Business Failure Prediction Model based on Grey Prediction and Rough Set Theory

JAO-HONG CHENG[1], HUEI-PING CHEN[2], KAI-LUN CHENG[3]
[1, 2,3] Department of Information Management
National Yunlin University of Science and Technology
No.123, Sec. 3, Dasyue Road., Douliou City, Yulin Country, 640
Taiwan, R.O.C.
jhcheng@yuntech.edu.tw[1]; g9623802@yuntech.edu.tw[2]

Abstract: - A lot of methods have been used in the past for the prediction of failure business like Discriminant analysis, Logit analysis, Quadratic Function etc. Although some of these methods lead to models with a satisfactory ability to discriminate between healthy and bankrupt, they endure some limitations, often due to the unrealistic assumption of statistical hypotheses. This is why we have undertaken a hybrid advisable system aiming at weakening these limitations. A hybrid model that predicts the failure firms based on the past financial performance data, combining grey prediction and rough set approach is possible to predict using few data and quickly calculate. The results are very encouraging, compared with original rough set, and prove the usefulness and highlight the effectiveness of the proposed method for firm failure prediction.

Key-Words: - Grey Prediction, Rough Set Theory (RST), Business Failure, Financial Ratio

1 Introduction

The subprime mortgage crisis is an ongoing economic problem that is characterized by contracted liquidity in the global credit markets and became more apparent during 2007 and 2008. And at the end of 2001 Enron Corporation was revealed that it's reported financial condition was sustained substantially by institutionalized, systematic, and creatively planned accounting fraud, sometimes we called the "Enron scandal". Enron has since become a popular symbol of willful corporate fraud and corruption. It is not only quite recently that business failure prediction is a scientific field which many academic and professional people have been working for. Also, financial organizations, such as banks, credit institutions, clients, etc., need these predictions for firms in which they have an interest.

The first approach to predict business failure started with the use of empirical methods proposed by large banks in USA [35]. Then, the financial ratios methodology was developed for the business failure prediction problem. These ratios have been long considered as objective indicators of firm failure (insolvency risk, see [4],[9],[2]). The approach of the financial ratios (also called univariate statistical approach), gave rise to the methods for business failure prediction based on the multivariate statistical analysis. In 1968 already, Altman [2] proposed to use the discriminant analysis for predicting the business failure risk.

In 1982, Deng [10] introduced the Grey theory in order to treat any random variation in the stochastic system as the variation of "Grey" value in time-varying Grey system in a certain range. Conceptually, the Grey theory has been used as tool of data generating to predict the system behavior from those previous random data. In other word, the Grey prediction employs the Grey model, which is to describe the trend of the generating time sequence data from the previous dynamic behaviour of the system.

The concept of a rough set, introduced by Pawlak [24], proved to be an effective tool for the analysis of financial information tables describing a set of firms by a set of multi-valued attributes (financial ratios). Xaio & Zhang [32] indicated application of the variable precision rough sets as an example on remote sensing image classification and the result of performance of system could be found to achieve much improvement after using the variable precision rough sets.Rough set approach has already been used for the analysis and explanation of financing decisions in a Greek industrial development bank called ETEVA [29].

This paper presents a hybrid method which is combined grey prediction and rough set approach for the analysis and prediction of business failure. The aim of the present study is to test the ability of the hybrid method of grey prediction and rough set approach in predicting business failure, and to compare it with original rough set.

2 Literature review

The prediction of bankruptcy for financial firms especially banks has been extensively researched area since late 1960s [2] .The prediction of financial health company is similar to the problem of predicting bankruptcy, which is a well-researched area where several techniques have been used. There are multiple discriminant analysis by Altman [2], multicriteria decision aid methodology by Dimitras et al. [12], support vector machines by Fan and Palaniswami [14], recursive partitioning algorithm by Frydman et al. [15], logit analysis by Ohlson [20]and probit analysis by Zmijewski [35]. In the area of business, rough sets have been used for business failure prediction, database marketing, and financial investment.

Beaver [4] was the first academic in using statistic analysis for business failure prediction. Data based on 79 distress companies from 1954 to 1964 period. To match by a ratio of equality 79 healthy companies that has similar industry type and organization size. He used dichotomous classification to test 14 financial ratios. The study result showed cash flow/total liabilities has the best performance. Secondly order is net benefits/total assets, total liabilities/total assets. Beaver [4] argued that three variables are cash flow, net benefit and total dept can not be dressed easily, since showed permanent factors of companies. Thus, business failure announce majorly by those factors.

Dimitras et al. [12] used a sample of 80 Greek firms and compared the rough sets method to inductive learning, discriminant analysis, and logit analysis. They showed that the rough sets method performed significantly better than other methods in terms of classification accuracy when 12 financial ratios were used for predicting failure or success of companies. A hybrid technique consisting of a genetic algorithm coupled with rough sets was used by McKee and Lensberg [20] to predict bankruptcies for US public companies using a sample of 291 companies. It was shown to perform better than independent rough sets analysis.

Another example of a hybrid approach for business failure prediction was discussed in Ahn et al. [1], where the rough sets technique was used to reduce the number of independent variables and generate rules linking them with the dependent variable. For instances that matched any of the rules, classification was performed using rough sets. For those that did not, classification was made using a neural network. Using a sample of 2400 Korean firms, this hybrid approach was shown to perform better than approaches based on discriminant analysis or neural networks.

Bose [6] conducted an empirical investigation of dot-coms from a financial perspective. Data from the financial statements of 240 such businesses was used to compute financial ratios and the rough sets technique was used to evaluate whether the financial ratios could predict financial health of them based on available data. It was shown that rough sets performed a satisfactory job of predicting financial health and were more suitable for detecting unhealthy dot-coms than healthy ones.

Cheng et al. [8] identified significant indicators of business failure. Rough set models with decision rules for business failure prediction were constructed in order to examine critical attributes required for rule generation. 14 financial ratios commonly used in existing business failure models and particularly a non – financial variable, auditor switching, are used in rough set models to enhance the predictive performance. The results of the empirical study showed that auditor switching and cash flow ratio are the most significant indicators of business failure. The critical indicators included quick ratio, earnings per share, and cash flow adequacy ratio.

The purpose of our research describe as (1) discussing and reviewing about business failure model, (2) including time series data and using grey prediction to analysis financial distress, and (3) combining grey prediction and rough set theory in business failure prediction model. There are two major steps in this research which Computing predicted value by GM(1,1) model as input variable to construct business failure prediction model and Combining grey prediction and rough set theory as business failure prediction model.

2.1 Grey Prediction

Recently, there are advanced applications of Grey prediction in various areas such as stock price or control schemes in material processing [30],[16]. Conceptually, the Grey theory is used as tool of data generating to predict behaviour of the system from those previous random data. In other word, the Grey prediction employs the Grey model, which is to describe the trend of the generating time sequence data from the previous dynamic behaviour of the system.

Yu & Luo [33] indicated the advantages of the gray incident model were (1) reflecting the system condition in a completely dynamic way, (2) identifying problems in condition of inexplicit system tendency, which reduces the sensitivity to and dependence on decision-making environment, and (3) considering the relationship between absolute quantity and changing rate during the identification process, which complies with thinking patterns of decision-makers in reality. Chen [7] established forecast models for the cross-strait trade through the methods of grey forecasting, simple regression, and exponential smoothing and the results indicated that for model selection, it is more correct to perform the forecast on trade volume based on the statistical information released by Taiwan Customs.

Grey prediction is a predicted model to find changes in law and to construct predicted model by collected data of sample. Grey system assumed that any variables have a range and time period. In grey system, original discrete sequence will be found the law by generating operation and fitting new value by the grey differential equation. The grey prediction method only required a few sampled data to develop the grey model and go forecast the future. Grey prediction organized as follow stage:

(1) Accumulated Generating Operation (AGO for short), the purpose is decreasing the randomness by translating original sequence to an incrementally one and increase the smoothness of the sequence. AGO can be shown by equation as follow:

assume $x^{(0)}$ as a original sequence

$$x^{(0)}=\{x^{(0)}(1),x^{(0)}(2),...,x^{(0)}(n)\}$$

the first of generating operation of AGO is

$$x^{(1)}=\{x^{(1)}(1),x^{(1)}(2),...,x^{(1)}(n)\}$$

and $x^{(1)}(2)=x^{(0)}(1)+x^{(0)}(2)$

$$x^{(1)}(3)=x^{(0)}(1)+x^{(0)}(2)+x^{(0)}(3)$$

$$x^{(1)}(k)=\sum_{i=1}^{k}X^{(0)}(i),k=1,2,....,n$$

the second of generating operation of AGO is

$$x^{(2)}=\{x^{(2)}(1),x^{(2)}(2),...,x^{(2)}(n)\}$$

And so on

The standard equation of AGO is

$$\left(x^{(r)}(k);k=1,2,3,...,n\right)=\left(\sum_{m=1}^{k}x^{(r-1)}(m)\right)$$

r is accumulating times; $r\geq 1$

(2) Inverse Accumulated Generating Operation (IAGO for short), is an inversive process in AGO. The purpose is inverse accumulated sequence to orginal one in validating precision. IAGO can be shown by equation as follow:

Assume that r times IAGO of $x^{(k)}$ is $\alpha^{(r)}\left[x^{(k)}(i)\right]$,it can be shown

$$\alpha^{(0)}\left[x^{(k)}(i)\right]=x^{(k)}(i),$$

$$\alpha^{(1)}\left[x^{(k)}(i)\right]=\alpha^{(0)}\left[x^{(k)}(i)\right]-\alpha^{(0)}\left[x^{(k)}(i-1)\right],$$

$$\alpha^{(2)}\left[x^{(k)}(i)\right]=\alpha^{(1)}\left[x^{(k)}(i)\right]-\alpha^{(1)}\left[x^{(k)}(i-1)\right],$$

....

$$\alpha^{(j)}\left[x^{(k)}(i)\right]=\alpha^{(j-1)}\left[x^{(k)}(i)\right]-\alpha^{(j-1)}\left[x^{(k)}(i-1)\right],$$

then

$$\alpha^{(1)}\left[x^{(k)}(i)\right]=\alpha^{(0)}\left[x^{(k)}(i)\right]-\alpha^{(0)}\left[x^{(k)}(i-1)\right],$$

$$=x^{(k)}(i)-x^{(k)}(i-1)$$

$$=x^{(k-1)}(i),$$

$$\alpha^{(2)}\left[x^{(k)}(i)\right]=\alpha^{(1)}\left[x^{(k)}(i)\right]-\alpha^{(1)}\left[x^{(k)}(i-1)\right],$$

$$=x^{(k-1)}(i)-x^{(k-1)}(i-1)$$

$$=x^{(k-2)}(i),$$

....

$$\alpha^{(r)}\left[x^{(k)}(i)\right]=x^{(k-r)}(i)$$

Using $\alpha^{(r)}\left[x^{(k)}(i)\right]=x^{(k-r)}(i)$, can inverse k times accumulated sequence to original one.

(3) Grey differential equation, because of accumulated sequence has clear exponential law, it can be fitting by grey differential equation. Differential equation included multi-order differentiation: GM(n,n) and first-order differentiation: GM(1,n), when n=1, GM(1,1) is specialization of GM(1,n). In this research, we use GM(1,1) to construct business failure prediction model because of the sample consisted of annual data of companies. GM(1,1) can be shown by equation as follow:

Assume original sequence as:

$$x^{(0)}=\{x^{(0)}(1),x^{(0)}(2),...,x^{(0)}(n)\}$$

to generate AGO once

$$x^{(1)}=\{x^{(1)}(1),x^{(1)}(2),...,x^{(1)}(n)\},$$

then

$$x^{(1)}(k)=\sum_{i=1}^{k}X^{(0)}(i),k=1,2,....,n$$

Contributing $x^{(1)}$ differential equation

$$\frac{dx^{(1)}(t)}{dt} + aX^{(1)}(t) = b$$

It can be replaced in

$$x^{(0)}(j) + az^{(1)}(j) = b, \forall \quad j = 2,3,4...,n$$

Then $z^{(1)}(j) = 0.5x^{(1)}(j) + 0.5x^{(1)}(j-1)$

commuting with least squares method

$$\hat{a} = \begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T Y_N,$$

$$B = \begin{bmatrix} -z^{(1)}(2),1 \\ -z^{(1)}(3),1 \\ -z^{(1)}(4),1 \\ \vdots \\ -z^{(1)}(n),1 \end{bmatrix}; \quad Y_N = \begin{bmatrix} -x^{(0)}(2) \\ -x^{(0)}(3) \\ -x^{(0)}(4) \\ \vdots \\ -x^{(0)}(n) \end{bmatrix}$$

As description above

$$x^{(1)}(1) = x^{(0)}(1)$$

We can calculate

$$\hat{x}^{(0)}(j) = \left[x^{(0)}(1) - \frac{b}{a} \right] e^{-a(j-1)} + \frac{b}{a}; \text{ j=2,3,…n}$$

2.2 Rough Set Theory

Rough sets theory was proposed by Pawlak [24]. The central concept of rough sets is a collection of rows that have the same values for one or more attributes. These sets called elementary sets are indiscernible [27]. Rough sets incorporate the use of indiscernible relations to approximate sets of objects by upper and lower set approximations. The upper and lower approximations represent the classes of indiscernible objects that possess sharp descriptions on concepts but without sharp boundaries.

The elementary set forms a basic granule of knowledge about the universe. Any subset of the universe can be expressed either precisely or roughly [13]. A certain subset of the universe can be characterized by two ordinary sets which are called the lower and the upper approximation set. For each subset X of the universe, the lower approximation set consists of all objects which certainly belong to X and the upper approximation contains objects which possibly belong to X.

The focus of initial rough sets applications is mainly placed on medical diagnosis, drug research, and process control. In fact, in recent years there has been a rapid growth of interest in rough sets theory and its applications, worldwide [28].

2.2.1. A formal treatment

This section recalls necessary rough sets notions used in the paper. Formally, an information table or information system [26] is the 4-tuples, S = (U, Q, V, f), where U is a non-empty, finite set of objects which are called the universe, Q is a non-empty, finite set of attributes, $V = U_{q \in Q} \; V_q$ is a set of attribute values and V_q is a domain of the attribute q, and f: U × Q→V is a total function such that f(u, q) ∈ V_q for every u ∈ U, q ∈Q, called an information function [25]. The attributes in Q are composed of two disjoint subsets, condition attributes C and decision attributes D, such that $Q = C \bigcup D$ and C∩D = Φ.

Given an information system S, any attribute subset P of Q, and any x and y is members of U. A binary relation R(P) on U, called an indiscernible relation, is defined as $R(P) = \{(x, y) \in U \times U : \;\; \forall p \in P, \;\; f(x, p) = f(y, p)\}$. If f(x, p) = f(y, p) for every p∈P, then x and y are said to be indiscernible by the set of attributes P. Obviously, R (P) is an equivalence relation. The family of all equivalence classes of R (P), a partition determined by P, will be denoted by U/P. An equivalence classes of R (P) containing x will be denoted by $[x]_{p.}$ $Let [x]_p = \{y \in U : \forall_x \in U, xR(P)y\}$.

Information system contains knowledge about the universe in terms of a predefined set of attributes. A subset of the universe is called a concept in rough sets theory. For representation or approximation of these concepts, an equivalence relation is defined. The equivalence classes of the equivalence relation, which are the minimal blocks of the information system, can be used to approximate these concepts. Concepts that can be constructed from these blocks are called definable sets. Upper and lower approximation set can construct to approximate the concepts called indefinable sets. To this end we define two operations assigning to every X⊂U two set $P*(X)$ and $P^*(X)$ called the P-lower and P-upper approximation of X, respectively, and defined as follows.

The conditional probability of a concept X on the equivalence class $[x]_p$ can be defined as $\Pr(X \mid [x]_p) = card(X \cap [x]_p) / card([x]_p)$, where card denotes cardinality of the set.

$$P*(X) = \{x \in U : \forall [x]_p \in U / P, \Pr(X \mid [x]_p) = 1\}.$$

$$P^*(X) = \{x \in U : \forall [x]_p \in U / P, \Pr(X \mid [x]_p) > 0\}.$$

In addition, the set $BND_p(X) = P^*(X) - P*(X)$ will be referred to as the P-boundary region of X. A boundary region contains a set of objects that a set of attributes P can't classify as belonging to or not belonging to concept X. Therefore, if the boundary region of X is the empty set, then the set X will be called crisp with respect to P, and otherwise the set X will be referred to as rough with respect to P [26].

2.2.2. The variable precision rough sets model

The variable precision rough sets (VPRS) model is a generalized model of rough sets [34]. The generalization is aimed at handling uncertain information and is directly derived from the original rough sets model without any additional assumptions. By making use of the statistical information in the data, this model can extend upon the original rough sets model and handle incomplete and ambiguous data [3].

The VPRS model deals with partial classification by introducing a probability value βwhich measures the size of the largest group of objects as a proportion of the total number of objects in an equivalence class [5]. The βis a real number in the range (0.5, 1]. If the value of βis 1, then the VPRS model is the standard model of rough sets. Thus, the standard model of rough sets becomes a special case of the VPRS model.

Given an information system S, let $Y \subseteq U$ be a concept and P a selected set of attributes. The β-positive region of a concept X with respect to P is defined as

$$POS_P^\beta(Y) = \{x \in U : \forall[x]_P \in U/P, \Pr(Y \mid [x]_P) \geq \beta\}.$$

The β-negative region of concept Y with respect to P is defined as

$$NEG_P^\beta(Y) = \{x \in U : \forall[x]_P \in U/P, \Pr(Y \mid [x]_P) \leq 1-\beta\}.$$

The β-boundary region of concept Y with respect to P is defined as

$$BND_P^\beta(Y) = \{x \in U \forall[x]_P \in U/P, 1-\beta < \Pr(Y \mid [x]_P) < \beta\}.$$

2.2.3. Feature selection based on rough sets

Feature selection is a process of finding an optimal subset of features according to the given goal of processing and criterion from the original set of features. A solution of an optimal feature selection does not need to be unique. Dash et al. [11] recommended four basic steps in the process of selecting:

(1) A generation procedure to generate the next candidate subset;

(2) An evaluation function to evaluate the subset under examination;

(3) A stopping criterion to decide when to stop; and

(4) A validation procedure to check whether the subset is valid.

A rough sets approach to feature selection is based on calculation of a core for discrete attribute data set, containing strongly relevant features, reducts, and a core plus additional weakly relevant features, such that each reduct is satisfactory to determine concepts in the data set. A minimal set of features is satisfactory to describe concepts in a given data set, including a core and possibly some weakly relevant features, form a reduct. A core is an intersection of reducts. The core is a collection of the most significant attributes for the classification in the system.

A feature a_i is relevant if there exists some value v_i of that feature, a decision value v, and value $\mathbf{v_i}$ for which $P(a_i = v_i) > 0$ such that $P(d = v, a_i = v_i | a_i = v_i) \neq P(d = v, a_i = v_i)$ where $\mathbf{a_i}$ is a vector of features (a1, a2,…,ai-1, a_{i+1}, …, a_n) obtained from the original feature vector a by removing a_i and $\mathbf{v_i}$ is denoted a value of $\mathbf{a_i}$ [17].

A feature a_i is strongly relevant if there exists some value v_i of that feature, a decision value v, and value $\mathbf{v_i}$ for which $P(a_i = v_i, a_i = v_i) > 0$ such that $P(d = v, a_i = v_i | a_i = v_i) \neq P(d = v, a_i = v_i)$ [17].

Strong relevance implies that the removal of a feature from a feature vector will change prediction accuracy. A feature a_i is weakly relevant if it is not strong. A weakly relevant feature sometimes may improve prediction accuracy. A feature is relevant whether it's strong or weak, otherwise it is irrelevant. An irrelevant or redundant feature that has no any classification ability thus can be removed.

In addition, inconsistency can be used as a measure criterion for feature subset evaluation. An important idea in data analysis is discovering dependencies between attributes. Let C and D be subsets of Q. To say that C depends on D in a degree k $(0 \leq k \leq 1)$, denoted C ⇒k D, if $k = \gamma(C, D) = card(U_{z \in U/D} POS_c^\beta(Z) / card(U)$. If k=1 we say that D depends totally on C, and if k<1, we say that D depends partially (in a degree k) on C. The coefficient k expresses the ratio of all elements of the universe, which can be properly classified to blocks of the partition U/D, employing attributes C and will be called the degree of the dependency. The coefficient $1-\gamma$ (C, D) can be called the inconsistency degree of information system [19].

It is now possible to define the significance of a feature. Let D⊆Q be a decision attribute and c∈P be a condition attribute. If γ(P, D) =γ(P-{c}, D), then the condition attribute c is regarded as superfluous factor and is removed. The remaining set of condition attributes is a reduct. If there is more than one reduct, then select a best reduct that depends on the optimality criterion associated with the attributes [3]. Some researchers select the reduct with the smallest number of attributes or the maximum number of objects classified (the strength) [30], and others select the reduct on the basis of the increase in the quality of classification by successive augmentation of the subset of attributes [21].

Clearly, not all attributes are necessary while preserving approximate quality of original information system. A reduct is the minimal set of attribute preserving approximate quality. A probabilistic reduct of P with respect to D, $\text{RED}^{\beta}(P,X)$, referred to as a β -reduct. $\text{RED}^{\beta}(P,X)$ is a set of attributes P' ⊆ P ⊆ Q such that (1)γ(P, D) =γ(P', D), and (2) there is no O ⊆ P' such thatγ(P, D) =γ(O, D).

Reduct is one of the most important notions in rough sets. A reduct is the minimal attribute set which preserves classification power of original dataset. All reducts of a dataset can be found by constructing and simplifying discernible function [23]. Unfortunately, it has been shown that finding minimal reduct and all reducts are both NP-hard problems.

2.2.4. Decision rules

Let S = (U, Q, V, f) be an information system. With every P ⊆ Q, a set of formulas For (P) is built up from attribute-value pairs (a, v) where a∈P and $v \in V_a$ by means of logical connectives ∧, ∨, ~ in the standard way.

A decision rule in S is an expression ϕ→Ψ, read if ϕ thenΨ, where ϕ∈ For(C), Ψ∈For(D) and C and D are condition and decision attributes respectively; ϕ and Ψ are referred to as conditions and decisions of the rule respectively. The number $\text{supp}s(\phi,\Psi)=card(\|\phi\wedge\Psi\|s)$ will be called the support of the rule ϕ→Ψ in S. The meaning $\|\phi\wedge\Psi\|s$ is $\|\phi\|s \bigcap \|\Psi\|s$, where $\|\phi\|s$ is defined as $=\{x\in U:\forall_a \in C, v\in V_a, a(v)=x\}$ and $\|\Psi\|s=\{x\in U:\forall_a \in D, v\in V_a, a(v)=x\}$.

The coefficient $cerS(\phi,\Psi)=card(\|\phi\wedge\Psi\|s)/card(\|\phi\|s)$ will be called the certainly factor of the rule ϕ→Ψ in S. The coefficient is now widely used in data mining and is called confidence coefficient. Obviously, $cerS(\phi,\Psi)=1$ if and only if ϕ→Ψ is true in S. If $cerS(\phi,\Psi)=1$, then ϕ→Ψ will be called a certain decision rule; if $0< cerS(\phi,\Psi)<1$ the decision rule will be referred to as a uncertain decision rule. The number $\text{supp}s(\phi,\Psi)/card(U)$ will be called the strength of the decision rule ϕ→Ψ in S.

2.3 Financial Ratios Variables for Business Failure Analysis

A large number of financial ratios having the ability to predict the failure of business have thus been proposed in the literature. For example, Business failure research starts from Altman's pioneering work using multivariate discriminant analysis with a set of five financial ratios for distinguishing failed firms from non-failed firms [2]. He used multivariate discriminat analysis in business failure prediction, and combines various financial ratios into single indicator. Altman [2] selected 33 financial distress companies period form 1946 to 1966, and selected 33 healthy by matching industry type and organization size are similar. Using 22 financial ratios as explanatory variables to build multiple discriminant analysis model. Translating 22 explanatory variables into 5 independent factors and constructing discriminat function.

Courtis [9] identifies 79 financial ratios for analyzing corporate performance and structure, including 28 ratios useful for predicting various forms of corporate difficult, 34 additional ratios from contemporary textbook literature, and 17 ratios used in specific studies or organizations. To represent corporate financial phenomena which express the relationships between the ratios and the characteristics that they purport to reveal, these financial ratios are grouped into three categories: (1) profitability, (2) managerial performance, and (3) solvency. The most important financial ratios are in the solvency category (e.g. working capital/total assets, total debt/assets), and the next is in the profitability category, suggesting that the viability of a business depends on profit making to a large extent. In addition, the performance and survival of a business are influenced by several factors, including environment, national and international economic situations.

In business failure prediction studies using rough sets as McKee & Lensberg [20], workable models normally use 10 to 86 if–then decision rules with 2 to 5 variables (financial ratios). A total of 13 different financial variables are used in these models and rough set models with financial variables produce good predictive ability of business failure.

The 14 financial ratios considered cover all the categories suggested by previous studies[8], including (a) solvency: there are 4 ratios (current ratio; quick ratio; liabilities/assets ratio; times interest earned ratio), (b) managerial performance : there are 1 ratios (average collection turnover), (c) profitability : there are 5 ratios (return on total assets; return on shareholders' equity; operating income to paid-in capital; profit before tax to paid-in capital; earnings per share), (d) financial structure : there are 1 ratio (shareholder's equity/total assets ratio), and (e) cash flow : there are 3 ratios (cash flow ratio; cash flow adequacy ratio; cash flow reinvestment ratio).

3 Research design

The purpose of paper is constructing a business failure model including time series and cross section data. Using grey prediction to combine time series with cross section data in data process then rough set to classify sample data.

3.1 Sample collection and pre-processing

The failed companies are so defined by Taiwan Economic Journal (TEJ) database since it may have been cited as (1) Corporate Bankruptcy, (2) Reorganization Plan, (3) The cheque bounced, (4) To ask for help to free from distress, (5)Take over, (6)Opinions of Certified Public Accountant (CPA), (7) Net worth less than zero, (8) Brokerage, and (9)To suspend work because of some problem on financial turnover.

3.2 The data

Firms in the population that were announced as failed during the period from 1971 to 2008 are selected as the samples of the study and collected from TEJ database.

Except for missing and incomplete data, there are 76 companies that each one have 7 years long history data including 38 financial distress companies and 38 healthy one matching by industry type and firm size. Industries of companies include 16 general industries except for securities and financial industry. Financial industry was excluded because the variables we used are not adequate to indicate the state of financial industry.

3.3 Grey Prediction

In order to add a time series information to the model of business failure prediction, we use Visaul Basic Application (VBA) embed in Microsoft EXCEL 2003 to build GM (1,1) model of grey prediction. We computer prediction value of the fourth year, the fifth year, the sixth year and the seventh year history data respectively for announcement and comparing to history data.

3.4 Rough Set Theory

In this section we used ROSE2 Entropy Based method provided by ROSE2 to discrete continuous data to discreted data. There are two steps to implement:

Step1. Data discretization

The interval of discretization is shown as Table1.

Step2. Reducts computing

After discreting, we use Discernibility Matrix as table 2 shown.

Table 1 Data Discretization and Interval Code

Vaiables	Interval code				
	0	1	2	3	4
X_1	(-∞, -0.0334397)	<-0.0334397, 0.168262)	<0.168262, +∞)		
X_2	(-∞, -0.0317657)	<-0.0317657, 0.0606566)	<0.0606566, +∞)		
X_3	(-∞, -0.0283678)	<-0.0283678, 0.00548595)	<0.00548595, 0.0304666)	<0.0304666, 0.296736)	<0.296736, +∞)
X_4	(-∞, 0.000434135)	<0.000434135, 0.000766877)	<0.000766877, 0.0020475)	<0.0020475, +∞)	
X_5	(-∞, 0.453716)	<0.453716, +∞)			

Table 2 Discernibility Matrix

U	a	b	c	d
O1	0	1	2	0
O 2	1	2	0	2
O 3	1	0	1	0
O 4	2	1	0	1
O 5	1	1	0	2

Translation information table to 6×6 discernibility matrix as table 3 shown.

Table 3 6×6 Discernibility Matrix

	O1	O 2	O 3	O 4	O 5
O1					
O 2	a,b,c,d				
O 3	a,b,c	b,c,d			
O 4	a,c,d	a,b,d	a,b,c,d		
O 5	a,c,d	b	b,c,d	a,d	

In order to find minimun set of atrributes, then using discernibility function $f(a)$ to replace discernibility matrix:

$$f(a) = (a+b+c+d)(a+b+c)(a+c+d)(a+c+d)$$
$$\times (b+c+d)(a+b+d)d$$
$$\times (a+b+c+d)(b+c+d)$$
$$\times (a+d)$$

To calculate the final form of f(a) the absorption law of Boolean algebra is applied:

(1) $X + XY = X$

(2) $X \bullet (X+Y) = X$

According to the absorption law:

$$f(a) = (a+d)b = ab + bd$$

We can find reducts of this example are { a , b } and { b , d }.

3.5 Variables selection

There were lots of financial ratios in use and various ratios are different in prior research respectively. Then variables in Z-score provided by Altman [2] in 1968 are often been chosen as variables afterward [17].

As mentioned above, we use five variables provided to build business failure prediction model:

(1) Working Capital/Total Assets is defined as the difference between current assets and current liabilities. Liquidity and size characteristics are explicitly considered. Ordinarily, a firm experiencing consistent operating losses will have shrinking current assets in relation to total assets.

(2) Retained Earning/Total Assets is the ratio measures the leverage of a firm. Those firms with high RE, relative to TA, have financed their assets through retention of profits and have not utilized as much debt. This ratio highlights either the use of internally generated funds for growth (low risk capital) or OPM (other people's money) - higher risk capital.

(3) Earnings is before Interest and Taxes/Total Assets is a measure of the true productivity of the firm's assets, independent of any tax or leverage factors. Since a firm's ultimate existence is based on the earning power of its assets, this ratio appears to be particularly appropriate for studies dealing with credit risk.

(4) Market Value Equity/Book Value of Total Liabilities measure shows how much the firm's assets can decline in value (measured by market value of equity plus debt) before the liabilities exceed the assets and the firm becomes insolvent.

(5) Sales/Total Assets is a standard financial ratio illustrating the sales generating ability of the firm's assets. It is one measure of management's capacity in dealing with competitive conditions.

4 Result

Data experiment organized as three parts.

(1) Building business failure prediction model according to data sets were during from Y-1 to Y-4.

(2) To combine grey prediction and rough set for building model. Acquirement of next year predicted value as input sample is according to several year data by using grey prediction. And then put the predicted value into prediction model building by RST. Finally we can get sub-rules according to sub-reducts.

(3) To put the predicted value, that be calculated by grey prediction, into rough set business failure prediction model. The ratio of training and testing sample is 70:30. In other words, there were 54 data for training and 27 data for testing. Later sections will compare to each results.

The period of sample selection is from one to seven years before announcement. Choosing history data for 3 years respectively, and using GM(1,1) model building by the 4th, 5th, 6th, and 7th dimension time series data to generate predicted value about next year as input variable to business failure prediction model. Accuracy of both models will be compared to each other, and show which one being higher as table 4 shown.

Table 4 Classification Results of the Matrix for Business Failure Prediction Model

Year-1 in 4th dimension		prediction		
		Failure company	No-failure company	None
actuality	Failure company	9	2	0
	No-failure company	0	11	0
Correct rate： 90.91 %				

Year-1 in 5th dimension		prediction		
		Failure company	No-failure company	None
actuality	Failure company	8	3	0
	No-failure company	0	11	0
Correct rate：86.36 %				

Year-1 in 6th dimension		prediction		
		Failure company	No-failure company	None
actuality	Failure company	9	2	0
	No-failure company	2	9	0
Correct rate：81.82 %				

Year-1 in 7th dimension		prediction		
		Failure company	No-failure company	None
actuality	Failure company	9	2	0
	No-failure company	1	10	0
Correct rate：86.36 %				

Year-2 in 4th dimension		prediction		
		Failure company	No-failure company	None
actuality	Failure company	11	0	0
	No-failure company	2	9	0
Correct rate：90.91 %				

Year-2 in 5th dimension		prediction		
		Failure company	No-failure company	None
actuality	Failure company	10	1	0
	No-failure company	3	8	0
Correct rate：81.82 %				

Year-2 in 6th dimension		prediction		
		Failure company	No-failure company	None
actuality	Failure company	9	2	0
	No-failure company	4	7	0
Correct rate：72.73 %				

Year-3 in 4th dimension		prediction		
		Failure company	No-failure company	None
actuality	Failure company	10	1	0
	No-failure company	6	5	0
Correct rate：68.18 %				

Year-3 in 5th dimension		prediction		
		Failure company	No failure company	None
actuality	Failure company	11	0	0
	No-failure company	5	6	0
Correct rate：77.27 %				

As table 5 mentioned above, after including time series information the accuracy at Year-1, Year-2, and Year-3 of grey prediction business failure prediction models in the 4th and the 5th dimensions have better performance than history business failure prediction models. Specially, accuracy of grey prediction business failure prediction model in the 5th dimension has the best performance.

Accuracy of model in the 6th and the 7th dimensions are less than history model. There are two reasons of occurrence:

(1) The length of data collection is over the business cycle therefore the performance is not good enough although enhancing information.

(2) The other reason may be influenced by window dressing that made a fund appear more attractive.

The result indicates that comparison of the rough set approach with a hybrid of grey prediction and rough set analysis has better performance.

Table 5 Comparison of History and Prediction Model

	Year-1 correct rate	Year-2 correct rate	Year-3 correct rate
History value (Original Rough Set data)	77.27	63.64	59.09
Predicted value in 4th dimension	**81.82**	**77.27**	**63.64**
Predicted value in 5th dimension	**86.36**	**77.27**	**68.18**
Predicted value in 6th dimension	72.73	54.44	
Predicted value in 7th dimension	72.73		

5 Conclusion and Limitation

Summarily, compared to the original rough set method, the hybrid intelligent system of grey prediction and rough set approach offers the following advantages:

· It discovers some important facts hidden in data and expresses them in the natural language of decision rules.
· It can contribute to retrench the time and cost of the decision making process (rough set approach is an information processing system in real time).
· It offers transparency of classification decisions, allowing for their argumentation.
· It takes into account background knowledge of the decision maker.

We adapt Taiwan Stock Exchange listed company because of restricted by unlisted company financial information difficult to obtain. Another limitation of this study is the lack of data which needs seven years of the failure company. Also, the company's financial information may be due to artificial factors caused by Window Dressing. In addition, some activities on the financial impact can not be exposed on the company balance sheet such as high-risk activities, derivatives, foreign exchange hedge etc. Therefore, the results should be viewed with some caution.

References:

[1] Ahn, B. S., Cho, S. S., & Kim, C.Y., (2000). The integrated methodology of rough set theory and artificial neural network for business failure prediction, *Expert Systems with Application*, Vol.18, No.2, pp.65-74.

[2] Altman, E. I., (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, *Journal of Financial*, Vol.23, No.4, pp.589-609.

[3] An, A., Shan, N., Chan, C., Cercone, N, & Ziarko, W., (1996). Discovering rules for water demand prediction: an enhanced rough-set approach, *Engineering Applications of Artificial Intellgence*, Vol.9, No.6, pp.645-653.

[4] Beaver, W. H., (1966). Financial Ratios as Predictors of Failure, Empirical Research in Accounting: Selected Studies, *Journal of Accounting Research*, Vol.4, No.3, pp.71-111.

[5] Beynon, M., Curry, B., & Morgan, P., (2000). Classification and rule induction using rough set theory. *Expert Systems*, Vol.17, No.3, pp.136-147.

[6] Bose, I., (2006). Deciding the Financial Health of Dot-coms Using Rough Sets, *Information & Management*, Vol.43, pp.835-846.

[7] Chen, C. Y., (2008). A study of Forecasting Methods for the Cross-Straits Trade, *WSEAS TRANSACTIONS on BUSINESS and ECONOMICS*, Vol.5, No.5, pp.161-171.

[8] Cheng, J. H., Yeh, C. H. & Chiu, Y. W., (2007). Identifying significant indicators of business failure using rough sets, J*ournal of Harbin Institute of Technology (New Series)*, Vol.14, No.2, pp.42-47.

[9] Courtis, J. K., (1978). Modelling a Financial Ratios Categoric Framework, *Journal of Business Finance and Accounting*,Vol.5, pp.371-386.

[10] Deng, J., (1982). Control Problems for Grey System. *Systems and Control Letters*, Vol.1, pp. 288-294.

[11] Dash, M., & Liu, H., (1997). Feature selection for classification, *Intelligent Data Analysis*, Vol. 1, pp.131-156.

[12] Dimitras, A. I., Zanakis, S. H., & Zopounidis, C. (1996). A Survey of Business Failures with An Emphasis on Prediction Methods and Industrial Applications, *European Journal of Operational Research*, Vol.90, pp.487-513.

[13] Dimitras, A. I., Slowinski, R., Susmaga, R. & Zopounidis, C., (1999). Business Failure Prediction Using Rough Sets, *European Journal of Operational Research*, Vol.114, pp.263-280.

[14] Fan, A., & Palaniswami, M., (2001). Stock Selection using Support Vector Machines, *IEEE*, Vol.3, pp.1793-1798.

[15] Frydman, H., Altman, E. I. & Kao, D. L., (1985). Introducing Recursive Partitioning for Financial Classification□The Case of Financial Distress, *Journal of Finance*, Vol.40, pp.269-291.

[16] Hsieh, C. H., Chou, J. H. & Wu, Y. J., (2002). Optimal Predicted Fuzzy Controller of a Constant Turning Force System with Fixed Metal Removal Rate, *Journal of Materials Processing Technology*, Vol.123, pp.22-30.

[17] John, G. H., Kohavi, R., & Pfleger, K., (1994). Irrelevant features and the subset selection problem. In: Proceedings of the Eleventh International Conference on Machine Learning, pp.121-129.

[18] Kumar, P. R. & Ravi, V., (2007). Bankruptcy Prediction in Banks and Firms via Statistical and Intelligent Techniques - A review, *European Journal of Operational Research*, Vol.180, pp.1-28.

[19] Liu, H., & Setiono, R., (1996). A probabilistic approach to feature selection A lter solution. In: Kumar, S. (Ed.), 13th International Conference on Machine Learning (ICML_96), Bari, Italy, pp.319-327.
[20] McKee, T. E. & Lensberg, T., (2002), Genetic programming and rough sets: A hybrid approach to bankruptcy classification, *European Journal of Operational Research*, Vol.138, No.2, pp.436-451
[21] Moraczewski, I. R., Sudnik-Wojcikowska, B., & Borkowski, W., (1996). Rough sets in floristic description of the inner-city of Warsaw, FLORA 191, (3)253-260.
[22] Ohlson, J. A, (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy, *Journal of Accounting Research*, Vol.18, pp.109-131.
[23] Pal, S. K., & Skowron, A., (1999). Rough Fuzzy Hybridization: A New Trend In Decision Making. Springer, Singapore.
[24] Pawlak, Z., (1982). Rough Sets, *International Journal of Computer and Information Sciences*, Vol.11, pp.341-356.
[25] Pawlak, Z., (1991). Rough Sets—Theoretical Aspects of Reasoning about Data. Kluwer Academic, Dordrecht.
[26] Pawlak, Z. (1997). Rough sets. In T. Y. Lin, & N. Cercone (Eds.), Rough sets and data mining—Analysis for imprecise data (pp.3-8). Boston, MA: Kluwer.
[27] Ruggiero, M. A., (1997). Cybernetics Trading Strategies: Developing Profitable Trading Systems with State-of-the-Art Technologies, New York: Wiley.
[28] Skowron, A., (2000). Rough sets in KDD (plenary talk). In: Shi, Z., Faltings, B., Musem, M. (Eds.), Proceedings of Conference on Intelligent Information Processing (IIP2000), 16-th World Computer Congress (IFIP_2000), Beijing, pp.19-25 August, 2000. Publishing House of Electronic Industry, Beijing, pp.1-17.
[29] Slowinski, R. & Zopounidis, C., (1995). Application of the rough set approach to evaluation of bankruptcy risk. *International Journal of Intelligent Systems in Accounting, Finance and Management*, Vol.4, pp.27-41
[30] Wakulicz D. A., & Paszek, P., (1997). Diagnose progressive encephalopathy applying the rough set approach. *International Journal of Medical Informatics*, Vol.46, pp.119-127.
[31] Wang, Y. F., (2002). Predicting Stock Price Using Fuzzy Grey Prediction System, Expert Systems with Applications, Vol.22, pp.33-39.
[32] Xaio, H., & Zhang, X., Comparison Studies on Classification for Remote Sensing Image Based on Data Mining Method, *WSEAS TRANSACTIONS on COMPUTERS*, Vol.7, No.5, pp.552-558.
[33] Yu, C., & Luo, Y., (2008). Research on Models of Decision-making Problem Recognition, WSEAS TRANSACTIONS on BUSINESS and ECONOMICS, Vol.5, No.3, pp.37-43.
[34] Ziarko, W., (1993). Variable precision rough set model. *Journal of Computer and System Sciences*, Vol.46, pp.39-59.
[35] Zmijewski, M. E., (1984). Methodological issues related to the estimation of financial distress prediction models. *Studies on Current Econometric Issues in Accounting Research*, pp. 59-82.
[36] Zopounidis, C., Dimitras, A. I., & Le Rudulier, L., (1995). A multicriteria approach for the analysis and prediction of business failure in Greece. Cahier du LAMSADE, no. 132, Université de Paris Dauphine