# COMBINING CLUSTER ANALYSIS WITH ROUGH SET THEORY FOR FINANCIAL CRISIS PREDICTION OF ELECTRONICS INDUSTRY

WEI-WUN SYU

*Department of Banking and Finance, Kainan University, Kainan University , No. 1, Kainan Rd., Luchu, Taoyuan 338, Taiwan*

GWO-HSHIUNG TZENG

*Department of Business and Entrepreneurial Management, Kainan University, Kainan University , No. 1, Kainan Rd., Luchu, Taoyuan 338, Taiwan; Institute of Management of Technology, National Chiao Tung University, Ta-Hsueh Road, Hsinchu 300, Taiwan*

**Abstract**

This paper propose a hybrid intelligent system for prediction the failure of business based on the financial performance data combining cluster analysis with rough set technique. Most variables often can not satisfy the statistical assumptions in real world, so we propose a novel approach to predict financial distress prediction of electronics companies based in rough set theory (RST). This paper is regards to distinguish between failed and healthy in electronic companies, then how we focuses in failed companies to propose the improving strategies to raise profit for companies and satisfy the customers' needs. Therefore, in this paper we will distinguish three class of "maybe crisis", "common" and "best healthy" in electronic industry by different financial features/characteristics, and can obtain reducts and core by through rough set techniques. And then, this reduced information is used to infer classification rules form appropriate variables, and can be discussed and compared with distress companies. An empirical example of electronic industry of Taiwan is illustrated to demonstrate to infer condition rules can effective identify discrepancy in the groups of industry, due to there rules based on the experience of real examples are well so this the argumentation of the decisions we make. Therefore, this paper is proposing a prediction model based on rough set theory for Taiwan electronic companies for distress prediction.

**Keywords**: Business failure prediction, Financial crisis, Cluster Analysis, Rough Set Theory, Decision rules

## 1. *Introduction*

Evaluation of the business failure has been, for a long time, a major preoccupation of researchers. Companies of financial distress is a general problem of companies and, according to a widespread definition, is the situation that a firm cannot pay lenders, preferred stock shareholders, suppliers, etc., or a bill is overdrawn, or the firm is bankrupt. All these situations result in a discontinuity of the firm's operations, so that these companies occur crisis. The number of failing firms is an important indicator for the health of the economy and it can be considered as an index of the development and robustness of the economy. Clearly, failure affects a firm's entire existence and it has high cost to the firm, the collaborators (firms and organizations), the society and finally the country's economy (Warner, 1977). In order to avoid crisis, it needs understand the cause of financial crisis early, and these effect factor can be considered a symptom of financial crisis. The development and use of models, able to predict failure in advance, can be very important for the firms in two different ways. First, as "early warning systems", such models can be very useful for those (i.e. managers, authorities, etc.) who have to prevent failure. Second, such models can be useful in aiding decision-makers of financial institutions in charge of evaluation and selection of the firms.

In this paper, we propose a two-stage intelligent system combining cluster analysis and rough set approach. We use cluster techniques for surveying the various financial status of companies of electronics industry, which represent different factors of the most several constituents, so that this financial diagnostic on their status can be to distinguish and, besides, can be effective in(to) explained the difference between from the status of those discrepancy. In addition, we propose to predict financial crisis prediction model based on rough set theory which classifies companies into healthy and failed performs a selection among the financial ratio. A sample of Taiwan electronics companies is used and general financial ratios as well as those that are specifically proposed for predicting electronics industry are employed. The results in that can be to distinguish different groups for electronics industry certainly, where this paper infer to them which they have certainly common attributes in different pattern, such as failed and healthy firms or maybe crisis, common and best healthy patterns from healthy firms of electronics industry. And then, we can be found that a firm will to bankrupt if it has a lot of debts to on credit. To change another words, a firm needs to bear risks that can't be to pay. Other, there are three pattern were divided up from healthy of electric industry, where we can also found that operating expenses is a key factor. When firms want to invest new financial program or dilate factory…etc, they will increase operating expenses to improve or convert so that can increase nice "image" or raise company's income, but firms will be to bankrupt if operating expenses to overflow and no have abundant current capital to pay so that income to less more than minus.

According to the aforementioned, we can result in liabilities and operating expenses are important ratios extremely. They can discrimination different groups of electronic industry effective; and we infer some useful rules to make out different patterns of electronic industry by rough set theory. This paper is organized as follows. In Section 2 the financial crisis prediction models are mainly discussed in the past, this paragraph describes development of financial crisis, it relate to many researcher to propose models of financial crisis prediction. In Section 3 basic concept of rough set theory is described. In Section 4, a case of financial crisis prediction for decision-making is illustrated to demonstrate the proposed methods; and its proposed data preprocessing algorithm by rough set theory. In Section 5, we analyze and compare the results of each groups and conclusions.

## 2. *Financial crisis prediction*

Previous studies on financial crisis prediction have used financial ratios under the assumption that these variables are random variables. Multivariate statistics indicate significant differences between the average financial ratios of crisis and healthy firms. Financial ratios maybe informational representative for users and have been extensively used for crisis classification.

Most early, Altman (1968) introduced Multiple Discrtiminant Analysis (MDA) and the issue of selecting independent accounting variables for predicting financial crisis. Other researches have introduced the logit model as a probabilistic model of bankruptcy, to discuss different ratios, and to evaluate different probabilities for classification errors and predictive accuracy. Studies provide extensive discussions and detailed analyses of logit, probit, MDA, ordinary least square (OLS) regression, and other techniques of limited dependent variables.

On the above of describe, financial crisis prediction methods such as traditional statistical methods, multiple discriminant analysis, linear probability models, and logit and probit analysis have been mainly used for business classification problems (Altman, 1968; Altman, Haldeman & Narayanan, 1977; Collins & Green, 1972). Later, the development and application of artificial intelligence led some researchers to employ inductive learning and neural networks in business domain (Chung & Tam, 1992; Fletcher & Goss, 1993; Odom & Sharda, 1990; Raghupathi, Schkade & Raju, 1991; Salchenberger, Cinar & Lash, 1992; Tam & Kiang, 1992). Many other methods such as multiple criteria decision analysis (MCDA) and rough set approach have been successfully applied to real world classification problems (Siegel, de Korvin &

Omer, 1993; Slowinski & Zopounidis, 1995). In recent, Min & Lee (2005) use machine learning for business failureprediction. Specifically, they propose a methodology based on support vector machines (SVM) and they to compare with multiple discriminant analysis, logistic regression analysis and neural networks. It use a large set of 1888 cases and also use stepwise logistic regression as a feature selection technique, which gives 11 attributes to be used for further modeling of bankruptcy. Support vector machines outperform the other techniques and their performance lies between 71 and 83% for the holdout dataset. Besides, Wu et al. (2007) succeed to use a real-valued genetic algorithm to optimize the parameters based on SVM for predicting bankruptcy in non-linear problems; but these methods is difficult to show decision-maker by using if-then rule in logic reasoning, wherefore just have rough set theory to rise and develop.

Recently, Rough set method are the most popular tool used for financial crisis prediction and has been reported that its accuracy is superior to that of traditional statistical methods in dealing with financial crisis problems, especially in regards to nonlinear patterns. Rough set theory, introduced by Pawlak (1982) and Pawlak, Grzymala-Busse, Slowinski and Ziarko (1995) is a mathematical tool to deal with vagueness and uncertainty of information and proved to be an effective tool for the analysis of financial information system comprised of a set of objects described by a set of multi-valued financial ratios and qualitative variables. Zopounidis (1995) employed rough set approach in business failure prediction. They used 12 financial ratios and compared rough set approach with statistical approaches. Rough sets theory based model has the following advantages: (1) the rough sets data analysis process results in the information contained in a large number of cases being reduced to a model containing a generalized description of knowledge, (2) the model is a set of easily understandable decision rules which do not normally need interpretation, (3) each decision rule is supported by a set of real examples in world, (4) additional information like probabilities in statistics or grade of membership in fuzzy set theory is not required (Mckee, 2000).

## 3. *The basic concepts of Rough Sets*

### 3.1. *Basic concepts of Rough Sets Theory*

Often, information on the surrounding world is imprecise, incomplete or uncertain. This means that to draw conclusion of information, we need to have a way of thinking and concluding which can be process uncertain and/or incomplete information. Rough set theory was developed by Pawlak (1982, 1984, 2004). The philosophy of the method is based on the assumption that with every object some information (data) can be associated. Objects characterized by the same information are indiscernible in view of the available information. Rough set theory can deal with inexact, uncertain, and vague datasets (Walczak & Massart, 1999). Both fuzzy set theory and rough set theory are used with the indiscernibility relation and perceptible knowledge. The major difference between them is that rough set theory does not need a membership function, thus, it does not need tradition assumption, too.

The rough set methodology assume that is based lowering the degree of precision in the data makes the data pattern more visible (Slowinski, 1992), this approach can be considered as a formal framework for discovering facts from imperfect data.

In this section we briefly introduce rough set theory. In Section 3.1 basic concepts of rough set theory. In Section 3.2 is Information system and Section 3.3 is Approximation of sets. In Section 3.4 main describes reduction of attributes and core, Section 3.5 Decision tables.

### 3.2. *Information system*

By an information system, IS (or an approximation space), we understand the 4-tuple $IS = (U, A, V, f)$, where $U$ is a finite set of objects (the universe), $A$ is a finite set of attributes (features), $V = \bigcup_{a\in A} V_a$ and $V_a$ is a domain of the attribute $a$, and $f_a : U \times A \rightarrow V_a$ is a total function such that $f(x,a) \in V_a$ for every $a \in A$, $x \in U$, defines an information function, where $V_a$ is the set of values of $a$, called the domain of attribute $a$.

Let $IS = (U, A, V, f)$ be an information system and let $B \subseteq A$ and $x_i, x_j \in U$. *IND*(*B*) is defined in the following way, $x_i$ and $x_j$ are indiscernible by the set of attributes $B$ in $A$, if $b(x_i) = b(x_j)$ for every $b \subset B$. Thus every $B \subseteq A$ generates a binary relation on $U$ which will be called an indiscernibility relation, denoted by *IND*(*B*). Obviously, The equivalence class of *IND*(*B*) is called elementary set in B because it represents the smallest discernible groups of objects. For any element $x_i$ of $U$, the equivalence class of $x$ in relation *IND*(*B*) is represented as $[x_i]_{IND(B)}$.

### 3.3. *Approximation of sets*

The rough sets approach to data analysis hinges on two basic concepts, namely the *lower* and the *upper approximations* of a set, referring to the elements that doubtlessly belong to the set, and the elements that possibly belong to the set. Let $X$ denote the subset of elements of the universe $U$ ($X \subset U$). The lower approximation of $X$ in $B$ $(B \subseteq A)$, denoted as $\underline{BX}$, and the upper approximation of the set $X$, denoted as $BX$, more formally:

$$\underline{BX} = \left\{ x_i \in U \left| [x_i]_{Ind(B)} \subseteq X \right. \right\} \tag{1}$$

$$BX = \left\{ x_i \in U \left| [x_i]_{Ind(B)} \cap X \neq \phi \right. \right\} \tag{2}$$

The boundary of set $X$ in $U$ is defined as

$$BNX = BX - \underline{BX} \tag{3}$$

The set $BNX$ is the set of objects which cannot be certainly classified to $X$ using the set of attributes $B$ only. Inexactness of a set is due to the existence of the boundary. The greater the doubtful region of a set, the lower the accuracy of that set, and an accuracy measure of the set $X$ in $B \subseteq A$ is defined as:

$$u_B(X) = card(\underline{BX}) / card(BX) \tag{4}$$

This ratio expresses the percentage of possible correct decisions when classifying objects employing knowledge available. Therefore, using the lower and the upper approximation we can define those subsets that cannot be expressed exactly using the available attributes- precisely.

Because we are interested in classifications in addition, it expresses the percentage of objects which can be correctly classified to classes employing the knowledge available. The quality of classification is defined as:

$$\gamma_B(F) = \sum_{i=1}^{n} card(\underline{BX_i}) / card(U) \tag{5}$$

### 3.4. ***Reduction of attributes and core***

The concepts of core and reduct are two fundamental concepts of the rough sets theory. If the set of attributes is dependent, one want to discussed in all possible minimal subsets of attributes, ، which lead to the same number of elementary sets as the whole set of attributes: reducts and the core is common to all them (reducts).

We say that the set of attributes $R \subseteq A$ depends on the set of attributes $B \subseteq A$ in *IS* iff $IND(B) \subseteq IND(R)$. Discovering dependencies between attributes is of primary importance in the rough set approach to knowledge analysis. Another important issue is that of attribute reduction, in such a way that the reduced set of attributes provides the same quality of sorting as the original set of attributes.

### 3.5. *Decision tables*

An information system can be seen as a decision table assuming that $A$= $C \cup D$ and $C \cap D = \phi$; where *C* are called condition attributes, and *D*, decision attributes. Decision table $IS = \{U, C \cup D, V, f\}$ is deterministic iff $C \rightarrow D$; otherwise it is non-deterministic. The deterministic decision table uniquely describes the decisions to be made when some conditions are satisfied. In the case of a non-deterministic table, decisions are not uniquely determined by the conditions. Instead, a subset of decisions is defined which could be taken under circumstances determined by conditions.

From the decision table, a set of decision rules can be derived. Let *IND*(*C*) be a family of all *C*-elementary sets called condition classes, denoted by $X_i$ (*i* =1,2,…,*k*; where *k* is the number of *IND*(*C*)) Let, moreover, *IND*(*D*) be the family of all *D*-elementary sets called decision classes, denoted by $Y_j$ (*j* =1,2,…, *n*; where *n* is the number of *IND*(*D*)).

$Des_c(X_i) \Rightarrow Des_D(Y_j)$ is called the (*C*,*D*)- decision rule. The rules are logical statements "if…then…" relating descriptions of condition and decision classes. The set of decision rules for each decision class $Y_j$ $(j = 1, 2, ..., n)$ is denoted by $\{r_{ij}\}$. More precisely, $\{r_{ij} = Des_c(X_i) \Rightarrow Des_D(Y_j): X_i \cap Y_i = \phi, i = 1, 2, ..., k\}$, rule $\{r_{ij}\}$ is deterministic iff $X_i \neq X_j$; and $r_{ij}$ is non-deterministic otherwise.

## 4. *Empirical study: a case of financial distress prediction decision - making*

Evaluation of the business failure has been, for a long time, a major preoccupation of researchers. Companies of financial distress is a general problem of companies and, according to a widespread definition, is the situation that a firm cannot pay lenders, preferred stock shareholders, suppliers, etc., or a bill is overdrawn, or the firm is bankrupt. All these situations result in a discontinuity of the firm's operations, so that these companies occurred crisis. In order to avoid crisis, these firms need understand the cause of financial crisis early, and they can be considered a symptom of financial crisis.

Most methods applied in the past to predict business failure are techniques of statistical nature and use financial ratios as explicative variables. These variables do not normally satisfy statistical assumptions so we propose an approach to predict electronic corporations based on rough set theory. Therefore, in this section we apply the financial information of companies using rough set theory to explore the classification problem.

### 4.1. *Problem descriptions*

Now, many corporation was confronted with financial crisis on situation of possible, especially in real world exist many uncertain element. Therefore, we search out attributes to recognize as pattern disagree, and construct a method that can be effective predict financial crisis, then we focuses on failed companies to propose the improving strategies to raise profit for companies. In addition, the change in administrative policy allowed financial crisis companies to increase competitiveness in the electronics industry.

The successful corporation does not only want to see early crisis portent from financial information/data, but also discovers corporation strategies and/or improve it. In this paper, we give a series of ratios designed to find information of quarterly reports, such as the return & income, short-term liquidity, resources utilization ratio, grow ratio and capital structure.

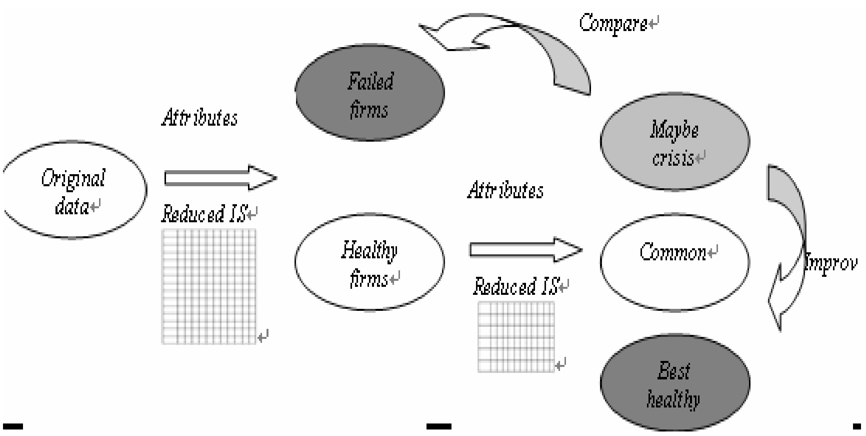


Figure 1. Concept plot

### 4.2. *Research data and selection of variables*

The work presented in this paper used a data by electronics industry from financial information of quarterly reports. The data comprises from 119 Taiwan electronic companies from the year 2001 to the year 2006, where 30 companies had been failed and 89 companies was healthy. The data includes financial information such as liquidity, solvency, profit margin and resources utilization etc. These variables are the inputs to the classification, the desired output of the classification is the variable that states if the companies were bankrupt or not.

The Table 1 of financial features table list below. According to many research points out some important variables in the past, these will affect companies result in financial distress. Therefore, this paper uses them to discuss in this research area, such as: Liquidity, Capital structure, Resources utilization ratio, Return & Income, Cash flow, Growth.

Table 1 Features of financial crisis prediction

| Financial features | Related items |
|---|---|
| Liquidity | Current ratio (%); Acid Test (%) |
| Capital structure | Liabilities (%); Times Interest Earne; Debt / Eqity (%); (L-T Liab.+SE) / FA (%) |
| Resources utilization ratio | A/R&N/R Turnover; Total Asset Turnover; Inventory Turnover; Equity Turnover |
| Return & Income | Gross Margin (%); Operating Exp. (%); Return on Assets (%); Net Income (%) |
| Cash flow | CFO / CL (%) |
| Growth | YOY(%)Gross Margin; YOY(%) Oper. Income; YOY(%) Pre-Tax Income |

CFO / CL (%): Cash flow ratio; Operating Exp. (%): Operating expenses (%); YOY(%) Gross Margin: Gross Margin Growth ratio; YOY(%) Oper. Income: Operating Income Growth ratio; YOY(%) Pre-Tax Income: Preceding of Tax Income Growth ratio

### 4.3. *Factor analysis and cluster analysis*

This section is to discuss variables of financial distress prediction to selection. It work has been acquired several financial features by Factor analysis, and it is extraction method of principal component analysis, where each Eigen value extract over 1 and it is from great to little order. There are five main components are composes of 12 financial features in Rotated Component Matrix, and they are return & income, liquidity, resources utilization ratio, growth, solvency. Therefore, the first component’s percentage of variance is greatest, others as well. The original Rotated Component Matrix is detailed in Table A.1 of Appendix A.

According to priority all of the above, we has acquired five important components and will them to discriminate several different groups by cluster analysis, this five components are respectively as ***Return & Income, Liquidity, Resources Utilization ratio, Growth*** and ***Solvency***, where ***Return & Income*** is main comprises Return on Assets(%), CFO/CL, Net Income(%) and Gross Margin(%), and can see that Return on Assets is most important components (it's factor loading is 0.854); ***Liquidity*** comprises Current ratio(%) and Acid Test(%), both they are important components (they're factor loadings are 0.949 & 0.946); and then, Operating Exp.(%) and Equity Turnover(%) are constitute ***Resources Utilization ratio,*** their effect are the reverse direction (their factor loadings are 0.862 & -0.745), and Equity Turnover(%) can understand that a per equity could created how much sales revenue are, this is mean that when it's high more investors will to gain profit more; ***Growth*** comprises YOY(%) - Oper. Income and YOY(%) - Pre-Tax Income (their factor loadings are 0.851 & 0.666); finally, ***Solvency*** is compose of Debt/Equity and Liabilities, and they are all important variables that use to measure degree of financial leverage (DFL), they could measured a firm have liquidation ability or not (their factor loading are 0.927 & 0.800). Other we are also sees that Current ratio(%) and Acid Test(%) are 0.967 & 0.968 the most high, and cumulative% of variance is 84.018.

this paper researches to obtain three groups of maybe crisis, common and best healthy firms, where common pattern in healthy firms has 73.3% market share ratio most, and market share ratios of maybe crisis firms was 18% more high than its best healthy of firms 6.7% in electronic industry. The original companies of electric industry in the market obtain three groups by Cluster analysis described in Table A.2

In addition, the component of radar plot made us to discover different characteristics between each groups certainly, best healthy firms higher obviously than others in the short – term liquidity. On one hand, maybe crisis firms has higher in resources utilization ratio and capital structure; on the other hand, it has lower return ＆ income and grow ratio relatively. Wherefore, best healthy firms and maybe crisis firms are difference mostly in Fig 2.

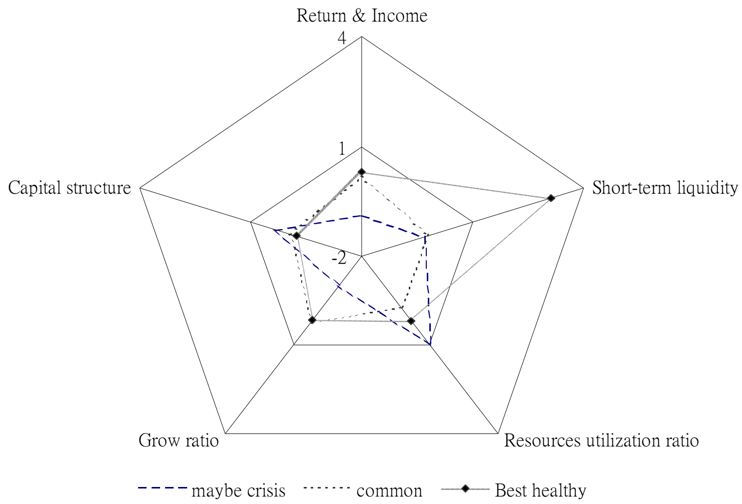


Figure 2. Five component of radar plot

### 4.4. *Rough sets approach and empirical results*

The information system for season 5 which consisted of 119 firms described with 7 ratios was entered into input file in ROSE. First step, we have made was to recode the continuous variables into qualitative terms (very low, low, medium, high, very high) with corresponding numeric values 0, 1, 2, 3 and 4. We have decided to recode the information table using 5 subintervals based on the actual mark ratio average values $\overline{X} = \{\overline{x}_1, \overline{x}_2, ..., \overline{x}_{12}\}$ of $\pm$ 1/2 $S$ ( $S$ shows a standard deviation set, $S = \{s_1, s_2, ..., s_{12}\}$) and $\pm$ 3/2 $S$ of the whole sample.

#### 4.4.1. *Empirical results – healthy & failed firms*

Table 5 shows the results after rough set analysis was performed in five seasons. As we can see, in 5 experiments we obtained one minimal reducts { $a_1$ $a_3$ , $a_5$ , $a_7$ , $a_8$ , $a_{10}$ , $a_{11}$ }, as: ***Current ratio(%), Liabilities (%), Equity Turnover, Operating Exp.(%), Return on Assets(%), CFO / CL(%)*** and ***YOY(%) - Oper. Income.*** This result mean that the reducts holds fewer attributes but ensuring the same value of the quality of approximation as the whole set of attributes, which 7 attributes are redundant and they could be eliminated. Consequently, this result shows the strong support of this approach in feature selection. This indicates that these variables of financial ratios are highly discriminatory between failed and healthy firms in our sample.

According to all of the above, we can to acquire that the reducts should have a small number of attributes as possible, and it should have the most significant attributes to present in our opinion for the evaluation of the companies; besides, after having selected a few reducts containing the most significant attributes. Therefore, in this test that we discover above seven features are important for financial crisis prediction of area and the test discovers quality of classification to present all is high more, where first experiment is most high.

The accuracy of the approximation for the two decision classes is shown in Table 2. The results indicate good accuracies for different classes. In general, high values for the quality of classification and accuracies mean that the attributes selected are adequate for approximating the classification. Recall that the rating of corporation crisis or no takes on the classification of H = healthy, F = failed. As shown in Table 2, the experiment #1 of accuracy of approximation is 1, other are 84.87%, 88.24%, 89.08% and 82.35%.

Table 2 Results after rough set data analysis (H, healthy; F, failed)

| Experiment # | Minimal reduct | Lower Approximation | | Upper Approximation | | Accuracy | | Quality of classification |
|---|---|---|---|---|---|---|---|---|
| | | H | F | H | F | H | F | |
| 1 | | 89 | 30 | 89 | 30 | 1.00% | 1.00% | 1.00% |
| 2 | | 77 | 24 | 95 | 42 | 81.5% | 57.14% | 84.87% |
| 3 | $\{a_1, a_3, a_5, a_7, a_8, a_{10}, a_{11}\}$ | 82 | 23 | 96 | 37 | 85.42% | 62.16% | 88.24% |
| 4 | | 82 | 24 | 95 | 37 | 86.32% | 64.86% | 89.08% |
| 5 | | 79 | 19 | 100 | 40 | 79.00% | 47.50% | 82.35% |

The program generates a set of "*if… then…*"decision rules called *minimal covering rules*. If they are complete and non-redundant, the decision rules are minimal. By complete we mean that the set of rules cover all the objects or respondents in the data set. By non-redundant, we mean that there are no other rules with an antecedent of at least the same weakness and a consequent of at least the same strength. Thus, the exclusion of any one rule makes the remaining rules non-complete.

Table 3 shows classification rules of case that we have obtained several rules, all of them are deterministic because the quality of the classification is equal to 1 and this means that the doubtful region is empty, so all the firms are highly discriminated among them, as follow:

Table 3 Classification rules (1, Healthy; 2, Failed)

| Rule# | Conditions | Decision | Strength% |
|---|---|---|---|
| 1 | (Current %=>2) & (Operating Exp.%=>1) & (CFO / CL %=>2) | 1 | 13.48 |
| 2 | (Liabilities %=>2) & (Operating Exp.%=>2) | 1 | 22.47 |
| 3 | (Liabilities %=>1) & ( CFO / CL %=>2) | 1 | 20.22 |
| 4 | (Current %=>2) & (Equity Turnover=>3) | 1 | 12.36 |
| 5 | (Liabilities %=>1) & (Return on Assets %=>2) | 1 | 20.22 |
| 6 | (Return on Assets %=>3) | 1 | 26.97 |
| 7 | (CFO/CL %=>4) | 1 | 11.24 |
| 8 | (YOY% Oper. Income=>3) | 1 | 12.36 |
| 9 | (Current %=>1) & (Liabilities %=> 3) & (Return on Assets %=> 2) & (YOY% Oper. Income=>2) | 2 | 16.67 |
| 10 | (Liabilities %=>4) & (CFO / CL=>2) | 2 | 23.33 |
| 11 | (Liabilities %=>2) & (Operating Exp.%=>3) | 2 | 6.67 |
| 12 | (Liabilities %=>3) & (Return on Assets %=>0) | 2 | 16.67 |
| 13 | (Operating Exp.%=>4) & (Return on Assets %=>0) | 2 | 13.33 |
| 14 | (Liabilities %=>3) & (Equity Turnover=>2) & (YOY% Oper. Income=>1) | 2 | 6.67 |
| 15 | (Return on Assets %=>1) & ( CFO / CL %=>1) | 2 | 6.67 |
| 16 | (Liabilities %=>3) & (Equity Turnover=>2) & (Return on Assets %=> 2) & (YOY% Oper. Income=>2) | 2 | 16.67 |

As we have mentioned earlier, on the below percentage of strength all over 20%, Table 4 converts the original rules into a meaningful explanation them using the strength rate (%), where Rule 2, 3, 5, 6 are all equivalent to a rating of healthy. Rule 2 (percentage of strength 22.47%) states that if corporation of liabilities(%) $\leq 2$ and operating exp(%) $\leq 2$, then corporation will be rated as healthy; Rule 3 and Rule 5 (all the percentage of strength 22.22%) state that if corporations of liabilities(%) $\leq 1$ and CFO/CL(%) $\geq 2$ or return on assets(%) $\geq 2$, then they will also be rated as healthy; in addition, Rule 6 state that if corporation of return on assets(%) $\geq 3$, then corporation will be rated as healthy. This means that nearly whole of the companies which rated the corporation's financial quality as healthy did so because they have several characteristic can be described; they identify liabilities(%), operating exp.(%), return on assets(%) and CFO/CL(%) as essential financial ratios.

Table 4 Classification rules with cover 20% strength (1, Healthy; 2, Failed)

| Rules | Current ratio (%) | Liabilities (%) | Equity Turnover | Operating Exp. (%) | Return on Assets (%) | CFO / CL (%) | YOY-Oper. Income | Strength % |
|---|---|---|---|---|---|---|---|---|
| 2 | | 2 | | 2 | | | | 22.47 |
| 3 | | 1 | | | | 2 | | 20.22 |
| 5 | | 1 | | | 2 | | | 20.22 |
| 6 | | | | | 3 | | | 26.97 |
| 18 | | 4 | | | | 2 | | 23.33 |

However, we can find out that liabilities (%) is common feature in Fig .3, if liabilities(%) of corporation $\leq$ 1or2, then as healthy firms; if liabilities(%) of corporation $\geq$ 4, then as failed firms. Thus it is can to differentiate between healthy and failed firms. Besides, this means that current ratio(%), equity turnover and YOY(%)-oper. Income are influence to low.

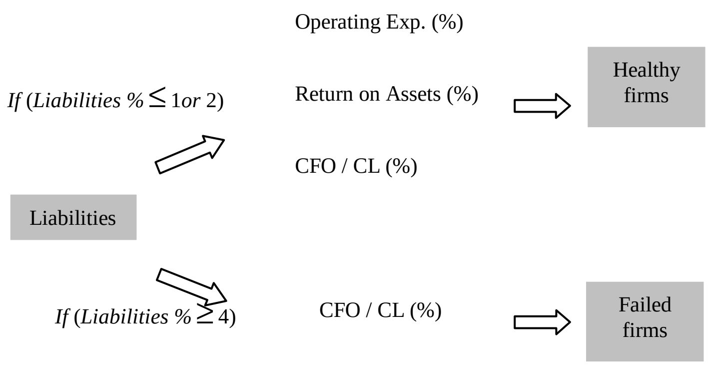


Figure 3. Line plot ( Healthy and Failed firms)

In additional, a more reliable approach is to use a resampling method, such as *k-fold cross validation*, *Double cross validation* or *LOOCV*, which main to consider training set can accepted or not, were *k-fold cross validation* could used samples to less so that increase credibility. In the test, we used the special case of the *k-fold cross-validation method* with *k*: 12. With 199 test cases, number of fold 10 and repetitions were used. In each repetition (199-(199/10)) of samples were used for the training set and (199/10) of samples as a holdout set for testing. Holdout sets were selected so that their union over all repetitions was the entire training set. In this way, every case was to participate in training and testing certainly.

This paper discovers better accuracy of each seasons tested with healthy and failed firms' holdout sample, we can see first season outperform the others. Wherefore, accuracy of classifying failed and healthy firms mainly are all higher more. (in Table A.3)

### 4.4.2. ***Empirical results – maybe crisis, common and best healthy firms***

Table 5 shows the results with "maybe crisis"," common" and"best healthy" firms, and let minimal reducts { $a_1$ $a_3$ , $a_5$ , $a_7$ , $a_8$ , $a_{10}$ , $a_{11}$ } into data of healthy firms. The accuracy of the approximation for the three decision classes is shown in follow. The results indicate good accuracies for different classes. In general, high values for the quality of classification and accuracies mean that the attributes selected are adequate for approximating the classification. Recall that the rating of corporation crisis or no takes on the classification of *M* = maybe crisis, *C* = common and *B* = best healthy. As shown in Table 5, the experiment #1 of accuracy of approximation is 96.63%, other are 95.51%, 89.89%, 89.89% and 89.89%.

Table 5 Results after rough set data analysis (*M*, Maybe crisis; *C*, Common; *B*,Best healthy firms)

| Experiment # | Lower Approximation | | | Upper Approximation | | | Accuracy (%) | | | Quality of classification |
|---|---|---|---|---|---|---|---|---|---|---|
| | *M* | *C* | *B* | *M* | *C* | *B* | *M* | *C* | *B* | |
| 1 | 15 | 65 | 6 | 18 | 68 | 6 | 83.33 | 95.59 | 100 | 96.63% |
| 2 | 14 | 65 | 6 | 18 | 69 | 6 | 77.89 | 94.20 | 100 | 95.51% |
| 3 | 12 | 62 | 6 | 21 | 71 | 6 | 57.14 | 87.32 | 100 | 89.89% |
| 4 | 12 | 62 | 6 | 21 | 71 | 6 | 57.14 | 87.32 | 100 | 89.89% |
| 5 | 12 | 62 | 6 | 21 | 71 | 6 | 57.14 | 87.32 | 100 | 89.89% |

Base on Table 5, we can see classification rules of case that we have obtained several rules, all of them are deterministic because the quality of the classification is equal to 96.63 and this means that the doubtful region is empty, so all the firms are highly discriminated among them, Table 6 as follow:

Table 6 Classification rules (1, Maybe crisis; 2, Common; 3,Best healthy firms)

| Rule# | Conditions | Decision | Strength |
|---|---|---|---|
| 1 | (Operating Exp.%=>2) & (YOY% Oper. Income=>1) | 1 | 25.00% |
| 2 | (Equity Turnover=>1) & (Operating Exp.%=>3) & (Return on Assets %=>2) & (YOY% Oper. Income=>2) | 1 | 12.50% |
| 3 | (Return on Assets %=>1) | 1 | 31.25% |
| 4 | (Liabilities %=>2) & (Equity Turnover=>1) & (CFO/CL %=>1) | 1 | 12.50% |
| 5 | (Liabilities %=>3) & (YOY% Oper. Income=>3) | 1 | 6.25% |
| 6 | (Liabilities %=>1) & (Equity Turnover=>2) | 2 | 17.91% |
| 7 | (Operating Exp.%=>1) & (YOY% Oper. Income=>2) | 2 | 34.33% |
| 8 | (Current %=>1) & (Equity Turnover=>1) & (Operating Exp.%=>2) & (YOY% Oper. Income=>2) | 2 | 8.96% |
| 9 | (Current %=>2) & (Return on Assets %=>3) | 2 | 20.90% |
| 10 | (Current %=>3) & (Operating Exp.%=>2) | 2 | 8.96% |
| 11 | (Equity Turnover=>3) | 2 | 19.40% |
| 12 | (Current %=>2) & (Liabilities %=>2) & (Equity Turnover=>1) & (YOY% Oper. Income=>2) | 2 | 4.48% |
| 13 | (CFO / CL %=>3) | 2 | 13.43% |
| 14 | (Current %=>2) & (YOY% Oper. Income=>3) | 2 | 10.45% |
| 15 | (YOY% Oper. Income=>4) | 2 | 1.49% |
| 16 | (Equity Turnover=>4) | 2 | 8.96% |
| 17 | (Liabilities %=>1) & (Return on Assets %=>4) | 2 | 1.49% |
| 18 | (Liabilities %=>2) & (YOY% Oper. Income=>3) | 2 | 8.96% |
| 19 | (Current %=>4) & (Operating Exp.%=>2) | 3 | 83.33% |
| 20 | (CFO / CL %=>0) & (YOY% Oper. Income=>1) | 3 | 16.67% |

As we have mentioned earlier, on the above of Table 7 percentage of strength all over 20%, where Rule 1, 3, 8, 10 and 20 are all equivalent to a rating of healthy. Rule 3 (percentage of strength 25.00%) states that if corporation of operating exp.(%)$\geq$2 and oper. income(%)$\leq$1, then corporation will be rated as maybe crisis, and Rule 3 (all the percentage of strength 31.25%) state that if corporations of return on assets(%)$\leq$1, then they will also be rated as maybe crisis; in addition; Rule 8 (the percentage of strength 34.33%) state that if corporation of operating exp.(%)$\leq$1 and oper.income(%)$\geq$2, and Rule 10 (all the percentage of strength 20.90%) state that if corporations of current ratio(%)$\geq$2 and return on assets(%)$\geq$3, then they will all be rated as common;

in addition, Rule 20 (the percentage of strength 83.33%)state that if corporation of current ratio(%)≥4 and operating Exp.(%)≥2, then corporation will be rated as best healthy.

Fig 6 This means that nearly whole of the companies which rated the corporation's financial quality as healthy did so because they have several characteristic can be described; they identify current ratio(%), operating exp.(%), return on assets(%) and YOY(%)-oper income as essential financial ratios.

Table 7 Classification rules with cover 20% strength (1, Maybe crisis; 2, Common; 3,Best healthy firms)

| Rule | Current ratio (%) | Liabilities (%) | Equity Turnover | Operating Exp. (%) | Return on Assets (%) | CFO / CL (%) | YOY-Oper. Income | % |
|---|---|---|---|---|---|---|---|---|
| 1 | | | | 2 | | | 1 | 25.00 % |
| 3 | | | | | 1 | | | 31.25 % |
| 8 | | | | 1 | | | 2 | 34.33 % |
| 10 | 2 | | | | 3 | | | 20.90 % |
| 20 | 4 | | | 2 | | | | 83.33 % |

However, we can find out that operating exp.(%) is common feature, if operating exp.(%) of corporation≤1or2, then as best healthy and common firms; if operating exp.(%) of corporation≥2, then as maybe crisis firms. Thus it is can to differentiate between healthy and failed firms. Besides, this means that liabilities(%), equity turnover and CFO/CL(%) are influence to low.

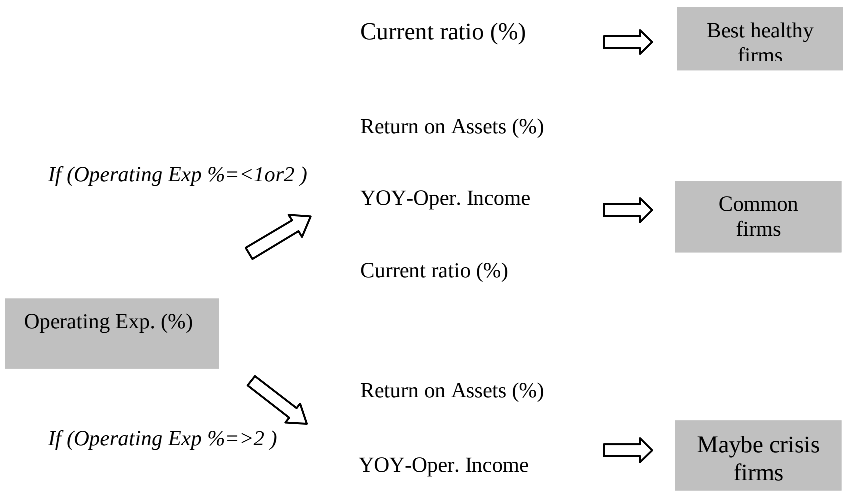


Figure 4. Line plot ( Maybe crisis, Common andBest healthy firms)

### 4.5. *Discussions*

We have presented a new approach to financial crisis prediction using rough sets. Through the exposition we have mentioned some advantages of this approach so we can conclude that this method is an effective tool for supporting managerial decision making in general, and for deal with financial crisis to occur, in particular.

In the light of the experiments carried out, this method is a effective tool that is existing bankruptcy prediction models in corporation and have great potential capacities that undoubtedly make it attractive for application to the field of business classification. Besides, these empirical results show that rough set model offers better predictive accuracy than the discriminant one we have given. And it neither requires the pre-specification of a functional form, nor the adoption of restrictive assumptions about the characteristics of statistical distributions of the variables and errors of the model. Consequently, for some real-world problems, the method we have presented is more attractive showing that it is a very robust technique especially in the areas of forecasting and classification decision problems.

We can to get some information by financial crisis predict model that regarding to difference between failed and healthy firms, the main attribute that liabilities can effective to discriminant and it is common attribute with failed and healthy firms, when the liabilities(%) is higher, firms are not easy to have crisis; on the other hand, if the liabilities is lower, firms are occur crisis easily. This condition indicates a firm when liabilities(%) are rising that it needs have huge flow cash, operating exp.(%) and return on assets(%) are all influence. Because a firm had borrowed huge debt, it needs paid huge debt interest lead to income decrease; therefore return on assets(%) is down. And then, we can be found that a firm will to bankrupt if it has a lot of debts to on credit. To change another words, a firm needs to bear risks that can't be to pay. On the other hand, we discuss healthy firms of have 3 groups, include maybe crisis, common and best healthy, where can discover that they have also common attribute, it is operating exp.(%). The attribute can effective to discriminant with this three groups, when the operating exp.(%) is lower, firms are not easy to have crisis; on the other hand, firms are occur easily if the operating exp.(%) is higher.

According to on above to describe that can to understand maybe crisis firms are different from others, how maybe crisis firms will to improve and how to prevent financial crisis to occur, maybe crisis firms will how to do that is important topic. Because there are lower profits for maybe crisis firms cause to oper. income(%) is the lower more. Therefore, when profits are very lower, return on assets(%) is lower more. On the above we can get to maybe crisis firms different others, and it can to compare with had bankrupted firms. When firms are not to have profits, they have not huge flow cash certainly. Therefore, when firms aren't to have huge money, they aren't investing any useful programs, and they needs to raise capital to do, they needs to shoulder interest of debts. In another word, when electronics corporation want to invest new financial program or dilate factory…etc, they will increase operating exp.(%) to improve or convert so that can increase nice "image" or raise company's income, but firms will be to bankrupt if operating exp. (%) to overflow and no have abundant current capital to pay so that income to less more than minus. It is can also measures have abundant capital or profits to pay or not by Return on assets and YOY(%)-oper income.

Finally, this paper main proposes traditional statistical method and rough set method apply to predict problems of financial crisis corporation, and compare rough set method with traditional statistical method, we can certainly discover to several similarity. We can easily to illustrated to demonstrate on the above of way by in Fig.8. First, we found out current ratio(%) of have evident divergence from failed and healthy firms in Fig.8 (a), this is mean that a healthy firm needs have huge capital and liquidity quickly (it's mean needs plenty free money to deal with any risks may be occur). We can also discriminate failed firms have highest debts of in Fig.8. (b), where Debt/Equity(%) represents each per equity needs to bear how much debt has, if it's higher that represents was more risks of bankruptcy on each investor’s shoulders. In another word, Liabilities(%) can also to represent between asset and debt relationship of configuration management. We can also see obviously that failed firms have higher operating exp.(%) in Fig.8. (d), where common firms have stably. And then, Fig.8. (e) describes that healthy firms have plus return on assets and failed firms have minus return on assets, to mean healthy firms’ assets can to create plus benefit. Certainly, it’s not easy occur bankruptcy if a firms have huge net cash flow. For this reason, healthy firms have ample net cash flow in Fig.8. (f). Finally, a firms’ growth ability can also shows to whether a firm have make profit ability, we can see that firms will to maybe occur bankruptcy when it has negative growth of profits in Fig.8. (g).

According to on the above demonstration to result in both traditional statistical method and rough set method can distinguish certainly between failed and healthy firms of electronics industry, where they can found out reason of financial crisis in time and predict bankruptcy occur to ahead of time by five financial components.

## 5. *Conclusions*

This paper proposes a hybrid intelligent system for predicting the failure of corporation based on the financial performance data combining cluster analysis with rough set technique. Through the exposition we have mentioned some advantages of this approach so we can conclude that this method is an effective tool for supporting managerial decision making in general, and for deal with financial crisis to occur, in particular.

In the light of the experiments carried out, this method is a effective tool that is existing bankruptcy prediction models in corporation and have great potential capacities that undoubtedly make it attractive for application to the field of business classification. Besides, these empirical results show that rough set model offers better predictive accuracy than the discriminant one we have gaved.

The decision rules generated can be used to companies to examine more thoroughly, quickly and inexpensively, therefore, management can solve financial problem in time efficiently. They can also be used to check and monitor corporation as a “warning system” for investors, management, financial analysts, banks, auditors, policy holders and consumers.

We know the model obtained has some problems and limitations but in spite of them, our objective is to show the suitability of this methodology as a support decision method for corporate crisis prediction. In short, we believe that rough set method, without replacing analyst's opinion and in combination with other methods, whether statistical or otherwise, will play a bright role in the decision making process in financial crisis predict model of electronic corporations.

## Appendix A

The original variables specification by Factor analysis described in Table A.1

Table A.1
Rotated Component Matrix

| Component | Return & Income | Liquidity | Resources utilization ratio | Growth | Solvency | Communility |
|---|---|---|---|---|---|---|
| Return on Assets (%) | ***0.854*** | -0.005 | -0.181 | 0.193 | -0.174 | 0.829 |
| CFO/CL (%) | ***0.830*** | 0.058 | 0.240 | -0.054 | -0.131 | 0.770 |
| Net Income (%) | ***0.731*** | 0.375 | -0.234 | 0.322 | 0.077 | 0.840 |
| Gross Margin (%) | ***0.617*** | 0.298 | 0.364 | 0.495 | -0.045 | 0.849 |
| Current ratio (%) | 0.096 | ***0.949*** | 0.110 | 0.024 | -0.210 | 0.967 |
| Acid Test (%) | 0.145 | ***0.946*** | 0.119 | 0.042 | -0.189 | 0.968 |
| Operating Exp. (%) | -0.187 | 0.089 | ***0.862*** | 0.288 | -0.021 | 0.869 |
| Equity Turnover (time) | -0.182 | -0.143 | ***-0.745*** | 0.179 | 0.163 | 0.668 |
| YOY(%) - Oper. income | 0.132 | -0.093 | 0.238 | ***0.851*** | -0.019 | 0.807 |
| YOY(%) - Pre -Tax Income | 0.190 | 0.168 | -0.333 | ***0.666*** | -0.179 | 0.651 |
| Debt/Equity (%) | -0.095 | -0.144 | 0.035 | -0.170 | ***0.927*** | 0.919 |
| Liabilities (%) | -0.163 | -0.376 | -0.371 | 0.047 | ***0.800*** | 0.948 |
| Eigenvalue | 2.520 | 2.255 | 1.908 | 1.704 | 1.695 | - |
| Contri. Rate | 20.997 | 18.795 | 15.899 | 14.202 | 14.125 | - |
| Accu.Con.Rate | 20.997 | 39.792 | 55.691 | 69.893 | ***84.018*** | - |

The original companies of electric industry in the market obtain three groups by Cluster analysis described in Table A.2

Table A.2
Cluster analysis data

| Failed firms | Healthy firms | | |
|---|---|---|---|
| had failed | maybe crisis | common | best healthy |
| 30 | 16 | 67 | 6 |
| | 18% | 75.3% | 6.7% |
| Total : 119 | | | |

we used the special case of the *k-fold cross-validation method* with *k*: 12. With 199 test cases, number of fold 10 and repetitions were used in Table A.3.

Table A.3
Cluster analysis data

| | first | second | third | fourth | fifth |
|---|---|---|---|---|---|
| Failed firms | 0.87395 | 0.722689 | 0.815126 | 0.7731 | 0.7647 |
| Healthy firms | 0.77528 | 0.719 | 0.82 | 0.77528 | 0.67415 |

**References**


1. Altman, E.I, Financial ratios, discriminant analysis, and the prediction of corporation bankruptcy. Journal of Finance. 23 (4), 589-609 (1968).
2. Ahn, B.S., Cho, S.S. and Kim, C.Y., The integrated methodology of rough set theory and artificial neural network for business failure prediction. Expert Systems with Applications. 18 (10), 65-74 (2000).
3. Beaver, W.H., Financial ratios as predictors of failure. Journal of Accounting Research. 4, 71-111 (1966).
4. Beaver, W.H., Alternative accounting measures as predictors of failure. Accounting Review. 43 (1), 113-122 (1968).
5. Blum, M., Failing company discriminant analysis. Journal of Accounting Research. 12 (1), 1-25 (1974).
6. Beynon, M.J. and Peel, M.J., Variable precision rough set theory and data discrimination: An application to corporate failure prediction. OMEGA (the International Journal of Management Science). 29 (6), 561-576 (2001).
7. Edward B. Deakin, Discriminant analysis of predictor of business failure. Journal of Accounting Research. 10 (1), 167-179 (1972).
8. Dimitras, A. I., Slowinski, R., Susmaga, R. and Zopounidis, C., Business failure prediction using rough sets. European Journal of Operational Research. 114 (2), 263–280 (1999).

9. Huang, J.J., Ong, C.S. and Tzeng, G.H., Building credit scoring models using genetic programming. Expert Systems Whit Applications. 29 (1), 41-47 (2005).
10. Lo, A.W., Logit versus discriminant analysis: A specification test and application to corporate bankruptcy. Journal of Econometries. 31 (2), 151~178 (1986).
11. Min, J.H., & Lee, Y.C., Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. Expert systems with Applications. Uncorrected proof, available online 5 (2005).
12. Ohlson, J.A., Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research. 18 (1), 109~131 (1980).
13. Patton A. and Baker J.C. Why won’t directors rock the boat. Harvard Business Review. 65 (6), 10-18 (1987).
14. Pawlak, Z., Rough sets. International Journal of Computer and Information Science. 11 (5), 341–356 (1982).
15. Pawlak, Z. Rough classification. International Journal of Man-Machine Studies. 20 (5), 469–483 (1984).
16. Pawlak, Z., Grzymala-Busse, J., Słowiński, R., and Ziarko, W. Rough sets. *Communications of the ACM*. 38 (11), 89-95 (1995).
17. Segovia, M.J., Gil, J.A., Heras, A., Vilar, J.L. and Sanchis, A, Using rough sets to predict insolvency of Spanish nonlife insurance companies. Documentos de trabajo de la facultad de ciencias económicas y empresariales. Avaliable: http://econpapers.repec.org/paper/ucmdoctra/ (2003).
18. Shuai, J.J. and LI, H.L., Using rough set and worst practice DEA in business failure prediction. Lecture Notes in Computer Science. Vol.3642, 503-510 (2005).
19. Walczak, B. and Massart, D.L., Rough set theory. Chemometrics and Intelligent Laboratory Systems. 47 (1), 1-16 (1999).
20. Wu, C.H., Tzeng, G.H., Goo, Y.J. and Fang, W.C., A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy”. Expert Systems with Applications. 32 (2), 397-408 (2007).