The emerging financial pre-warning systems

Mushang Lee[1*], Ching-Hui Shih[2], Tsui Chih Wu[3]

[1]Department of Accounting, Chinese Culture University, 55, Hwa-Kang Rd., Yang-Ming-Shan, Taipei, Taiwan (ROC) (leemushan@yahoo.com.tw)

(* Corresponding Author. Tel.: +886 2-28610511x35501)

[2] Department of Accounting and Information Systems, Asia University, 500, Liufeng Rd., Wufeng, Taichung, Taiwan (ROC) (chshih@asia.edu.tw)

[3] Department of Accounting, Shih Chien University, 70, Dazhi St., Zhongshan Dist., Taipei, Taiwan (ROC) (tsui@mail.usc.edu.tw)

Abstract - **The exact prediction of financial crises is an essential research task for decision makers. In recent years, data mining techniques have been used to tackle the related problems and perform a satisfactory job in various domains. However, in the information age, utilizing straightforward data mining techniques to predict financial crises has many shortcomings and limitations. Thus, this investigation utilized the random forest (RF) technique as a pre-processing procedure to determine the most representative features. Then, the selected features were fed into rough set theory to yield interpretable information for decision makers, who can use it to make suitable judgments in a turbulent economic climate. The proposed model is a promising alternative for predicting financial crisis, and it can assist in regard to both taxation and financial institutions.**

Keywords - **Financial crisis; Random forest; Rough set theory; Decision making.**

I. Introduction

With the radical changes in global financial markets and multinational trade, the prediction of corporate financial distress is playing an increasingly important role. Financial distress often takes place when a corporation has chronic and serious losses or when the corporation becomes insolvent with liabilities that are disproportionate to assets (Gestel et al., 2006). Corporate financial distress has a devastating effect on a company's shareholders and employees, and it can ruin a firm's reputation and credibility. In recent years, the financial tsunami started to impair the economic development of many countries, including Taiwan. When the world economy slips into depression, it remains a very important issue holding the public's attention.

Taiwan, which is located in East Asia, plays an important role in the global supply chain of electronic products; it is also an important capital market among global investors (Lin, 2009). However, there are serious shortcomings with regard to corporate governance in East Asia, as La Porta et al. (1999) have indicated. First the ownership structure is less dispersed than in Europe and America, which increases the likelihood of corporate financial distress. A second shortcoming of many corporations in East Asia is that the ultimate controllers (i.e., directors or managers) often maximize their power by means of a pyramid structure and cross shareholding. They also tend to pledge their shareholdings as loan collateral (Lin, 2009). When top managers or directors of Taiwanese corporations manipulate financial statements, they frequently do so to prevent a decrease in the share price. The "window dressing" of financial reports does not reflect the true value of corporate financial situations and causes serious damage to society as a whole.

Several models have been suggested to predict financial distress, using ratios and data originating from these financial statements. For example: univariate approaches (Beaver, 1966), multivariate approaches, linear multiple discriminant approaches (MDA) (Altman, 1986; Altman and Levallee, 1980), multiple regression (Meyer and Pifer, 1970), logistic regression (Dimitras et al., 1996), factor analysis (Blum, 1974) and stepwise (Laitinen and Laitinen, 2000). Nevertheless, traditional statistics have obeyed strict assumptions, such as linearity, normality, independence among predictor variables and pre-existing functional forms related to the criterion variable; the predictor variable limited their application in the real world (Hua et al., 2007).

With the progress of computational intelligence, artificial intelligence has been proposed to overcome the aforementioned problem. Random forest (RF) is one of the most successful ensembles for learning proposed by

978-0-7695-4684-1/12 $26.00 © 2012 IEEE
DOI 10.1109/IMIS.2012.156

Breiman (2001). The algorithm is composed of numerous tree-shaped classifiers. Each tree-shaped classifier utilizes a specific dataset established by a bootstrap procedure. In addition, the random character is based on the bagging embedded in the algorithm, which leads to the further development of bagging; it has no problem with over-fitting due to the strong law of large numbers (Brieman, 2001). Thus, RF has become a well-known data analysis algorithm. Additionally, identifying the most essential features is another byproduct of RF. In the information age, we encounter databases in which the number of objects becomes greater, and their dimensionality (features or attributes) increases too. Attributes that are useless or redundant to recognition tasks may deteriorate the performance of learning algorithms (Qian et al., 2011). Finding how to determine the useful information for decision makers to make reliable judgments is a challenging task in the turbulent economic environment. Executing feature selection can lead to many potential advantages; these include facilitating data visualization and data understanding, eliminating the assessment and storage requirements, decreasing the training and operational times, and overcoming the curse of dimensionality to enhance prediction performance.

Rough set theory (RST), introduced by Pawlak (1982) in the early 1980s, is an emerging mathematical tool that can be utilized to tackle uncertainty and vagueness. It focuses on the discovery of patterns in inconsistent data structures and can be employed on this basis to perform formal reasoning under uncertainty, machine learning and knowledge discovery (Zhai et al., 2002). This method has its specific merits compared with other machine learning methods. RST does not require any pre-defined or additional information about the empirical training data, such as probability distribution in statistics or grades of membership in fuzzy set theory (Pawlak, 1992). Furthermore, the RST can generate knowledge in logic statements and express comprehensive rules in "if (condition), then (decision)" format. The rules are easily interpretable. Generating the best reduction is one of the challenges in RST. Thus, this study presents emerging financial pre-warning systems which hybridize RF and RST to overcome the related problems. The former is utilized to determine the best reduction, and then the best reduction is fed into RST to analyze the financial risk tasks. The advantage of this system pertains to its ability to minimize decision maker risks by predicting corporate financial distress, as well as presenting the decision makers with a set of comprehensible decision rules. The evidence provided by this study also offers policymakers the ability to evaluate the policy implications of corporate governance mechanisms, as well as to formulate future policies. The remainder of this study is organized as follows: Section 2 introduces the methodologies used in this study. The experimental results are presented in Section 3, and a summary of the findings appears in Section 4.

II. Methodologies

2.1 Random forest

Random forest is one of the ensemble learning techniques utilizing tree-shaped classifiers $\{h(x,\vartheta_k), k=1,\ldots,n\}$ where the independent identical distributed random vector and input pattern were expressed as $\{\vartheta_k\}$ and x, respectively. In model construction process, this technique generates numerous CART-like trees, each trained on a bootstrapped sample of the original training data, and searches only across a randomly selected subset of the input variables to determine a split (Gislason et al., 2006). For classification task, each tree in RF casts a unit vote for the most popular class at input x. The final output of classifier is determined by majority vote of trees. The number of variables is a user-defined parameter, but the technique is not sensitive to it. Normally, the value is set to the square root of the number of inputs. By constraining the amount of variables employed for a split, the operational time is eliminated and the correlation between trees is also decreased. Eventually, the trees in RF are not pruned, further decreasing the computational complexity.

To assess the performance (forecasting accuracy) of test set, the out of bag samples of each tree can be run down through the tree. The outcomes are integrated with a majority vote as before, and when compared to the actual class labels, they generate a lower estimate of the forecasting error, as each sample can only be evaluated on the trees for which it was out-of-bag (Gislason et al., 2006).

Simultaneously, the importance of variable m can be evaluated by randomly permuting all the values of the m th variable in the out of bag samples for each classifier. If an increased out-of-bag error is yielded, that is an identification of the importance of that variable. Thus, the best reduct determination was conduct by RF.

2.2 Rough set theory

The RST proposed by Pawlak (1982) is an useful mathematical tool in identifying hidden knowledge, characterized by vague and uncertain information, and in generating decision rules. RST adopts information systems to represent knowledge and deal with vague data. An information system containing condition attributes and decision attributes is depicted as follow:

$$IS = (U, \Omega, V, g) \quad (1)$$

where U denotes a nonempty finite set with n objects $\{p_1, \dots, p_n\}$, Ω is a nonempty finite set with m attributes $\{q_1, \dots, q_m\}$. V is called the range of $U \times \Omega$, and $f : U \times \Omega \to V$ is an information function where $g(p, q) \in V$ for every $p \in U, q \in \Omega$. In addition, let $Q \subseteq \Omega$ and $(x, y) \in U \times U$. At present, x and y are two objects.

Indiscernibility arises from an inability to distinguish among objects in a distinct set, and results in identical information derived from different observations. The indiscernibility relation of x and y in terms of Q is defined as follows:

$$IND(Q) = \left\{ \begin{matrix} (X, Y) \in U \times U : \\ g(x, q) = g(y, q) \forall q \in Q \end{matrix} \right\} \quad (2)$$

The indiscernibility relation partitions the universe U into a family of equivalence classes. The equivalence classes of the relation, $IND(Q)$, are called the Q-elementary sets in IS, and $[x]_{IND(Q)}$ represents the Q-elementary set containing the objective $x \in U$. In RST, knowledge of objects is presented in a decision table.

Lower and upper approximation is the second essential concept of RST. Let $Q \subseteq \Omega$, and $X \subseteq U$. The Q-lower approximation of $X(Q_L)$ and the Q-upper approximation of $X(Q_U)$ are then defined as follows respectively:

$$X(Q_L) = \{x \in U : [x]_{IND(Q)} \subseteq X\} \quad (3)$$

$$X(Q_U) = \{x \in U : [x]_{IND(Q)} \cap X \neq \phi\} \quad (4)$$

Each object in the lower approximation set of X must be in X. If an objective is in the upper approximation set of X, then it might or might not be in X.

Two essential RST concepts in knowledge reduction are the reduct, represented by $RED(Q)$ and core, depicted as $CORE(Q)$ The reduction of attributes eliminates some irrelevant or redundant attributes without reducing the quality of the approximation of the information system based on the original set of attributes. When redundant attributes are removed, the indiscernibility relation of an attribute set remains unchanged. A reduct is a basic component of an information table, of which the core is the intersectionof all reducts. The relation between reducts and the core can be represented as follows:

$$CORE(Q) = \cap RED(Q) \quad (5)$$

Reducts can be derived through the discernibility matrix and Boolean operations (Nguyen and Skowron, 1995). The discernibility matrix is a set that can identified between two objects or sets.

The rule generation from decision table to classify new objects is one of the most significant functions of RST. Rules are derived from the condition attributes based on the decision table. Furthermore, a decision rule can be written as “IF condition(s), THEN decision(s).” The prediction of the new objective is performed by matching its description to one of the rules. A support of the rule is represented as Eq. (6)

$$Supp = CARD(\|\Omega \cap V\|) \quad (6)$$

where $CARD$ is the cardinality of the set.

Rules with higher support values are more general, and express more information in the data set which can be classified by decision rules. In addition, the coverage factor expressed as Eq. (7) is another essential criterion for measuring the performance of the RST. The coverage shows the degree to which the reasons for a decision can be trusted in terms of the data (Pawlak, 2002).

$$Coverage = Supp / CARD(\|V\|) \quad (7)$$

Because many classification algorithms, including RST, only accept discrete features, the investigation adopts entropy to discretize the continuous features of data.

III. Experimental results

The Taiwan Stock Exchange (TSE) and the database of the Taiwan Economic Journal (TEJ) provided the data in this study; Taiwan experienced financial distress from 2007 to 2010. The financial distress criteria for sampling required a firm to announce that stocks needed to be “Traded” or “Terminated.” In other words, it may have been cited as (1) having a credit crisis, (2) having a net operating loss, (3) failing to pay debts, or (4) violating regulations. Corporations experiencing financial distress were paired with financially sound corporations by (1) industry, (2) products, (3) capitalization and (4) the value of total assets. The size of the matched sample was 194 firms, including 97 corporations in financial distress, and 97 financially sound corporations.

After deleting features with missing values, the previous research, experiences from past decisions and the domain knowledge of financial experts in that industry, there were 14 financial features used in this study. Table 1 represents the definition of each feature.

TABLE 1. DEFINITION OF FEATURES USED IN THE STUDY

Category	Features	Definition
Stability	X1	Debt ratio
	X2	Long-term debt ratio
	X4	Cash and cash equivalents to current liabilities
	X5	Quick Ratio
Profitability	X10	Net income to total asset
	X11	Financial expenses to sales
	X13	EBIT
Growth	X3	Sales to total assets
	X6	Subsistence income to total assets
Activity	X7	Inventory turnover
	X8	Account receivable turnover
	X9	Total asset turnover
Cash flow	X14	Cash flow ratio

Feature selection has key steps in designing pattern recognition and machine learning systems because of the need to identify and/or extract the most relevant features for a given classification or recognition task (Sotoca and Pla, 2010). Thanks to the feature selection process, the computational cost decreases while the classification performance can increase (Polat and Güneş, 2009). In this study, the applied feature selection method was used, called the RF method, to identify the informative features and then was fed into the RST model to construct an emerging pre-warning system. Selected features from RF are indicated in Fig 1.

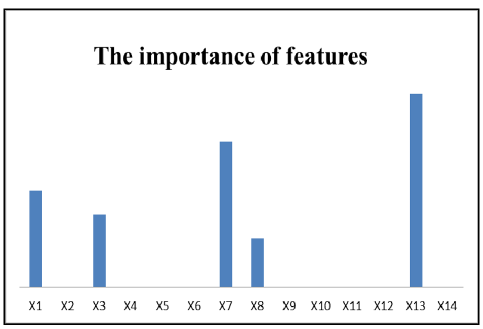

Figure 1. The selected features determined by RF.

To obviate the "over-fitting" problem, a 5-fold cross-validation procedure was conducted whereby training data were divided into subsets of equal size. Each subset was tested sequentially by adopting the classifier trained on the remaining four subsets. An instance of the whole training set was predicted once; the cross-validation (CV) accuracy was the percentage of data that was correctly classified. To examine the effectiveness of feature selection, the evaluation process was divided into two conditions: with feature selection and without feature selection. The result is expressed in Table 2. According to the research findings, the data that have undergone the feature selection process can perform a satisfactory job in terms of forecasting accuracy. Fig. 2 expresses the result under the two different conditions of accuracy improvement.

TABLE 2. TESTING ACCURACY OF PRE-WARNING MODEL UNDER TWO CONDITIONS

Cross Validation	Emerging pre-warning model	
	With feature selection	Without feature selection
	Accuracy	Accuracy
CV-1	84.21	81.58
CV-2	86.84	78.95
CV-3	84.21	81.58
CV-4	89.47	78.95
CV-5	83.33	71.43
AVG	85.61	78.5

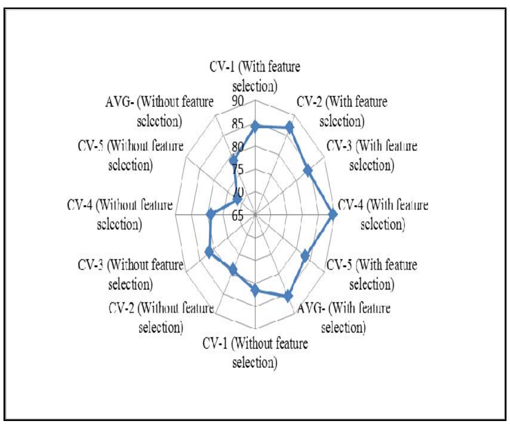

Figure 2. The accuracy improvement results of pre-warning model under two different conditions

Another indicator used in this study, for assessing the performance of the feature selection is the misclassification rate. A Type I error means that a financially sound corporation has been misclassified as being in financial distress. A Type II error means that a corporation in financial distress has been misclassified as being financially sound. Type I and Type II errors have different costs. Type I errors may lead to additional investigation. Type II errors may result in unacceptable economic losses. Table 3 depicts both types of errors obtained from the pre-warning model under the two different conditions.

TABLE 3. TYPE I AND TYPE II ERRORS FOR PRE-WARNING MODEL UNDER TWO DIFFERENT CONDITIONS

Cross Validation	Pre-warning model (%)					
	With feature selection		Without feature selection		Error improvement	
	Type I	Type II	Type I	Type II	Type I	Type II
CV-1	21.05	10.53	21.05	15.79	0.00	5.26
CV-2	21.05	5.26	17.24	15.79	-3.81	10.53
CV-3	15.79	15.79	10.53	26.32	-5.26	10.53
CV-4	15.79	5.26	26.32	15.79	10.53	10.53
CV-5	19.05	14.29	47.62	9.52	28.57	-4.77
AVG	18.55	10.23	24.55	16.64	6.00	6.42

Rather than resorting to complicated mathematical functions, it is easier for decision makers to intuitively realize the relation as well as the strengths of features pertaining to financial distress from a set of decision rules. Table 4 shows five rules from financial pre-warning models. The decision makers can take the rules as a guideline to lay much more emphasis on the corporation which poses a higher possibility of falling into financial distress. A well-established pre-warning model not only strengthens the investors' confidence in the stock market, but also ensures the stability of the economic environment.

TABLE 4. THE DECISION RULES DERIVE FROM FINANCIAL PRE-WARNING MODEL

Decision rules
Rule 1: If "Sales to total assets (X3)" is "39.37%~47.13%", and "EBIT (X13)" is "675,000~1,080,000", then "Non-financial distress"
Rule 2: If "Debt ratio (X1)" is "36.3%~40.6%", and "Inventory turnover (X7)" is "4.17~5.09", then "Non-financial distress"
Rule 3: If "Debt ratio (X1)" is 41.5%~42.4%",and "Sales to total assets (X3)" is "2%~7%", then "Financial distress"
Rule 4: If "Inventory turnover (X7)" is "0~0.66", and "Debt ratio (X1)" is "40.7~49.53", then "Financial distress"
Rule 5: If "Debt ratio (X1)" is "42.8%~45.9%", and "Inventory turnover (X7)" is "0.85~1.12", then "Financial distress"

IV. Conclusion

The prediction of financial distress is an important and challenging issue that has been rigorously investigated in recent years as the amount of financial distress has increased. Many different kinds of technology have been introduced to deal with the related problems, and the attempt to improve on those models continues. It is noteworthy that feature selection is quite essential for financial pre-warning model construction. Therefore, the current investigation presents an emerging financial pre-warning model which hybridizes RF and RST for analyzing the financial distress problem. The empirical results indicate that the proposed model is an effective and efficient alternative in predicting financial distress. Rather than presenting complicated mathematical functions, the present model provides a set of comprehensible decision rules for decision makers to make reliable judgments.

This investigation has the following limitations that need further research. First, adopting other feature selection techniques would give a satisfactory result. Second, the other informative features such as audit committees, boards with a significant proportion of outside members and CEO duality might be included in the pre-warning model to further enhance its effectiveness in predicting financial distress.

REFERENCES

[1] Altman, E.I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. Journal of Finance, 23, 589-609.

[2] Altman, E.I., Levallee, M.Y. 1980. Failure classification in Canada. Journal of Business Administration, 12, 147-162.

[3] Beaver, W.H. 1966. Financial ratios as predictors of failure. Journal of Accounting Research, 4, 71-111.

[4] Breiman, L. 2001. Random forests. Machine Learning 45(1), 5-32.

[5] Blum, M. 1974. Failure company discriminated analysis. Journal of Accounting Research, 12, 1-25.

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

[7] Gestel, T. V., Baesens, B. Suykens, J.A.K., Poel, D. V., Baestaens, D.E., Willekens, M. 2006. Bayesian kernel based classification for financial distress detection. European Journal of Operational Research, 172(3), 979-1003.

[8] Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R. 2006. RFs for land cover classification. Pattern Recognition Letters, 27, 294-300.

[9] Hua, Z., Wang, Y., Xu, X., Zhang, B., Liang, L. 2007. Predicting corporate financial distress based on integration of support vector machine and logistic regression. Expert Systems with Applications, 33 (2), 434–440.

[10]La Porta, R., Lopez-de-Silanes, F., Shleifer, A. 1999. Corporate ownership around the world. Journal of Finance 54(2), 471-517.

[11]Laitinen, E.K., Laitinen, T. 2000. Bankruptcy prediction: application of the Taylor's expansion in logictic regression. Journal of International Review of Financial Analysis, 9, 327-349.

[12]Lin, T. H. 2009. A cross model study of corporate financial distress prediction in Taiwan: Multiple discriminant analysis, logit, probit and neural networks. Neurocomputing, 72(16-18), 3507-3516.

[13]Meyer, P.A., Pifer, H.W. 1970. Prediction of bank failures. Journal of Finance, 25, 853-868.

[14]Nguten, S.H., Skoworn, A. 1995. Quantization of real value attributes-rough set and Boolean reasoning approach. Proc of the Second Joint Conference on Information Sciences, 34-37.

[15]Pawlak, Z. Rough sets.1982. International journal of computer and information sciences, 11, 341-356.

[16]Pawlak, Z. 1992. Rough sets: theoretical aspects of reasoning about data. Kluwer Academic Publisher, Dordrecht, 1st Edition.

[17]Polat, K., Günes, S. 2009. A novel hybrid intelligent method based on C4.5 decision tree classifier and one-against-all approach for multi-class classification problems. Expert Systems with Applications, 36(2), 1587-1592.

[18]Qian, Y., Liang, J., Pedrycz, W., Dang, C. 2011. An efficient accelerator for attribute reduction from incomplete data in rough set framework. Pattern Recognition, 44(8), 1658-1670.

[19]Sotoca, J.M., Pla, F. 2010. Supervised feature selection by clustering using conditional mutual information-based distances. Pattern Recognition, 43 (6), 2068–2081

[20]Zhai, L.Y., Khoo, L. P., Fok, S.C. 2002. Feature extraction using rough set theory and genetic algorithms—an application for the simplification of product quality evaluation. Computer and Industrial Engineering, 43, 661–676.