Predicting Bankruptcy After The Sarbanes-Oxley Act Using The Most Current Data Mining Approaches

Wikil Kwak, Ph.D., University of Nebraska at Omaha, USA
Yong Shi, University of the Chinese Academy of Science, China and University of Nebraska at Omaha, USA
Gang Kou, University of Electronic Science and Technology of China, China

ABSTRACT

Our study proposes several current data mining methods to predict bankruptcy after the Sarbanes-Oxley Act (2002) using 2007-2008 U.S. data. The Sarbanes-Oxley Act (SOX) of 2002 was introduced to improve the quality of financial reporting and minimize corporate fraud in the U.S. Because of this SOX implementation, a company's financial statements are assumed to provide higher quality financial information for investors and other stakeholders. The results of our data mining approaches in our bankruptcy prediction study show that Bayesian Net method performs the best (85% overall prediction rate with 94% in AUC), followed by J48 (85% with 82% AUC), Decision Table (83.52%), and Decision Tree (82%) methods using financial and other data from the 10-K report and Compustat. These results are better than previous bankruptcy prediction studies before the SOX implementation using most current data mining approaches.

Keywords: Sarbanes-Oxley; Data Mining; Bankruptcy; Decision Tree; Bayesian Net; Decision Table

INTRODUCTION

$\mathcal{B}$ecause of the major scandals at Enron and WorldCom in early 2000, the Sarbanes-Oxley Act (SOX) of 2002 was introduced to improve the quality of financial reporting and minimize corporate fraud in the U.S. Two sections of SOX in particular, Section 302 and Section 404, focus on required reporting on the effectiveness of and any material weaknesses in internal controls over financial reporting by companies' managements and certifications by specific corporate officers. SOX Section 404 requires management to report to shareholders acknowledging its responsibility to detect and prevent material weaknesses in internal controls over financial reporting and indicating the effectiveness of these internal controls. In addition, auditors are required to report on client assertions regarding the effectiveness of these internal controls and the existence or absence of material weaknesses in their internal control systems. After the SOX implementation, the quality of financial statements was supposed to improve because compliance costs of SOX are not trivial. Therefore, if we apply the bankruptcy prediction model using financial data with the internal control factor, the overall prediction rates will be better after SOX is implemented fully.

Data mining is becoming popular in business applications and its techniques are getting more powerful in terms of overall prediction rates. We apply the most current data mining tools for our study and try to determine which method is the best for bankruptcy prediction.

Our paper is organized as follows: The next section presents the background and prior research relevant to our study. The third section presents current data mining models we used. The fourth section describes our sample data and reports our empirical results using other data mining tools. The last section concludes with a summary of our findings and future research avenues.

BACKGROUND AND PRIOR RESEARCH

Section 404 of SOX includes two major objectives focusing on internal controls over financial reporting. The first one is the inclusion in a firm's annual report by management stating its responsibility for internal controls over finance reporting and its assessment of the effectiveness of these internal controls (Kwak *et al.*, 2009). The second one is requiring the firm's independent auditors to attest to management's internal control assessment as part of the annual audit engagement. These changes in SEC rules and regulations in accordance with SOX Section 404 were effective after November 15, 2004 for accelerated filers. Since we use data for three years before filing bankruptcy, 2008 is the first year of fully implementing the SOX regulation.

Even with these SOX provisions in place, recent significant events such as the 2008 capital market crash in the U.S. and European debt woes negatively impacted investor confidence. These continuous issues of financial troubles, deteriorating job markets, political turmoil globally make U.S. capital market and investor confidence uneasy. So, our study attempts to understand the effect of SOX by assuming better internal controls over financial reporting using the bankruptcy prediction study framework.

SOX Section 302 increases disclosure requirements through officer certifications regarding the reporting by officers on the effectiveness of a firm's internal control system and on significant changes of its internal control system (Ge and McVay 2005). The SEC requires an issuer's principal executive officers and principal financial officers to certify each quarterly and annual report, including transition reports, filed or submitted by the issuer (Kwak et al., 2009). The certifying officers have to acknowledge whether or not there were significant changes in internal controls or in other factors that could significantly affect internal controls subsequent to the date of their evaluation, including any corrective actions with regard to significant deficiencies and material weaknesses. We apply the current data mining approaches in our prediction study using Section 302 disclosures to propose an efficient and effective assessment of internal control weaknesses of a firm after SOX.

Altman (1968) used multiple discriminant analysis to predict bankruptcy. Another major study is Ohlson's (1980) study which used a logit model. His model does not require any assumptions about the prior probability of bankruptcy or the distribution of predictor variables.

Gupta et al. (1990) proposed linear goal programming, but his approach may not be practical because of computational complexity. The decision tree approach to develop bankruptcy prediction models was proposed by Sung et al. (1999). Their results show that prediction ratios are not stable in different economic conditions. Therefore, economic conditions are important factors in bankruptcy prediction studies. The accuracy of their model is 72.4% for bankruptcy under normal conditions while it was 66.7% under crisis conditions using 1991 to 1998 Korean data. A genetic algorithm (GA) in bankruptcy prediction modeling was used by Shin and Lee (2002). Their accuracy rate is 80.8% in both training and hold-out samples. However, this study used only one industry and there may be an upward bias for prediction accuracy. A case-based reasoning (CBR) with weights derived by an analytic hierarchy process (AHP) for bankruptcy prediction since AHP incorporates both financial ratios as well as non-financial variables into the model was proposed by Park and Han (2002). They reported an 84.52% accuracy rate even after incorporating non-financial data. Chi and Tang (2006) recently used 24 variables to apply logit analysis and the prediction rate was 85%. From previous bankruptcy studies, any study more than the overall prediction rate of 85% is a reasonable prediction study.

The most prominent bankruptcy studies are Altman's and Ohlson's and our study uses their financial variables. In addition, internal control weakness, missing dividend payout, and stock market return rate are added to improve the prediction rate. These variables are pulled from each bankrupt firm's financial statements three years before they file bankruptcy. Recently, there are other intelligent techniques such as neural network or Support Vector Machine (SVM) used in bankruptcy studies (Elish and Elish 2008). Baek and Cho (2003) used a neural network approach for Korean firms to predict bankruptcy. Their approach showed 80.45% for solvent and 50.6% for default firms in terms of prediction accuracy. Min and Lee (2005) proposed SVM for the bankruptcy prediction study and found their holdout data classification rate was 83.06%.

Artificial neural networks and a fuzzy set approach have both been used in management fraud and internal controls research. Fanning *et al*. (1995) used artificial neural networks to detect management fraud and developed the discriminant function that provides a set of questions to detect management fraud. Deshmukh *et al*. (1997) proposed a fuzzy set approach that provides guidelines to measure and combine red flags, and it can be used to build fuzzy reasoning systems that assess the risk of management fraud. Korvin *et al*. (2004) recently proposed a pragmatic approach to assess risks of internal control using the fuzzy set theory. Their approach is useful for the computer-based accounting information system's internal control. However, their approach could be expensive and time consuming.

Hammersley *et al*. (2008) find results using the stock price reaction to the SOX Section 302 disclosures. They find that the information content of internal control weakness disclosed in Section 302 disclosures depends on the factors of internal control weaknesses, such as their severity, auditability, and the vagueness of disclosures. This study's results indicate that investors react to information about the existence of internal control weaknesses, so the results of our study may benefit investors by providing better predictions on which firms will enter bankruptcy in the near future.

We use 13 of the most current data mining tools including the decision tree approach in this study for comparison.

SAMPLE DATA, VARIABLES, AND EMPIRICAL RESULTS

In this paper, we used DirectEDGAR (2008) to identify 130 firms that filed bankruptcy in 2008 and 2009. These are years that SOX has fully implemented each firm's financial statements. We then identified more than twice the number of control firms that had no bankruptcy filing. Our control firms were matched with our bankrupt firms using size and two-digit industry codes. Our final sample is composed of 306 firm-year observations that have available financial and other data in Form 10-K filings using DirectEDGAR and Compustat.

In selecting variables that may help predict bankruptcy over financial reporting, we included Altman's (1968) and Ohlson's (1980) variables because these variables have proven to be useful for bankruptcy prediction in previous studies, as discussed in section two of this paper. In addition, we include internal control weakness, dividend payout, and stock return variables as proposed by previous bankruptcy studies (e.g., Duffie *et al*., 2007, Sun, 2007, and Shumway, 2001).

Table 1 shows the descriptive statistics. As you can see in Table 1, all variables between bankrupt firms and non-bankrupt firms are quite different except for Size, Funds flow from operations/Total liabilities, CHIN (change in income), Market value/Total debt, Sales/Total assets, and Return variables. From our results, bankrupt firms have higher Debt to Total assets, Current liability to Current asset ratios are higher, more Losses, higher Total liabilities/Total assets ratios greater than 1, higher Internal control weakness, missing Dividend more frequently. In contrast, bankrupt firms have lower Working capital to Total assets, more negative Net income to Total assets, more negative Earnings before interest and taxes to Total assets, and more negative Retained earnings to Total assets ratios.

Size is similar as we matched the control firms based on size and industry. Stock market return data are missing a lot for near bankrupt firms for our sample firms and this variable is not significant because of missing values.

Table 1: Part 1: Descriptive data

Bankrupt =1

Variable	N	Mean	Std Dev	Minimum	Maximum	t-diff_stat
SIZE	130	3.036415	0.670017	1.066699	4.251368	1.87984
TDEBT_TA	130	0.522197	0.372747	4.8E-05	2.365381	6.539735***
WCA_TA	130	0.003749	0.365557	-2.35556	0.573819	-5.29695***
CL_CA	130	1.325495	1.917049	0.139244	13.66695	3.459551***
NI_TA	130	-0.12756	0.238387	-1.36716	0.544211	-4.98748***
FU_TL	130	33.56841	347.9683	-52.72	3949.74	-0.69753
LOSS	130	0.815385	0.389486	0	1	11.59204***
OENEG	130	0.084615	0.279385	0	1	2.97523***
CHIN	130	-3.04477	27.96642	-198.828	110.2286	-1.24302
EBIT_TA	130	-0.01393	0.130669	-0.92413	0.171389	-5.49235***
MKV_TD	130	42.64569	410.2507	0.000499	4648.89	-1.34875
SALES_TA	130	1.234936	0.857946	0.016094	4.438618	1.823911
RE_TA	130	-0.59322	1.167781	-7.32708	0.469771	-5.04411***
IC	130	0.207692	0.407225	0	1	4.712419***
DIV	128	0.757813	0.430091	0	1	3.380889***
RETX	28	-0.48858	0.938442	-2.37081	1.085964	-1.83609

Non-Bankrupt =0

Variable	N	Mean	Std Dev	Minimum	Maximum
SIZE	306	2.898546	0.767627	-0.2204	5.30122
TDEBT_TA	306	0.291959	0.228544	1.31E-05	1.080334
WCA_TA	306	0.185622	0.21494	-0.66999	0.824403
CL_CA	306	0.728434	0.680917	0.093013	9.784977
NI_TA	306	-0.00602	0.218965	-2.18817	0.588664
FU_TL	306	83.02064	1119.38	-78.3581	19433.31
LOSS	306	0.313726	0.464766	0	1
OENEG	306	0.009804	0.09869	0	1
CHIN	304	0.004126	0.068398	-0.24818	0.365246
EBIT_TA	306	0.071082	0.181967	-2.15954	0.888582
MKV_TD	306	386.9547	4421	0.001869	75722.15
SALES_TA	306	1.081045	0.667693	0.080171	3.137091
RE_TA	306	-0.01265	0.91857	-11.4386	1.557583
IC	306	0.03268	0.178088	0	1
DIV	306	0.598039	0.491097	0	1
RETX	212	-0.15298	0.643907	-4.07103	1.608485

*: p< 0.10
**: p < 0.05
***: p < 0.001

Variable Descriptions:
Size =Log (Total Assets/Gross Domestic Products);
TDEBT_TA=Total Liabilities/Total Assets;
WCA_TA=Working Capital Divided by Total Assets;
CL_CA=Total Current Liabilities/Total Current Assets;
NI_TA=Net Income/Total Assets;
FU_TL=Funds from Operations/Total Liabilities;
LOSS= if Net Income<**0** or lag (Net Income) <**0** then INTWO=**1**; else INTWO=**0**;
OENEG=if TL/TA>**1** then OENEG=**1**; else OENEG=**0;**
CHIN= (Net Income- lag (Net Income))/ [absolute (Net Income) + absolute (lag Net Income)];
EBI_TA = Earnings Before Interest and Taxes/Total Assets;
MKV_TD = Market Value of Equity/Book Value of Total Debt;
SALES_TA = Sales/Total Assets;
RE_TA = Retained Earnings/Total Assets;
IC = if Internal control weakness is disclosed then 1, 0 otherwise;
DIV = if dividend is missing then 1, 0 otherwise; and
RETX = annual market returns of the firm.

Table 2 presents our results using thirteen prediction models using both years. Panel A shows results with only 13 financial ratios and Panel B shows results with 13 financial ratios and other variables. Both results are similar and we will focus on using all variables.

Table 2: Panel A: Classification results with 13 financial ratios

Algorithm	AUC	Overall Accuracy	F-measure	Mean Absolute Error
decision table	0.88	85.11	0.78	0.22
simple CART	0.85	83.84	0.77	0.21
Jrip	0.83	83.21	0.77	0.21
logistic	0.87	81.33	0.72	0.2
J48	0.81	85.1	0.77	0.16
bayesnet	0.93	84.45	0.78	0.16
naivebayes	0.74	42.06	0.54	0.58
multilayer perceptron	0.86	81.33	0.73	0.21
nbtree	0.89	83.53	0.76	0.19
SMO	0.74	76.84	0.65	0.23
RBFNetwork	0.61	66.45	0.13	0.42
libSVM	0.66	75.6	0.48	0.24
ibk	0.76	78.75	0.68	0.21

Kappa	TPR	FPR	TNR	FNR	Precision	Recall	Time Training	Time Testing
0.67	0.79	0.12	0.88	0.21	0.78	0.79	0.19	0
0.65	0.79	0.13	0.87	0.21	0.76	0.79	0.19	0
0.64	0.81	0.16	0.84	0.19	0.73	0.81	0.07	0
0.58	0.72	0.14	0.86	0.28	0.74	0.72	0.28	0
0.66	0.74	0.09	0.91	0.26	0.81	0.74	0.04	0
0.66	0.8	0.14	0.86	0.2	0.77	0.8	0.01	0
0.09	0.96	0.86	0.14	0.04	0.39	0.96	0.01	0
0.59	0.74	0.15	0.85	0.26	0.73	0.74	9.36	0
0.63	0.78	0.13	0.87	0.22	0.75	0.78	4.45	0
0.48	0.65	0.17	0.83	0.35	0.67	0.65	0.13	0
0.07	0.14	0.06	0.94	0.86	0.15	0.14	0.07	0
0.36	0.34	0.03	0.97	0.66	0.88	0.34	0.15	0.01
0.52	0.68	0.16	0.84	0.32	0.69	0.68	0	0.01

Table 2: Panel B: Classification results with all variables

Algorithm	AUC	Overall accuracy	F-measure	Mean Absolute Error
decision table	0.89	83.52	0.75	0.23
simple CART	0.82	83.22	0.75	0.22
Jrip	0.83	83.51	0.77	0.2
logistic	0.86	80.69	0.72	0.21
J48	0.82	85.42	0.78	0.16
bayesnet	0.94	85.42	0.8	0.15
naivebayes	0.74	42.68	0.55	0.57
multilayer perceptron	0.87	77.56	0.68	0.23
nbtree	0.88	81.94	0.73	0.19
SMO	0.73	75.58	0.63	0.24
RBFNetwork	0.63	66.13	0.17	0.42
libSVM	0.65	74.97	0.47	0.25
ibk	0.77	78.43	0.68	0.22

Kappa	TPR	FPR	TN	FN	Precision	Recall	Time Train Time	Test
0.63	0.75	0.12	0.88	0.25	0.77	0.75	0.23	0
0.63	0.76	0.13	0.87	0.24	0.76	0.76	0.21	0
0.64	0.8	0.14	0.86	0.2	0.76	0.8	0.09	0
0.57	0.71	0.14	0.86	0.29	0.73	0.71	0.28	0
0.67	0.78	0.11	0.89	0.22	0.8	0.78	0.04	0
0.68	0.83	0.13	0.87	0.17	0.78	0.83	0.02	0
0.1	0.96	0.85	0.15	0.04	0.4	0.96	0.01	0
0.51	0.7	0.19	0.81	0.3	0.67	0.7	10.07	0
0.6	0.76	0.15	0.85	0.24	0.71	0.76	5.37	0
0.45	0.64	0.18	0.82	0.36	0.65	0.64	0.08	0
0.08	0.15	0.08	0.92	0.85	0.23	0.15	0.08	0
0.35	0.33	0.03	0.97	0.67	0.87	0.33	0.11	0.01
0.52	0.7	0.17	0.83	0.3	0.67	0.7	0	0.01

Variable Descriptions:

AUC: Area under Receiver Operating Characteristic;
Overall Accuracy: Percentage corrected classified as bankrupt firms;
F-measure: The harmonic mean of precision and recall;
Mean Absolute Error: The predictions deviate from the true probability;
Kappa: Measure that estimates the similarity between the members of ensemble in
       multi-classifier systems;
TPR: True positive rate in %;
FPR: False positive rate in %;
TN: True negative rate in %;
FN: False negative rate in %;
Precision: The number of classified fault-prone modules that actually are fault-prone;
Recall: Fault-prone modules that are correctly classified in %;
Time Training: Time training in seconds; and
Time Testing: Time testing in seconds.

Algorithm description:

Decision Table: Decision Table Rules;
Simple CART: Simple Classification and Regression Trees;
Jrip: Repeated Incremental Pruning Rules;
Logistic: Linear Logistic Regression Functions;
J48: C4.5 Decision Trees;
Bayesnet: Bayesian network;
Naivebayes: Naïve Bayes;
Multilayer perception: Neural Network Functions;
Nbtree: Naïve Bayes Trees;
SMO: Sequential Minimal Optimization Functions;
RBFNetwork: Radial Basis Functions Network;
LibSVM: Support Vector Machine Functions; and
Ibk: K-nearest-neighbor Lazy.

As a means of validation, the classification results of thirteen data mining models are compared with eleven performance metric (Peng *et al*. 2009)

There are five categories of classification models and these are trees, functions, Bayesian classifiers, lazy classifiers, and rules. We use Regression Tree (CART), Naïve Bayes tree, and C4.5 for tree category. The functions category uses Linear Logistic Analysis, Radial Basis Function (RBF) network, Sequential Optimization (SMO), Support Vector Machine (SVM), and Neural Networks. The Bayesian category includes Bayesian network and Naïve Bayes. The Lazy classifier uses K-nearest-neighbor. We use Decision Table and Repeated Incremental Pruning to Produce Error Reduction (RIPPER) for rules category (Peng *et al*., 2009). Each of these classifiers has pros and cons and we want to see which methods perform better in our bankruptcy context.

We use eleven performance criteria to evaluate the classification performance: Overall accuracy, True positive, False positive, True negative, False negative, Precision, Recall, F-measure, AUC, Kappa Statistics (KapS), and Mean absolute error (MAE) score. Accuracy is one the most widely used classification performance metrics. It equals the ratio of correct predicted records to the total records in the dataset/class. Overall Accuracy = $\dfrac{TN+TP}{TP+FP+FN+TN}$, True positive rate/Sensitivity $=\dfrac{TP}{TP+FN}$, False positive rate $=\dfrac{FP}{FP+TN}$, True negative rate/Specificity $=\dfrac{TN}{FP+TN}$, False negative rate $=\dfrac{FN}{FN+TP}$, where TP FP, TN and FN are defined below.

TP (True Positive): The number of bankrupt firms correctly classified.
FP (False Positive): The number of records non-bankrupt firms incorrectly classified as bankrupt firms.
TN (True Negative): The number of non-bankrupt firms correctly classified.
FN (False Negative): The number of bankrupt firms incorrectly classified as non-bankrupt firms.

Type I and Type II errors also are ratios using TP, FP, TN, and FN as defined above. A Type I error equals $\dfrac{FN}{FN+TN}$, and shows the percentage of Type II error $=\dfrac{FP}{FP+TP}$. There are two types of errors (Type-I and Type-II) involved in prediction studies. Type-I error refers to false rejection error. For example, in a bankruptcy prediction study, we reject the null hypothesis that a firm is a non-bankrupt firm even though the firm is actually a bankrupt firm. This type of error will be very costly for a decision maker. A Type-II error is a false acceptance error. For example, we predict a firm to be a bankrupt firm, even though the firm is a non-bankrupt firm. In the Type II error case, the cost of misclassification is not as severe as in the Type I error case. For our study, we focus on Type-I error because the cost of misclassifying can be significant. Receiver Operating Characteristic (ROC) measures trade-offs between TP and FP rates and area under ROC (AUC) shows the accuracy of the classifier and the larger the area, the better the classifier. For other metrics refer to Peng *et al*. (2009).

Since training results indicate how well the classification model fits the training set and test results reflect the real predicting power of the classification model, the following tables only summarize the eleven metrics for test sets. The standard 10-fold cross validation process is applied to all classifiers. The standard Leave-One-Out validation process is applied to LDA.

Based on the overall accuracy rate of both years using all variables, Bayesnet shows 85.42% of overall accuracy with 94% of AUC. Next, J48 Decision Tree using C4.5 shows 85.42% of overall accuracy, followed by Decision Table with 83.52% of overall accuracy. Interestingly, the overall accuracy rate of Decision Table is 85.11% with only 13 financial ratios compared with 83.52% with all variables. Other cases, usually overall accuracy rates are higher for classifiers using all variables. Therefore, we focus on results of overall variables for further discussions. For Bayesnet, F-measure is 80%, 15% of Mean Absolute Error, and Kappa Rate is 68%, Precision is 78%, Recall is 83%, Training Time is 0.02 and Testing Time is close to 0. Overall, our top data mining models are very effective and efficient and overall prediction rates are close to 85% using 2008 and 2009 U.S. data. The most previous studies show lower than 85% of bankruptcy prediction rates using various methods and our results imply that bankruptcy prediction rates after SOX is a little better than before SOX.

SUMMARY AND CONCLUSIONS

In this paper we used 13 data mining approaches to predict bankruptcy after SOX implementation using 2008-2009 U.S. data with 13 financial ratios and internal control weakness, dividend payout, and market return variables. The data mining approaches have recently been applied to several bankruptcy prediction studies (Kwak *et al*., 2012). This approach has proven to be robust and powerful even for a large sample size using a huge financial database. The results of the Bayesnet approach in our bankruptcy study are better in overall accuracy compared with previous bankruptcy studies before SOX implementations.

Based on overall accuracy and AUC, Bayesnet and Decision Table are promising classifiers, but based on Type I and Type II error rates, Bayesnet is the best classifier in our bankruptcy context. Type I error rate for

Bayesnet is 16.35% (23.15% for Decision Table) and Type II error rate is 13.54% (13.79% for Decision Table) as we calculated from Table 2, Panel B using all variables. Type I error rate is more costly for decision makers. We may need to add more data and variables to improve our study.

AUTHOR INFORMATION

Wikil Kwak is a Professor of Accounting at the University of Nebraska at Omaha. He received Ph.D. in Accounting from the University of Nebraska in Lincoln. Dr. Kwak's research interests include the areas of mathematical programming approaches in capital budgeting, transfer pricing, performance evaluation and Japanese capital market studies. He has published in the *Engineering Economist*, *Abacus*, *Contemporary Accounting Research*, *Review of Quantitative Finance and Accounting*, *Management Accountant*, *Journal of Petroleum Accounting and Financial Management*, *Business Intelligence and Data Mining, Review of Pacific Basin Financial Markets and Policies*, and *Multinational Business Review*. E-mail: wkwak@mail.unomaha.edu. Corresponding author.

Yong Shi is the director of Chinese Academy of Sciences Research Center on Data Technology & Knowledge Economy, President's Assistant of the Graduate University of Chinese Academy of Sciences. He had been the Charles W. and Margre H. Durham Distinguished Professor of Information Technology, College of Information Science and Technology, University of Nebraska at Omaha since1999. Dr. Shi's research interests covers data mining, information overload, multiple criteria decision making and telecommunication management. He has published seven books and more than sixty papers in various journals He is the Editor-in-Chief of *International Journal of Information Technology and Decision Making* (SCI) and an Area Editor of *International Journal of Operations and Quantitative Management*. Dr. Shi has received many distinguished awards including Outstanding Young Scientist Award, National Natural Science Foundation of China, 2001; Member of Overseas Assessor for the Chinese Academy of Sciences, May 2000; and Speaker of Distinguished Visitors Program (DVP) for 1997-2000, IEEE Computer Society. He has consulted in a number of famous companies.

Gang Kou is a professor of School of Management and Economics, University of Electric Science and Technology of China and managing editor of *International Journal of Information Technology & Decision Making*. Previously, he was a research scientist in Thomson Co., R&D. He received his Ph.D. in Information Technology from the College of Information Science & Technology, Univ. of Nebraska at Omaha. He got his master's degree in Computer science, University of Nebraska at Omaha, B.S. in Physics, Tsinghua University, Beijing, China.

REFERENCES

1. Altman E., Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, *The Journal of Finance* 23 (3), 589-609, 1968.
2. Baek, J. and Cho S. Bankruptcy Prediction for Credit Risk Using an Auto Associative Neural Networks in Korean Firms, IEEE International Conference on Computational Intelligence for Financial Engineering, Hong-Kong, 2003.
3. Chi, L. and T. Tang, Bankruptcy Prediction: Application of Logit Analysis in Export Credit Risks, *Australian Journal of Management* 31 (1), 17-27, 2006.
4. Deshmukh A. J. Romine, and P. Siegel, Measurement and Combination of Red Flags to Assess the Risk of Management Fraud: A Fuzzy Set Approach, *Managerial Finance* 23 (6), 35-48, 1997.
5. DirectEDGAR, Academic EDGAR+ LLC, 2008 North 55th Street, Omaha, NE 68104, 2008.
6. Duffie, D, L. Saita and K. Wang, Multi-period Corporate Default Prediction with Stochastic Covariates, *Journal of Financial Economics* 83, 635-665, 2007.
7. Elish, K. O. and M. O. Elish, Predicting Defect-Prone Software Modules Using Support Vector Machines, *Journal of Systems Software* 81 (5), 649-660, 2008,
8. Fanning, K., K. O. Cogger and R. Srivasava, Detection of Management Fraud: A Neural Network Approach, *Intelligent Systems in Accounting, Finance, and Management* 4, 113-126, 1995.
9. Ge W. and S. McVay, The Disclosure of Material Weaknesses in Internal Control after the Sarbanes-Oxley Act, *Accounting Horizons* 19(3), 137-158, 2005.

10.   Gupta Y. P., R. P. Rao, and P. K. Baggi, Linear Goal Programming as an Alternative to Multivariate Discriminant Analysis: A Note, *Journal of Business, Finance, and Accounting* 17(4), 593-598, 1990.
11.   Hammersley, J. S., L. A. Myers, C. Shakespeare, Market Reaction to the Disclosure of Internal Control Weakness and to the Characteristics of Those Weakness under the Section 302 of the Sarbanes Oxley Act of 2002, *Review of Accounting Studies*, 13(1), 141-165, 2008.
12.   Korvin, A. D., M. Shipley, and K. Omer, Assessing Risks due to Threats to Internal Control in a Computer-based Accounting Information System: A Pragmatic approach Based on Fuzzy Set Theory, *Intelligence Systems in Accounting, Finance and Management* 12, 139-152, 2004.
13.   Kwak, W., S. Eldridge, Y. Shi and G. Kou, Predicting Material Weakness in Internal Control Systems After the Sarbanes-Oxley Act using Multiple Criteria Linear Programming and Other Data Mining Approaches, *The Journal of Applied Business Research*, 25(6), 105-118, 2009.
14.   Kwak, W., Y. Shi, and G. Kou, Bankruptcy Prediction for Korean Firms after the 1997 Financial Crisis: Using a Multiple Criteria Linear Programming Data Mining Approach, forthcoming in *Review of Quantitative Finance and Accounting*, 2012.
15.   Min, J. H. and Lee Y. Bankruptcy Prediction Using Support Vector Machine (SVM) with Optimal Choice of Kernel Function Parameters, *Expert Systems with Applications* 28, 603-614, 2005.
16.   Ohlson J., Financial Ratios and the Probabilistic Prediction of Bankruptcy, *Journal of Accounting Research* 18(1), 109-131, 1980.
17.   Park, C. S. and I. Han, A Case-Based Reasoning with the Feature Weights Derived by Analytic Hierarchy Process for Bankruptcy Prediction, *Expert Systems with Applications* 23 (3), 255-264, 2002.
18.   Peng, Y., G. Kou, G. Wang, H. Wong, and F. Ko, Empirical Evaluation of Classifiers for Software Risk Management, *International Journal of Information Technology & Decision Making*, 8(4), 749-767, 2009.
19.   Shin, K. S. and Y. J. Lee, A Genetic Algorithm Application in Bankruptcy Prediction Modeling, *Expert Systems with Applications* 23 (3), 321-328, (Springer-Verlag, Berlin), 2002.
20.   Shumway, T., Forecasting Bankruptcy More Accurately: A Simple Hazard Model, *The Journal of Business* 74(1), 101-124, 2001.
21.   SOX: Sarbanes-Oxley Act. 2002.
22.   Sun, L., A Re-Evaluation of Auditor's Opinions versus Statistical Models in Bankruptcy Prediction, *Review of Quantitative Finance and Accounting* 28, 55-78, 2007.
23.   Sung, T. K., N. Chang, and G. Lee, Dynamics of Modeling in Data Mining: Interpretive Approach to Bankruptcy Prediction, *Journal of Management Information Systems* 16(1), 63-85, 1999.

NOTES