경영정보학연구
제9권 제4호
1999년 12월

도산 예측을 위한 러프집합이론과 인공신경망 통합방법론

김 창 연*, 안 병 석**, 조 성 식*, 김 성 희*

The Integrated Methodology of Rough Set Theory and Artificial Neural Network for Business Failure Prediction

Kim, Chang-Yun, Ahn, Byeong-Seok, Cho, Sung-Sik, Kim, Soung-Hie

This paper proposes a hybrid intelligent system that predicts the failure of firms based on the past financial performance data, combining neural network and rough set approach. We can get reduced information table, which implies that the number of evaluation criteria such as financial ratios and qualitative variables and objects (i.e., firms) is reduced with no information loss through rough set approach. And then, this reduced information is used to develop classification rules and train neural network to infer appropriate parameters. Through the reduction of information table, it is expected that the performance of the neural network improve. The rules developed by rough sets show the best prediction accuracy if a case does match any of the rules. The rationale of our hybrid system is using rules developed by rough sets for an object that matches any of the rules and neural network for one that does not match any of them. The effectiveness of our methodology was verified by experiments comparing traditional discriminant analysis and neural network approach with our hybrid approach. For the experiment, the financial data of 2,400 Korean firms during the period 1994-1996 were selected, and for the validation, k-fold validation was used.

* Graduate School of Management, Korea Advanced Institute of Science and Technology
** Department of Business Administration and Accounting, Suwon University

Ⅰ. Introduction

Evaluation of the business failure has been, for a long time, a major preoccupation of researchers and practitioners. Business failure is a general term 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 according to the law. All these situations result in a discontinuity of the firm's operations. 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].

The development and use of models, able to predict failure in advance, can be very important for 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.

At the beginning, statistical methods such as univariate statistical methods, multiple discriminant analysis, linear probability models, and logit and probit analysis have been mainly used for business classification problems [Altman, 1968; Altman et al., 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 et al., 1991; Salchen- berger et al., 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 et al., 1993; Slowinski & Zopounidis, 1995]. Rough set theory, introduced by Pawlak [1982] and Pawlak et al.[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 (firms) described by a set of multi-valued financial ratios and qualitative variables.

The composite models of neural network and rough set components, which take advantage of each method's generic characteristics, were constructed to predict a sample of bank holding patterns [Jelonek et al., 1995; Hashemi et al., 1998]. They used rough sets as a preprocessor for neural network. Using sorting rules developed by rough sets may lead to a burdensome situation where a record does not match any of the sorting or classification rules. On the other hand, the neural network approach classifies every object by its weighting mechanism, although in a black box fashion. Jelonek et al. employed 1-dimensional reduction (attribute reduction) and Hashemi et al. employed 2-dimensional reduction (attribute and object reduction). After reduction of information system (i.e. decision table) they trained neural network with the reduced information system. And then, they applied the neural network to prediction.

In this paper, we propose a hybrid intelligent system combining neural network and rough set approach. Rough set approach, by which redundant attributes in multi-attribute information system can be removed without any information loss, is utilized as a preprocessor to improve business prediction capability by neural network. The streamlined information system by rough set approach is fed into neural network for training. At the prediction step, we apply the rules developed by rough sets first, and then we apply the neural network to the objects that does not match any of the rules. The effectiveness of our hybrid approach was verified with experiments that compared traditional discriminant analysis and neural network approach with the hybrid approach.

This paper is organized as follows. Section 2 describes classification techniques used in previous researches concerned with our paper: rough set theory and neural network. In section 3, proposed data preprocessing algorithm by rough set and hybrid models are de- scribed. In section 4, we compare and analyze the results of each model. In the last section, limitations and further research issues are discussed.

Ⅱ. Rough sets and neural networks

A brief description of the neural network model and the backpropagation algorithm is presented here. More details can be found in Hertz et al. [1991] and Wasserman [1989].

2.1 Rough sets

Pawlak [1982] first introduced rough set theory. The philosophy of the method is based on the assumption that with every object some information (data, knowledge) can be associated. Because of its simplicity and usefulness, rough set approach has been applied in many domains.

Slowinski et al. [1997] applied the rough set approach in a real problem considered by a Greek bank which finances industrial and commercial firms in Greece presenting a great activity. The bank was interested in investing its funds in the best and dynamic firms. Slowinski & Zopounidis [1995] employed rough set approach in business failure prediction. They used 12 financial ratios and compared rough set approach with statistical approaches.

The following is a brief review of rough set theory

2.2.1. Information system

An information system is represented by 4-tuple $S=\langle U, Q, V, \rho\rangle$, where U is a finite set of objects, Q is a finite set of attributes, $V=\bigcup_{q\in Q} V_q$ and V_q is a domain of the attribute q, and $\rho: U\times Q \rightarrow V$ is a total function such that $\rho(x, q)\in V_q$ for every $q\in Q$, $x\in U$, called an information function.

Let $P\subseteq Q$ and x, $y\in U$. We say that x and y are indiscernible by the set of attributes P in S iff $\rho(x, q)=\rho(y, q)$ for every $q\in P$. Thus every $P\subseteq Q$ generates a binary relation on U which will be called an indiscernibility relation, denoted by $IND(P)$. Obviously $IND(P)$ is an equivalence relation for any P. Equivalence classes of $IND(P)$ are called P-elementary sets in S. The family of all equivalence classes of relation $IND(P)$ on U is denoted by $U|IND(P)$ or, in short, $U|P$.

$Des_p(X)$ denotes a description of P-elementary set $X \in U | P$ in termsx of values of attributes from P, i.e.

$$Des_p(X) = \{ (q,\ v): \rho(x,\ q) = v, \forall x \in X,\ \ \forall q \in P\}$$

2.2.2. Approximation of Sets

Let $P \subseteq Q$ and $Y \subseteq U$. The P-lower approximation of Y, denoted by $\underline{P}Y$, and the P -upper approximation of Y, denoted by $\underline{P}Y$, are defined as:

$$\overline{P}Y = \bigcup\{X \in U \mid P.X \subseteq Y\}$$

$$\underline{P}Y = \bigcup\{X \in U \mid P{:}X \cap Y \neq \emptyset\}$$

The P-boundary (doubtful region) of set Y is defined as

$$Bn_p(Y) = \overline{P}Y - \overline{P}Y$$

Set $\underline{P}Y$ is the set of all objects from U which can be certainly classified as elements of Y, employing the set of attributes P. Set $\overline{P}Y$ is the set of objects from U which can be possibly classified as elements of Y, using the set of attributes P. The set $Bn_p(Y)$ is the set of objects which cannot be certainly classified to Y using the set of attributes P only.

With every set $Y \subseteq U$, we can associate an accuracy of approximation of set Y and P in S, or in short, accuracy of Y, defined as:

$$a_P(Y) = \frac{card(\underline{P}Y)}{card(\overline{P}Y)}$$

card means cardinality.

2.2.3 Approximation of a Partition of U

Let S be an information system, $P \subseteq Q$, and let $\Psi = \{Y_1, Y_2, \ldots, Y_n\}$ be a partition of U. The origin of this partition is independent on attributes from P; it can follow from solving a sorting problem by an expert. Subsets Y_i, $i = 1, \ldots, n$, are categories of partition Ψ. By P-lower and P-upper approximation of Ψ in S we mean sets $\underline{P}\Psi = \{\underline{P}Y_1, \underline{P}Y_2, \ldots, \underline{P}Y_n\}$ and $\overline{P}\Psi = \{\overline{P}Y_1, \overline{P}Y_2, \ldots, \overline{P}Y_n\}$, respectively. The coefficient

$$r_p(\Psi) = \frac{\sum_{i=1}^{n} card(\underline{P}Y_i)}{card(U)}$$

is called the quality of approximation of partition Ψ by set of attributes Ψ, or in short, quality of sorting. It expresses the ratio of all P-correctly sorted objects to all objects in the system.

2.2.4 Reduction of Attributes

We say that the set of attributes $R \subseteq Q$ depends on the set of attributes $P \subseteq Q$ in S (denotation $P \rightarrow R$) iff $IND(P) \subseteq IND(R)$. 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. The minimal subset $R \subseteq P \subseteq Q$ such that $\gamma_P(\Psi) = \gamma_R(\Psi)$ is called Ψ-reduct of P(or, simply, reduct if there is no ambiguity in the understanding of Ψ)

and denoted by $RED_{\Psi}(P)$. Note that an information system may have more than one Ψ-reduct. Intersection of all Ψ-reducts is called the Ψ-core of P, i.e. $CORE_{\Psi}(P) = \bigcap RED_{\Psi}(P)$.

The core is a collection of the most significant attributes in the system. It can also be empty.

2.2.5 Decision Tables

An information system can be seen as a decision table assuming that $Q = C \bigcup D$ and $C \bigcap D = \emptyset$, where C are called condition attributes, and D, decision attributes. Decision table $S = \langle U, C, \bigcup D, V, \rho \rangle$ 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 $U \mid IND(C)$ be a family of all C-elementary sets called condition classes, denoted by X_i (i=1,...,k, *where k is the number of* $U \mid$ IND(C)). Let, moreover, $U \mid IND(D)$ be the family of all D-elementary sets called decision classes, denoted by Y_j (j=1,...,k, *where n is the number of* $U \mid$ 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 $(j = 1, \ldots, n)$ is denoted by $\{r_{ij}\}$. More precisely, $\{r_{ij}\} = \{Des_C(X_i) \Rightarrow Des_D(Y_j) : X_i \bigcap Y_j \neq \emptyset,\ i = 1, \ldots, k\}$ Rule r_{ij} is deterministic iff $X_i \subseteq Y_j$, and r_{ij} is non-deterministic otherwise.

Procedures for derivation of decision rules from decision tables were presented by Boryczka & Slowinski [1988], Slowinski & Stefanowski [1992], Grzymala-Busse [1992], Siegel et al. [1993] and Ziarko et al. [1993].

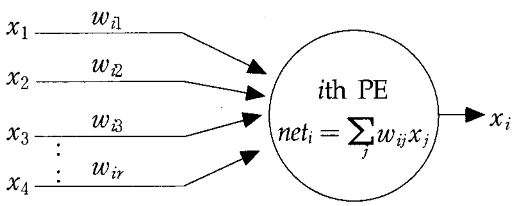
2.2.6 Decision support using decision rules

Decision rules derived from a decision table can be used for recommendations concerning new objects. Specifically, matching its description to one of the decision rules can support the classification of a new object. The matching may lead to one of four situations [Slowinski & Stefanowski, 1994]:

(a) the new object matches one deterministic rule,
(b) the new object matches more than one deterministic rules suggesting, however, the same decision class,
(c) the new object matches one non-deterministic rule or several rules suggesting different decision classes,
(d) the new object does not match any of the rules.

2.2 Neural networks

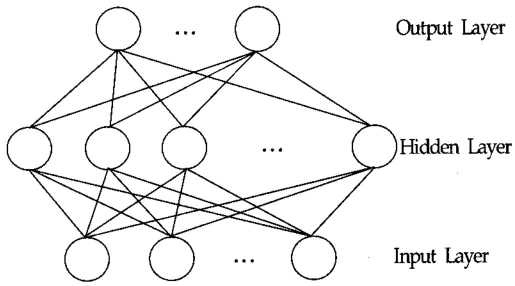
Neural network technology was developed in an attempt to mimic the acquisition of knowledge and organization skills of the human brain. It offers significant support in terms of organizing, classifying, and summarizing data. It also helps to discern patterns among input data, requires few assumptions, and achieves a high degree of prediction accuracy. These characteristics make neural network technology a potentially promising alternative tool for recognition, classification, and forecasting in the area of finance, in terms of accuracy, adaptability, robustness, effectiveness, and efficiency. Therefore, financial application areas that require pattern matching, classification and prediction, such as bankruptcy prediction, loan evaluation, credit scoring, and bond rating, are fruitful candidate areas for neural network technology.



<Figure 1> General symbol of PE (Processing Element)

The individual computational units that make up neural network are referred to as nodes, units, or processing elements (PEs). <Figure 1> shows general PE model. Each PE is numbered, the one in the figure being the i th. The PE has many inputs, but has only a single output, which can fan out to many other PEs in the network. The input the i th PE receives from the j th PE is indicated as x_j (note that this value is also the output of the j th node, just as the output generated by the i th node is labeled x_i). Each connection to the i th PE has associated with it a quantity called a weight or connection strength. Weights are adaptive coefficients within the network that determine the intensity of the input signal. The weight on the connection from the j th node to the i the node is denoted w_{ij}. Each PE determines a net-input value $net_i = \sum_j w_{ij}, x_j$, based on all its input connections. Once the net input is calculated, it is converted to an activation value, by activation function, $x_i = f_i(net_i)$, for the PE. Hard limit, threshold, and sigmoid functions are used as activation function and among them sigmoid function is most widely used

The set of PEs which run simultaneously, or in parallel, is called a layer and there are input layer, output layer, and hidden layer(s). <Figure 2> describes general neural network having a hidden layer.



<Figure 2> Multi-layered Neural Network

Financial applications of neural networks typically focus on pattern matching, classification and forecasting. These functions include mortgage underwriting judgements [Collins et.al., 1988], credit card fraud detection [Rochester, 1990], prediction of corporate bond ratings [Dutta & Shekhar, 1988], and forecasting credit card risk from customer applications [Trippi & Turban, 1989]. Salchenberger et.al. [1992] employed a neural network model that processed input data of financial ratios to discriminate between healthy and failing financial institutions. Similar studies with neural network for prediciting bankruptcy were conducted by Messier & Hansen [1988] and Fletcher & Goss [1993]. Tam & Kiang [1992] employed neural networks to predict cases of bank failure in Texas, and Altman et al.[1994] used neural network for the prediction of failure of Italian firms, providing encouraging results, comparable to those of discrinimant analysis models.

Ⅲ. Research model development

3.1 Rough set data preprocessing

In the process of rough set data analysis, attributes can be reduced, which implies that some redundant attributes that do not play any role in distinguishing an object from the others, can be eliminated without any information loss. And the final result to which rough set approaches direct is the production rule that is capable of predicting newly gathered data. Recently, a new trial of combining rough set and neural network was suggested by Jelonek et al. [1995]. In the conclusion of that paper the authors say: Rough set based reduction of data representation appears to cooperate well with backpropagation-learned neural networks due to the fact that attribute reduction by rough set approach is helpful for reducing the input variables for neural network approach.

This paper employs 2D (2-dimensional)-reduction data-preprocessing algorithm. The name 2D implies that information system is reduced both horizontally and vertically. Horizontal reduction is, in other words, attribute reduction. The reduced set of attributes provides the same quality of sorting as the original set of attributes. Vertical reduction is the reduction of *conflicting* objects that can result in deterioration of accuracy of neural network. Hashemi et al. [1998] also employed vertical reduction, but they removed *redundant* objects only. They defined the redundant objects as those with same attribute values and same decision value: for two objects, x_i and x_j, $(a_i = a_j)\,(d_i = d_j) \Rightarrow (x_i = x_j)$.

From the viewpoint of neural network, conflicting objects are noises. With same inputs but different outputs, neural network cannot decide which to return for an output. It's the power game of conflicting objects that decide the output. That is, if there are more objects with "failed" output than the objects with "not failed" output, then neural network will return "failed". If the numbers of conflicting objects are equal then the output of neural network does not have any meaning. In other words, after the reduction of redundant objects, the conflicting objects do not have any effect on neural network training. Thus, we removed the conflicting objects.

3.3.1 Horizontal Reduction

Find minimal subset $R \subseteq Q$ such that $\gamma_Q(Y)$, $= \gamma_R(Y)$ in other words, $RED_Y(Q)$. $RED_Y(Q)$ will be the reduced attribute set with which the neural network will be trained

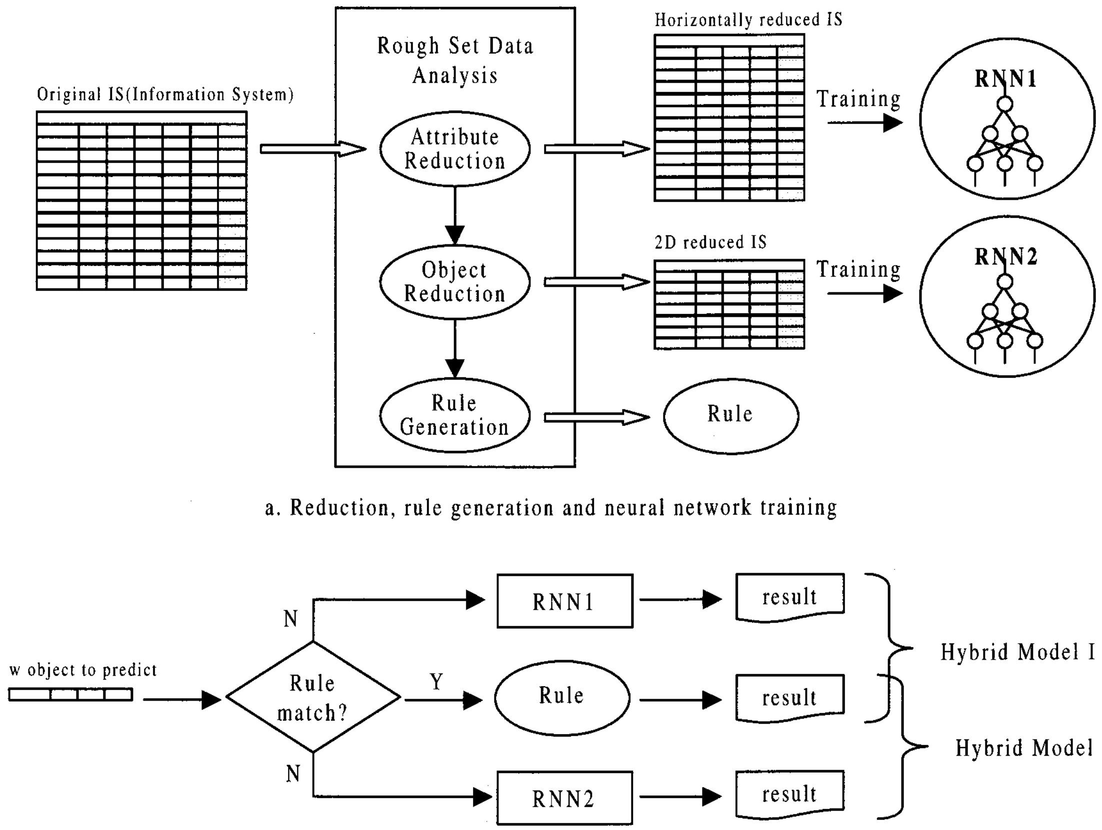
3.3.2 Vertical Reduction

Let $x_i, x_j \in BN_R(Y)$ $(R$ *is* $RED_Y(Q))$. If $a_i = a_j$ and $d_i \neq d_j$ for the set of two objects x_i, x_j we call them conflicting pair. Repeat removal of conflicting pairs until there is none left.

3.2 The hybrid models

Hybrid model proposed in this paper is composed of rough set component and neural network component. By rough set, some rules are extracted from the information system. Using rough set tool, we can discover knowledge in two kinds of rules: deterministic and non-deterministic. Usually, the rules generated by rough set approach fail to predict newly entered object because of non-deterministic rules. To handle this situation, some researchers reported that reduced data set (horizontally or vertically) is fed into neural network for complementing the limitation of rough set, which finally produces full prediction of new case data.

In this study, we tested two possible hybrid models. Rough set component combined with neural network trained only with horizontally reduced information system is one and rough set and neural network trained with 2D reduced information system is the other. We named the former Hybrid model I, and the latter Hybrid model II. These two hybrid models will be compared with traditional discriminant analysis model and neural network model.



a. Reduction, rule generation and neural network training

b. Hybrid models

<Figure 3> Hybrid models

<Figure 3> shows the configuration of the Hybrid model Ⅰ and Ⅱ. First, original information system enters the rough set component. Through attribute reduction, rough set component produces horizontally reduced information system. And then rules are generated based on reduced attributes and neural network is trained with horizontally reduced information system.

When a new object enters the system, the system first finds whether there is a rule that matches the object. If there is a rule that matches, follow the rule. In case of no match, the object is sent to neural network component, and neural network classifies it. The only difference between Hybrid model Ⅰ and Ⅱ is the training set of neural network. In Hybrid model Ⅱ vertical reduction is added and 2D-reduced information system is used as a training set of neural network. Hybrid model Ⅱ manages conflicting data by vertical reduction and resulting 2D-reduced information system is consistent.

Ⅳ. Experiment

4.1 Research data

For the experiment, the financial data of Korean companies from 1994 to 1996 were selected. We avoided using data of the current economic crisis because the business environments in this crisis are quite different from usual business environments before the crisis and this would make our samples biased to the special case. After the crisis many companies went bankrupt. That wouldnt happen if it were not the period of special crisis. Sample data were drawn randomly from the list of companies which a credit research company, KIS (Korea Information Service), has. The sample data sets consist of the equal number of healthy and failed cases. 2,400 companies (1,200 of healthy firms and 1,200 of failed firms) from various industries and of various size were used, excluding financial industry. Financial industry was excluded because the variables we used are not adequate to indicate the state of financial industry. We defined a dishonored company as a failed company and a company with high enough credit score to be classified as investment possible as a healthy one. Financial ratios we used were gathered one year before the firm was classified. In other words, when a company received investment possible credit level in a certain year, this company is classified as a healthy one and financial data one year before are used.

The variable set considered in this paper is comprised of 8 financial ratios such as, A_1 (net income to total asset), A_2(net income to sales), A_3(owner's equity to total asset), A_4 (total borrowings and bonds payable to total asset), A_5(net working capital to total asset), A_6 (cash flow to total liability), A_7(iventory turnover), and A_8(current asset to current liability). The selection of the financial ratios is based upon two main characteristics: their usefulness in previous studies [Altman, 1968; Beaver, 1966; Gilbert, 1990; Keasey et al., 1990; Lee et al., 1996; Odom & Sharda, 1989; Raghupathi et al., 1991], and the experiences from past decisions, the knowledge and the preferences of financing experts.

As rough set approach is concerned with discrete values, we have to transform quanti-

<Table 1> The definition of norms for quantitative attributes

Attributes	1	2	3	4	5
A_1	Below 0	(0,2]	(2,5]	Over 5	
A_2	Below 0	(0,2]	(2,5]	Over 5	
A_3	Below 0	(0,10]	(10,20]	(20,33.3]	Over 33.3
A_4	Over 100	(66.7,100]	(33.3,66.7]	Below 33.3	
A_5	Below 0	(0,0.33]	(0.33,1]	(1,10]	Over 10
A_6	Below 1 or Over 10	(-1,-0.5],[1,10)	(-0.5,-0.2],[0.5,1)	(-0.2,0],[0.2,0.5)	(0,0.2)
A_7	Below 5	(5,10]	(10,15]	(15,20]	Over 20
A_8	Below 80	(80,100]	(100,120]	(120,150]	Over 150

tative attributes into qualitative terms according to some norms following from the financial manager's experience and some standards of the corporate financial analysis.

The use of norms translating the quantitative attributes into qualitative terms is not imposed by the rough set approach but by a practical interpretation of the qualitative attributes. Even if an attribute represents a continuous measure, such as financial ratios, or blood pressure in medicine, the expert usually interprets the values of this attribute in qualitative terms, i.e. low, medium or high. The norms used for this interpretation come from tradition, habits or convention. As they are consequently used from the beginning of problem setting until final explanation of decision rules, they do not falsify the original image of the decision situation. The result of using the norms and the codes for translation of the original information system, one obtains the coded information system presented in <Table 1>.

4.2 Neural network configuration

In experiment, we used 3 neural networks. Ordinary neural network was denoted by NN. Neural network trained with horizontally reduced information system denoted by RNN1 (for Hybrid Model Ⅰ). Neural network trained with horizontally and vertically reduced information system denoted by RNN2 (for Hybrid Model Ⅱ). The network structure of NN was 8-5-1 for input layer, hidden layer and output layer respectively. Both RNN1 and RNN2 had 4-3-1 structure. NN, RNN1, RNN2 used sigmoid function for activation and backpropagation algorithm for learning.

Rough set analysis part of the experiment was performed with program we developed in JAVA. We used NeuroShell 2 release 3.0 for neural network and SPSS for statistical tests.

4.3 Experiment and results

4.3.1 K-Fold Cross Validation

A more reliable approach is to use a resampling method, such as k-fold cross validation or bootstrapping. In the test, we used the special case of the k-fold cross-validation method with k=12. With 2,400 test cases, 12 repetitions were used. In each repetition 2,200 cases were used for the training set and 200 cases 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 guaranteed to participate in training and testing.

For the comparison between methodologies, we separated holdout sample into two groups, after rough set analysis was finished and rules were generated. We classified cases that matched sorting rule into Group Ⅰ and the other Group Ⅱ. In other words, Group Ⅰ was group of cases that have matching rules and Group Ⅱ was group of cases that do not have matching rules. Group Ⅱ cases had no matching rule, thus they needed neural network component to classify them.

<Table 2> shows the results after rough set analysis was performed. As we can see, in 12 experiments we obtained one minimal reduct $\{A_4, A_5, A_6, A_8\}$. This minimal reduct is the result of horizontal reduction.

The column named Group I shows ratio of cases that have a matching rule. In the experiment, 89% of healthy firms and 81% of failed firms in the holdout sample were classified as Group Ⅰ.

DR objects reduced column shows the percentage of reduced object after vertical reduction.

After rough set analysis was finished and holdout sample was separated into two groups, we tested performance of each methodology. First, five methods were tested with Group Ⅰ holdout sample. Next, four methods except rule method were tested with Group Ⅱ sample. Because cases in Group Ⅱ have no matching rule, rules cannot classify them. <Table 3> shows Group Ⅰ hit ratios of each method and <Table 4> shows Group Ⅱ hit ratios.

<Table 2> Results after Rough set data analysis

Experiment #	Minimal Reduct	Quality of Sorting	DR Objects Reduced	Group Ⅰ		Group Ⅱ	
				H	F	H	F
1	$\{A_4, A_5, A_6, A_8\}$	0.897	0.731	0.89	0.81	0.11	0.19
2	$\{A_4, A_5, A_6, A_8\}$	0.895	0.693	0.9	0.82	0.1	0.18
3	$\{A_4, A_5, A_6, A_8\}$	0.903	0.784	0.85	0.71	0.15	0.29
4	$\{A_4, A_5, A_6, A_8\}$	0.9	0.768	0.87	0.73	0.13	0.27
5	$\{A_4, A_5, A_6, A_8\}$	0.891	0.692	0.92	0.89	0.08	0.11
6	$\{A_4, A_5, A_6, A_8\}$	0.896	0.703	0.94	0.85	0.07	0.14
7	$\{A_4, A_5, A_6, A_8\}$	0.9	0.773	0.86	0.76	0.14	0.24
8	$\{A_4, A_5, A_6, A_8\}$	0.898	0.729	0.9	0.77	0.11	0.22
9	$\{A_4, A_5, A_6, A_8\}$	0.895	0.749	0.92	0.82	0.1	0.16
10	$\{A_4, A_5, A_6, A_8\}$	0.894	0.756	0.91	0.84	0.1	0.15
11	$\{A_4, A_5, A_6, A_8\}$	0.895	0.706	0.93	0.84	0.08	0.15
12	$\{A_4, A_5, A_6, A_8\}$	0.897	0.744	0.89	0.8	0.11	0.2

DR : Doubtful Region, H: Healthy, F: Failed

<Table 3> Hit ratios of methods in Group I holdout subset in percent

Experiment #	Rule		DA		NN		RNN1		RNN2	
	H	F	H	F	H	F	H	F	H	F
1	95.5	97.5	67.4	88.9	83.1	86.4	88.8	92.6	88.8	92.6
2	95.6	97.6	73.3	89.0	83.3	87.8	86.7	95.1	87.8	95.1
3	95.3	95.8	68.2	85.9	84.7	83.1	87.1	95.8	85.9	95.8
4	95.4	95.9	69.0	87.7	81.6	87.7	87.4	94.5	87.4	95.9
5	93.5	96.6	69.6	91.0	85.9	85.4	84.8	96.6	85.9	96.6
6	95.7	96.5	75.5	89.4	85.1	87.1	86.2	91.8	86.2	92.9
7	96.5	96.1	69.8	85.5	83.7	86.8	81.4	96.1	81.4	96.1
8	95.6	97.4	70.0	88.3	82.2	88.3	86.7	94.8	86.7	94.8
9	95.7	96.3	72.8	89.0	85.9	82.9	82.6	92.7	83.7	92.7
10	94.5	96.4	71.4	90.5	81.3	85.7	85.7	89.3	85.7	90.5
11	95.7	96.4	68.8	88.1	83.9	84.5	88.2	95.2	88.2	95.2
12	94.4	97.5	66.3	86.3	80.9	87.5	84.3	92.5	84.3	93.8
Avg.	95.3	96.7	70.2	88.4	83.5	86.1	85.8	93.8	86.0	94.3
Overall Avg.	95.9		78.8		84.7		89.5		89.9	

<Table 4> Hit ratios of models in Group II holdout subset

Experiment #	DA		NN		RNN1		RNN2	
	H	F	H	F	H	F	H	F
1	72.7	84.2	81.8	84.2	81.8	94.7	81.8	94.7
2	60.0	83.3	80.0	88.9	80.0	94.4	80.0	94.4
3	66.7	89.7	80.0	86.2	86.7	93.1	86.7	93.1
4	69.2	88.9	76.9	92.6	84.6	88.9	84.6	92.6
5	62.5	90.9	75.0	81.8	87.5	81.8	87.5	81.8
6	71.4	85.7	85.7	85.7	85.7	92.9	85.7	92.9
7	64.3	83.3	78.6	83.3	78.6	91.7	78.6	91.7
8	72.7	86.4	81.8	86.4	90.9	95.5	90.9	95.5
9	60.0	81.3	80.0	87.5	80.0	93.8	80.0	93.8
10	70.0	86.7	80.0	80.0	80.0	86.7	80.0	86.7
11	75.0	73.3	87.5	86.7	87.5	93.3	87.5	93.3
12	63.6	80.0	81.8	85.0	81.8	90.0	81.8	90.0
Average	67.2	84.8	80.5	86.1	83.6	91.7	83.6	92.2
Overall Avg.	78.5		84.1		88.8		89.1	

<Table 5> Hit ratios of models

Experiment #	DA	NN	RNN1	RNN2	HYBRID I	HYBRID II
1	0.780	0.845	0.905	0.905	0.946798	0.946879
2	0.800	0.855	0.905	0.910	0.946778	0.948878
3	0.775	0.840	0.910	0.905	0.943294	0.940529
4	0.785	0.850	0.900	0.910	0.938218	0.947418
5	0.800	0.850	0.900	0.905	0.925978	0.916271
6	0.820	0.860	0.890	0.895	0.951915	0.940268
7	0.770	0.845	0.880	0.880	0.946744	0.949386
8	0.790	0.850	0.910	0.910	0.952889	0.955592
9	0.795	0.845	0.875	0.880	0.947174	0.946853
10	0.805	0.830	0.870	0.875	0.931923	0.949974
11	0.775	0.845	0.915	0.915	0.951935	0.942051
12	0.755	0.840	0.880	0.885	0.932528	0.937074
Average	0.7875	0.84625	0.895	0.897917	0.943015	0.943431

<Table 6> Paired t-test of methods (Group I)

	NN t-value (sig)	RNN1 t-value (sig)	RNN2 t-value (sig)	RULE t-value (sig)
DA	-11.66 (.000)	-13.52 (.000)	-15.17 (.000)	-28.68 (.000)
NN		-10.88 (.000)	-9.90 (.000)	-12.92 (.000)
RNN1			-2.44 (.017)	-14.13 (.000)
RNN2				-14.51 (.000)

After testing each methodology, we calculated performance of prediction models. These models were discriminant analysis model (DA), neural network model (NN), neural network trained with horizontally reduced information system (RNN1), neural network trained with horizontally and vertically reduced information system (RNN2), Hybrid model Ⅰ, and Hybrid model Ⅱ.

4.5 Analysis of the results

We can see that rule performs best in classifying Group I samples. And we can see also that RNN2 outperforms RNN1. In Group Ⅱ test, RNN2 outperforms the others.

<Table 7> Paired t-test of methods (Group II)

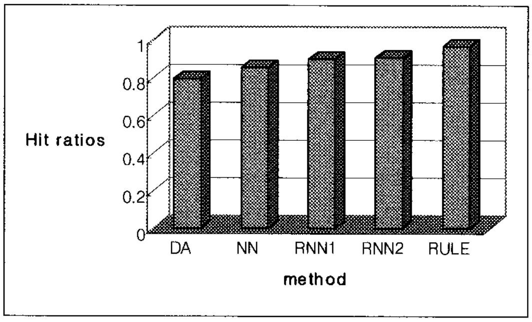
	NN t-value (sig)	RNN1 t-value (sig)	RNN2 t-value (sig)
DA	-4.43 (.001)	-8.56 (.000)	-9.25 (.000)
NN		-7.33 (.000)	-9.32 (.000)
RNN1			-1.00 (.169)

Six prediction models performances are also compared. Hybrid model Ⅰ and Hybrid model Ⅱ outperform the others, but the performance difference between hybrid models was statistically insignificant. That means we cannot say that Hybrid model Ⅱ is better than Hybrid model Ⅰ.

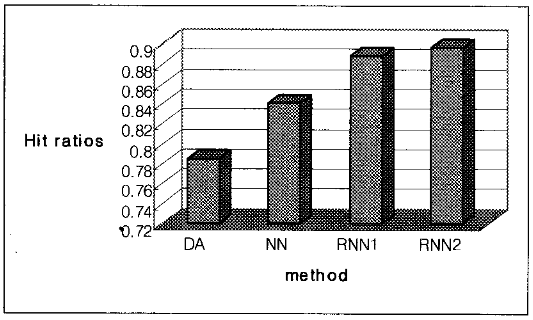
<Figure 4> shows hit ratios of each methods tested with Group I holdout sample. We can

<Table 8> Paired t-test of models

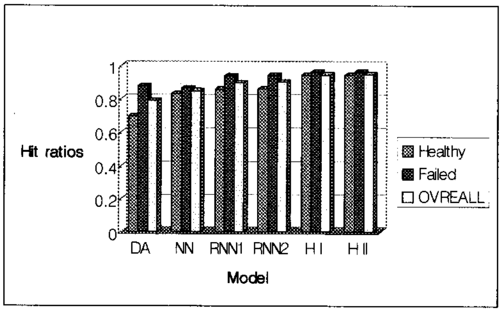
	NN t-value (sig)	RNN1 t-value (sig)	RNN2 t-value (sig)	HYBRID I t-value (sig)	HYBRID II t-value (sig)
DA	-12.63 (.000)	-15.12 (.000)	-16.77 (.000)	-27.98 (.000)	-27.98 (.000)
NN		-11.92 (.000)	-14.54 (.000)	37.89 (.000)	39.51 (.000)
RNN1			-2.55 (.014)	11.34 (.000)	11.62 (.000)
RNN2				10.79 (.000)	11.25 (.000)
HYBRID I					-1.00 (.170)



<Figure 4> Hit ratios of methods (Group I)



<Figure 5> Hit ratios of methods (Group II)



<Figure 6> Hit ratios of models

see Rule outperform the others. <Figure 5> shows hit ratios of methods tested with Group II holdout sample. Because Group II holdout sample cannot be classified by Rule, Rule is excluded in the graph.

Ⅴ. Conclusions

The experiment results show the effectiveness of rough set approach as a data preprocessor for neural network. We proposed 2D-reduction algorithm and proved its usefulness. Horizontal reduction reduced attributes and vertical reduction reduced conflicting objects. The reduction of information system has a great meaning for neural network in that reduction of attributes prevents overfitting problem and saves training time. Further, removing conflicting objects and train neural network with consistent cases can bring performance improvement as well as reduction of training time.

Proposed hybrid models were proven to outperform discriminant analysis model and neural network models. In comparison between Hybrid model Ⅰ and Hybrid model Ⅱ, the experiment results could not show that the Hybrid model Ⅱ performed better than Hybrid model Ⅰ. This was because both models used same rule set generated by rough set analysis, and this rule component covered rather large part compared to neural network component.

In our research model, firms of various industries and different size were mixed in case data set, so we neglected the difference among industries and difference in the sizes of companies. Further, in generating rules, we strictly applied rough set theory, so there existed a possibility that rule could not be generated because of just a few extraordinary objects. With more work in dealing with doubtful region we may discover more valuable knowledge. For example, we can make some rules in rough set module such that if the percentage of conflicting object is less than 5%, we can ignore them and generate rule from those objects.

References

[1] Altman, E.I., Financial Ratios, Discriminant Analysis and Prediction of Corporate Bankruptcy, *The Journal of Finance*, 23, 1968, pp. 589-609.

[2] Altman, E.I., Haldeman, R.G., and Narayanan, P., Zeta Analysis, *Journal of Banking and Finance*, June 1977, pp. 29-51.

[3] Altman, E.I., Marco, G., and Varetto, F., Corporate Distress Diagnosis: Comparisons using Discriminant Analysis and Neural Networks (the Italian Experience), *Journal of Banking and Finance*, 18, 1994, pp. 505-529.

[4] Beaver, W.H., Financial Ratios as Predictors of Failure, Empirical Research in Accounting: Selected Studies, *Journal of Accounting Research*, Supplement to Vol. 5, 1990, pp. 179-199.

[5] Blum, M., Failing Company Discriminant Analysis, *Journal of Accounting Research*, spring 1974, pp. 1-25.

[6] Boryczka, M., and Slowinski, R., Derivation of Optimal Decision Algorithms from Decision Tables using Rough Sets, *Bulletin of the Polish Academy of Sciences, ser. Technical Sciences*, 36, 1988, pp. 252-260.

[7] Chung, H., and Tam, K., A Comparative Analysis of Inductive-Learning Algorithm, *Intelligent Systems in Accounting, Finance and Management*, 2, 1992, pp.3-18.

[8] Collins, E., Ghosh, S., and Scofield, C., An Application of a Multiple Neural Network Learning System to Emulation of Mortgage Underwriting Judgments, *In: Proceedings of the IEEE International Conference on Neural Networks*, 1988, pp. 459-466.

[9] Collins, R.A., and Green, R.D., Statistical Methods for Bankruptcy Forecasting, Journal of Economics and Business, 32, 1972, pp. 349-354.

[10] Dutta, S., and Shekhar, S., Bond Rating: A Non-conservative Application of Neural Networks, *In: Proceedings of the IEEE International Conference on Neural Networks*, 1988, pp. 443-450.

[11] Fletcher, D., and Goss, E., Forecasting with Neural Networks: An Application Using Bankruptcy Data, *Information and Management*, 24, 1993, pp. 159-167.

[12] Gilbert, L.R., Menon, K., and Schwartz, K.B., Predicting Bankruptcy for Firms in Financial Distress, *Journal of Business Finance and Accounting*, 17(1), 1990, pp. 161-171.

[13] Grzymala-Busse, J.W., LERS-A System for

Learning from Examples Based on Rough Sets, in: R. Slowinski (ed.), *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory*, Kluwer Academic Publishers, Dordrecht, 1992, pp. 3-18.

[14] Hashemi, R., Le Blanc, L.A., Rucks, C.T., and Rajaratnam, A., A Hybrid Intelligent System for Predicting Bank Holding Structures, *European Journal of Operational Research*, 109(2), 1998, pp. 390-402.

[15] Hertz, J., Krogh, A., and Palmer, R.G., Introduction to the Theory of Neural Computation, 1991, Addison-Wesley, Redwood City, CA.

[16] Jelonek, J., Krawiec, K., and Slowinski, R., Rough Set Reduction of Attributes and their Domains for neural Networks, *Computational Intelligence*, 11(2), 1995, pp. 339-347.

[17] Kasey, K., McGuinness, P., and Short, H., Multilogit Approach to Predicting Corporate Failure Further Analysis and the Issue of Signal Consistency, *Omega*, 18(1), 1990, pp. 85-94.

[18] Lee, K, Han, I., and Kwon, Y., Hybrid Neural Network Models for Bankruptcy Prediction, *Decision Support Systems*, 18, 1996, pp. 63-72.

[19] Messier, W.F., and Hansen, J.V., Inducing Rules for Expert System Development: An Example Using Default and Bankruptcy Data, *Management Science*, 34(12), 1988, pp. 1403-1415.

[20] Odom, M., and Sharda, R., A Neural Network Model for Bankruptcy Prediction, Proceedings of the IEEE International Conference on Neural Network, 2, 1990, pp. 163-168.

[21] Pawlak, Z., Rough Sets, *International Journal of Information and Computer Sciences* 11, 1982, pp. 341-356.

[22] Pawlak, Z., Grzymala-Busse, J., Slowinski, R., and Ziarko, W., Rough Sets, *Communications of ACM*, 38(11), 1995, pp. 89-95.

[23] Raghupathi, W., Schkade, L., and Raju, B.S., A Neural Network Application for Bankruptcy Prediction, IEEE International Joint Conference on Neural Networks, 1991, pp. 147-155.

[24] Rochester, J.D., New Business Uses for Computing, *I/S Analyzer*, 28, 1990, pp. 1-16.

[25] Salchenberger, L.M., Cinar, E.M., and Lash, N.A., Neural Networks: A New Tool for Predicting Thrift Failures, *Decision Sciences*, 23, 1992, pp. 899-916.

[26] Siegel, P.H., de Korvin, A., and Omer, K., Detection of Irregularities by Auditors: A Rough Set Approach, *Indian Journal of Accounting*, 1993, pp. 44-56.

[27] Slowinski, R., and Stefanowski, J., RoughDAS and RoughClass Software Implementations of the Rough Set Approach, in R. Slowinski (ed.) *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory*, Kluwer Academic Publishers, Dordrecht, 1992, pp. 445-456.

[28] Slowinski, R., and Stefanowski, J., Rough Classification with Valued Closeness Relation, *In:Diday, E. et al. (eds.), New Approaches in Classification and Data Analysis*, Springer-Verlag, Berlin, 1994, pp. 482-488.

[29] Slowinski, R., and Zopounidis, C., Application of the Rough Set Approach to Evaluation of Bankruptcy Risk, *International Journal of Intelligent Systems in Accounting, Finance and Management*, 4, 1995, pp. 27-41.

[30] Slowinski, R., Zopounidis, C., and Dimitras, A.I., Prediction of Company Acquisition in Greece by Means of the Rough Set

Approach, *European Journal of Operational Research*, 100(1), 1997, pp. 1-15.

[31] Tam, K.Y., and Kiang, M.Y., Managerial Applications of Neural Networks: The Case of Bank Failure Predictions, *Management Science*, 38(7), 1992, pp. 926-947.

[32] Trippi, R., and Turban, E., The Impact of Parallel and Neural Computing on Managerial Decision Making, *Journal of Management Information System*, 6, 1989, pp. 85-97.

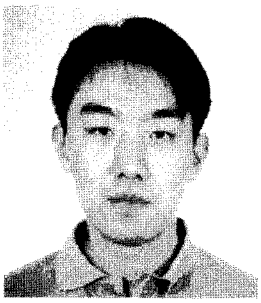
[33] Warner, J.B., Bankruptcy costs some evidence, *The Journal of Finance*, 32(2), 1977, pp. 337-347.

[34] Wasserman, P., *Neural Computing: Theory and Practice*, 1989, Van Nostrand Reinhold, New York.

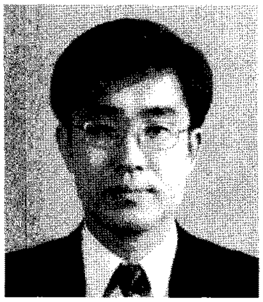
[35] Ziarko, W., Golan, R., and Edwards, D., An Application of DATALOGIC/R Knowledge Discovery Tool to Identify Strong Prediction Rules in Stock Market Data, *Proc. of AAAI Workshop on Knowledge Discovery in Databases*, Washington DC, 1993.

◈ 이 논문은 1999년 4월 26일 접수하여 1차 수정을 거쳐 1999년 9월 21일 게재확정되었습니다.

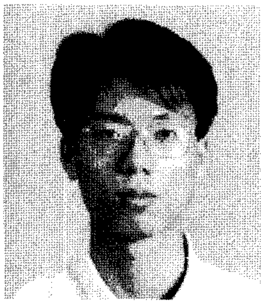
◆ 저자소개 ◆



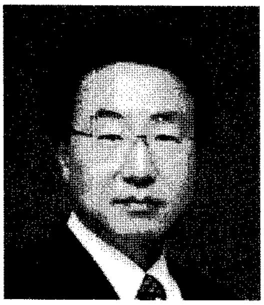
김창연 (Kim, Chang-Yun)
서울대학교 전자공학 학사
한국과학기술원 경영공학 석사
현재 SynTech㈜ 연구원
관심분야 : 기업 평가, Knowledge Based Decision Support System



안병석 (Ahn, Byeong-Seok)
연세대학교 응용통계학 학사
한국과학기술원 경영과학 석사
한국과학기술원 경영정보학 박사
현재 수원대학교 경영회계학부 교수
관심분야 : Evaluation of Information Systems, Knowledge Based Decision Support Systems, Multicriteria Decision Analysis



조성식 (Cho, Sung-Sik)
고려대학교 전산과학 학사, 한국과학기술원 경영공학 석사, 현재 한국과학기술원 경영공학 박사과정 재학
관심분야 : Data Mining, GDSS, Knowledge Based Decision Support System



김성희 (Kim, Soung-Hie)
서울대학교 공과대학을 졸업
University of Missouri-Columbia 석사
Stanford University 경영과학 박사
미국 Strategic Decision Group에서 컨설팅 연구원
University of Michigan 객원교수 역임
현재 KAIST 테크노경영대학원 교수로 재직 중이다.
관심분야 : 경영혁신/BPR, CALS/EC, GDSS, DA