Newton, 1979.
- [13] K. S. Shin, T. S. Lee, and H. J. Kim. An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications*, Vol. 28, pp. 127-135, 2005.
- [14] A. Fan and M. Palaniswami. Selecting bankruptcy predictors using a support vector machine approach. *Proceedings of the International Joint Conference on Neural Networks*, Vol. 6, pp. 354-359, 2000.
- [15] S. Salcedo-Sanz, M. Deprado-Cumplido, and M. J. Segovia. Feature selection methods involving support vector machines for prediction of insolvency in non-life insurance companies. *International Journal of Intelligent Systems in Accounting, Finance & Management*, Vol. 12, pp. 261-281, 2004.
- [16] TEJ: <http://www.tej.com.tw>
- [17] MATLAB: <http://www.mathwork.com>
- [18] C. C. Chang and C. J. Lin, LIBSVM: A library for support vector machines, 2001. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [19] P.-W. Huang and C.-L. Liu. Using genetic algorithms for feature selection in predicting financial distresses with support vector machines. *Proceedings of the 2006 IEEE International Conference on Systems, Man, and Cybernetics*, to appear, 2006.