Journal of the Korean
OR/MS Society
Vol. 22, No. 3, September 1997

The Hybrid Systems for Credit Rating*

Ingoo Han** · Hongkyu Jo** · Kyung Shik Shin**

Abstract

Although numerous studies demonstrate that one technique outperforms the others for a given data set, it is hard to tell a priori which of these techniques will be the most effective to solve a specific problem. It has been suggested that the better approach to classification problem might be to integrate several different forecasting techniques by combining their results.

The issues of interest are how to integrate different modeling techniques to increase the predictive performance. This paper proposes the post-model integration method, which tries to find the best combination of the results provided by individual techniques. To get the optimal or near optimal combination of different prediction techniques, Genetic Algorithms (GAs) are applied, which are particularly suitable for multi-parameter optimization problems with an objective function subject to numerous hard and soft constraints.

This study applies three individual classification techniques (Discriminant analysis, Logit model and Neural Networks) as base models for the corporate failure prediction. The results of composite predictions are compared with the individual models. Preliminary results suggest that the use of integrated methods improve the performance of business classification.

1. Introduction

Prediction of corporate failure using past financial data is a well-documented topic. Collins and Green [4], Altman [1], Jones [8] and Boritz [3] provide comparisons of major methods and the empirical results of these methodologies. In particular, a number of recent studies have demonstrated

* This paper was supported by NON DIRECTED RESEARCH FUND, Korea Research Foundation.
** Graduate School of Management, Korea Advanced Institute of Science and Technology

that Neural Networks can be an alternative methodology for classification problems to which traditional statistical method have long been applied [2].

Although these studies demonstrate that one technique outperforms the others for a given data set, it is hard to tell a priori which of these techniques will be the most effective to solve a specific classification problem. Thus, we may try several different techniques and select one that seems to provide the most accurate results for the specific problem. Alternatively, it has been suggested that the better approach to classification problem might be to integrate several different forecasting techniques by combining their results [7][10].

The remainder of this paper is organized as follows. The first section provides the description of problem; the organization of the sample cases, basic concepts and models of three classification methods (Discriminant Analysis, Logit, Neural Networks) which are used as individual prediction techniques. We then present a new approach, the composite prediction method, to integrate different techniques using Genetic Algorithms (GAs). The following section contains a general description of GAs. The fourth section presents the results of experiments. The final section discusses further research issues.

2. Problem Description

2.1 Data and Variable

The sample consists of the same number of bankrupt and non-bankrupt cases. The data set contains 1,274 companies which filed for bankruptcy during the period 1993-95 (239, 322, and 713 bankruptcy cases for 1993, 1994, and 1995 respectively). To select the non-bankrupt firms, pair matching method by industry is used. Table 1 shows the organization of data set.

The financial ratios are used as the independent variables. The whole set of independent variables consists of 67 ratios in six categories; stability, profitability, growth, cash flow, activity, and productivity. This study uses two methods for selection of input variable. One is to select the input variables which have the significant correlation with output variable (bankrupt/non-bankrupt). The other is statistical stepwise method. Independent variables of this experiment consist of 11 financial ratios, industry type, the number of years after establishment of firm, and the size of firm.

Table 1 Industry classification of data set

Industry	1993 Bankrupt	1993 Non-Bankrupt	1994 Bankrupt	1994 Non-Bankrupt	1995 Bankrupt	1995 Non-Bankrupt
1	56	56	82	82	145	145
2	61	61	77	77	188	188
3	54	54	81	81	152	152
4	36	36	43	43	144	144
5	12	12	21	21	41	41
6	20	20	18	18	43	43
Sum	239	239	322	322	713	713
Total	1,274 bankrupt firms/1,274 non-bankrupt firms					

2.2 Discriminant Analysis

Discriminant analysis (DA) is a representative of statistical classification methods. DA consists of three steps. The first step is estimating the coefficients of variables, the second step is calculating the discriminant score of each case, and the third step is classifying the case.

Linear discriminant function is as follow.

$$D = B_0 + B_1X_1 + B_2X_2 + \cdots + B_pX_p \tag{1}$$

where D is a discriminant score, B_0 is a estimated constant, B_n are the estimated coefficients, X_n are the independent variables of a case.

The probability that a case with a discriminant score of D belongs to group i is estimated by following equation.

$$P(G_i \mid D) = \frac{P(D\backslash G_i)P(G_i)}{\sum_{i=1}^{g} P(D\backslash G_i)P(G_i)} \tag{2}$$

The prior probability, represented by P(G_i) is an estimate of the likelihood that a case belongs to a particular group when no information is available. The prior probability can be estimated in observed proportions of cases in each group.

2.3 Logistic regression (Logit)

The regression model implicitly assumes that the dependent variable is continuous. But the general regression model is not useful for classification problems, because classification has categorical dependent variable. Logistic regression is a method to solve the discrete dependent variable problem. The model is shown below.

$$P_i = E(Y=1 \mid X_i) = \frac{1}{1+\exp^{-(\beta_0+\beta_1 X_1+\cdots+\beta_n X_n)}} \tag{3}$$

In the above equation, when the dependent variable of problem is binary, P_i or $E(Y=1 \mid X_i)$ is the expectation when the dependent variable is 1 in case X_i. P_i is known as the (cumulative) logistic distribution function, so this method is named as *logistic regression.*

2.4 Neural network

Neural network is the result of human effort to make a machine which has a reasoning process similar to the human brain. But we do not understand how the brain works or what intelligence really is [11]. In many theoretical and experimental studies, the usefulness of neural network was supported in classification studies.

It is an art to find an appropriate neural network model which can reflect problem characteristics because there are numerous NN architecture, learning methods, and parameters. Multi-layered feed-forward network has been frequently used. This study employed the 3-layered, fully connected, feedforward network with one hidden layer.

Generalized delta rule is used in this study. There is no offset added to the derivative of the transfer function such as η. This study uses the common asymmetric sigmoid and the common quadratic error function. The value of raw input data ranges continuously with different scales. The values of input node are standardized to be from 0 to 1.

3. Integrated Approach

Some studies have suggested integrated methods to increase the classification ability. Lee et. al. [9] introduced the new approach to integrate the results of inductive learning and neural network.

Composite prediction is to find the optimal combining weights of predicted outputs. Predicted outputs of each classification method are used as input. DA, Logit, and NN are used to generate the predicted outputs. Weight adjusting rule is represented as the following nonlinear programming problem.

$$\text{Minimize} \qquad CP = \sum_{i=1}^{n} (T_i - O_i)^2 \tag{4}$$

$$\text{s.t.} \qquad O_i = f\left\{ \sum_{j=1}^{m} W_i V_{ij} \right\}$$

$$f(X) = \begin{cases} 0 & X \le 0.5 \\ 1 & X > 0.5 \end{cases}$$

$$T_i, O_i = 0 \text{ or } 1$$

$$0 \le W_i, V_{ij} \le \ , \qquad \text{real value}$$

where

CP = square root error between target output T_i and O_i, hit ratio is calculated by CP, hit ratio is 100* $(1-(CP/n))$ when n is number of cases (patterns)

T_i = target output of ith data

O_i = translated value of weighted sum of predicted output of ith data by threshold function f

V_{ij} = predicted output of ith data and jth classification method

W_j = weight of jth classification method

f = threshold function to generate binary outputs

The values of V_{ij} have the same range [0,1] in order to apply the above nonlinear programming. NN produces the output with continuous value from zero to one, while DA and Logit do not. Therefore the outputs of DA and Logit are transformed to values of the range [0,1]. The objective function is to minimize the misclassification error shown in the equation (4). In the equation (4), the dependent variable T_i is binary. The function f is used for transforming the continuous value to binary value.

The other issue is to solve the above nonlinear programming. It has the step function to transform the weighted sum of predicted outputs of each method to binary value. Most algorithms for nonlinear programming use derivatives of objective function which includes the decision variables of the problem. The problem of this study is to be solved for the values of the variables w_j. It is hard to transform the step function f to the continuous equation. Thus we apply the genetic algorithm to this nonlinear programming. The constraints are used as the conditions of genetic algorithms.

4. Genetic Algorithms Methodology

Genetic Algorithms are stochastic techniques that can search large and complicated spaces. Based on genetic and evolutionary principles, GAs work by repeatedly modifying a population of artificial structures through the application of initialization, selection, crossover and mutation operators. They have been demonstrated to be effective and robust in searching very large spaces in a wide range of applications [6]. GAs are particularly suitable for multi-parameter optimization problems with an objective function subject to numerous hard and soft constraints.

In the initialization stage, a population of genetic structures that are randomly distributed in the solution space is selected as the starting point of the search. After the initialization stage, each structure is evaluated using a user-defined fitness function and assigned a utility value. In our optimization problems where we maximize hit ratio, the closer the score is to 1, the higher the score. The mating convention is such that only high scoring members will preserve and propagate their "worthy" characteristics from generations to generation and thereby help in continuing the search for an optimal solution.

Selection of parents for mating involves choosing some member with high scores by a "roulette wheel" approach and choosing other member randomly. The selected members are recombined using crossover, which operates by swapping corresponding segments of a string representation of the parents. Crossover serves two complementary search functions. First, it provides new points for further testing of schemata already presented in the population. Second, it introduces the instances of new schema into the population. A schema is a subset of strings which have similarities at certain positions. The crossover operation is repeated on the parents until at least one offspring is feasible.

Mutation is a GAs mechanism by which we randomly choose a member of the population and change one randomly chosen bit in its bit-string representation. If the mutant member is feasible, it replaces the member which was mutated in the population. The presence of mutation ensures that the probability of reaching any point in the search space is never zero.[5]

The weights of independent techniques are generated by the computer package "EVOLVER". The initial population is fixed to 50 and the crossover rate and mutation rate are changed to prevent the output from falling into the local optima. The crossover rate is ranged 0.5 ~ 0.6, and mutation rate is ranged 0.06 ~ 0.1.

5. Results

The results obtained by genetic algorithms are compared with results from individual classification techniques to study the effectiveness of integrated approach for the corporate failure prediction. Table 2 presents the hit ratios of the these models for training and holdout samples. The result of holdout sample represents the hit ratio when the weights of being searched in the training sample are applied to holdout sample.

For each set of data (bankrupt and non-bankrupt), the training subset consists of 90% of data

Table 2 Classification accuracy (Hit ratios: %)

Experiment set		DA	Logit	NN	Composit prediction	
						Weights (DA, Logit, NN)
Set 1	Training	76.4	77.1	79.0	79.4	(.053, .013, .934)
	Holdout	73.2	73.6	75.6	76.0	
Set 2	Training	76.3	77.1	79.3	79.8	(.208, .177, .615)
	Holdout	72.0	73.0	74.0	74.0	
Set 3	Training	75.8	76.5	78.7	79.1	(.181, .025, .794)
	Holdout	72.8	74.4	74.8	77.2	

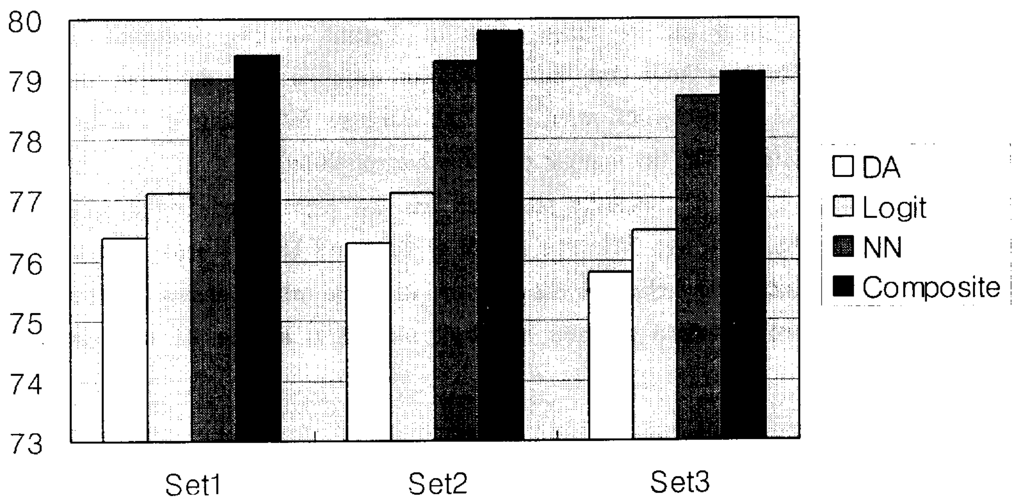

Figure 1 Classification accuracy of the training sample (%)

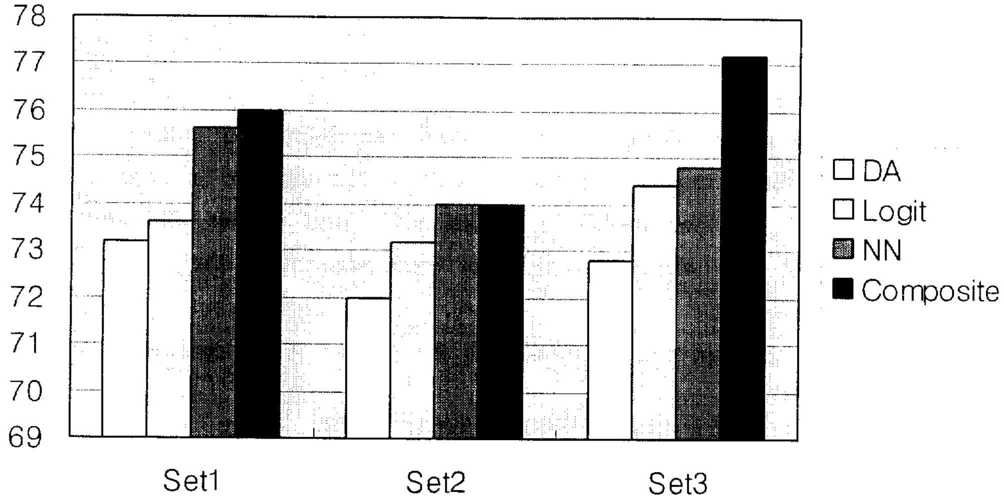

Figure 2 Classification accuracy of the validation sample (%)

while testing subset consists of the remaining 10%. We randomly replicate the selection of data subset three times (set 1, 2, and 3). Estimation and testing process are applied for each of three data sets. Among individual techniques, the neural networks have the highest level of accuracy, followed by Logit and Discriminant analysis.

We generate near optimal combining weights of three prediction outputs for the composite prediction using GAs. The integrated approach of composite prediction has a better predictive performance than any other individual techniques.

Table 3 shows the results of Cochran's Q tests to compare the predictive ability between three models and composite prediction for three kinds of holdout sample. Cochran's Q test is a nonparametric test of the hypothesis that the dichotomous variables have the same means. We are interested in the correct prediction of cases, thus the measure for testing is the hit ratios (the number of correct prediction from the number of whole holdout samples).

As shown in figure 1 and 2, composite prediction has the higher prediction accuracy than the individual method of DA, logit, and NN. The results of Cochran's Q tests show that the composite prediction outperforms individual techniques significantly. Also, the table 3 shows that neural network is more useful than two statistical methods and logit method is a little bit more accurate than discriminant analysis.

Table 3 Nonparametric (Cochran Q) test between DA, Logit, NN, and composite prediction

set 1

	DA Cochran Q (significance)	Logit Cochran Q (significance)	NN Cochran Q (significance)	Composite Cochran Q (significance)
DA	-	1(.3173)	6(.0143)*	7(.0082)**
Logit	-	-	5(.0253)*	6(0.0143)*
NN	-	-	-	1(.3173)
Composite	-	-	-	-

Set 2

	DA Cochran Q (significance)	Logit Cochran Q (significance)	NN Cochran Q (significance)	Composite Cochran Q (significance)
DA	-	3(.0833)	5(.0253)*	5(.0253)*
Logit	-	-	2(.1573)	2(.1573)
NN	-	-	-	0(1.0000)
Composite	-	-	-	-

set 3

	DA Cochran Q (significance)	Logit Cochran Q (significance)	NN Cochran Q (significance)	Composite Cochran Q (significance)
DA	-	4(.0455)*	5(.0253)*	11(.0009)**
Logit	-	-	1(.3173)	7(.0082)**
NN	-	-	-	6(.0143)*
Composite	-	-	-	-

* significant at 0.05 level

* significant at 0.01 level

6. Conclusion

This study proposes the post-model integration method by finding the best combination of the results of each method. The measure of misclassification cost was used to integrate the three different classification techniques. We modified the objective function of mathematical model to apply the measure of misclassification cost.

We applied genetic algorithms to search optimal set of weights among different prediction models. The empirical results demonstrate that the use of integrated methods will offer improved performance in business classification problems.

The limitation of this study is insufficient experimental simulations to compare the predictive performance statistically, although we replicated three times to reduce the instability due to sampling error.

References

[1] Altman, E., ***Corporate financial distress: A complete guide to predicting, avoiding, and dealing with bankruptcy***, John Wiley, New York, 1983.

[2] Boritz, J.E., D. B. Kennedy and A. Albuquerque, "Predicting Corporate Failure Using a Neural Network Approach," ***Intelligent Systems in Accounting, Finance and Management***, Vol. 4 (1995), pp. 95-111.

[3] Boritz, J.E., ***The going concern assumption***, Canadian Institute of Chartered Accountants, Toronto, 1991.

[4] Collins, R.A., & R. D. Green, "Statistical methods for bankruptcy forecasting," ***Journal of Economics and Business***, Vol. 32, (1982), pp. 349-354.

[5] Deboeck, G.J., ***Trading On The Edge*** (ed), 1994.

[6] Goldberg, D.E., ***Genetic algorithms in search, optimization and machine learning***, Addison-Wesley, Reading, MA, 1989.

[7] Jo, H., & I. Han, "Integration of case-based forecasting, neural network and discriminant analysis for bankruptcy prediction," ***Accepted in Expert Systems With Applications***, Vol. 11 (1996).

[8] Jones, F., "Current techniques in bankruptcy prediction," ***Journal of Accounting Literature*** (1987), pp. 131-164.

[9] Lee, K.C., H. Kim and M. Kim, "An inductive learning-assisted neural network approach to bankruptcy prediction: Comparison with MDA, inductive learning, and neural network models," ***Korean Management Study*** (1995).

[10] Liang, T., J. Chandler and I. Han, "Integrating Statistical and Inductive Learning Methods for Knowledge Acquisition," ***Expert Systems With Applications***, Vol. 5 (1990), pp. 391-401.

[11] Markham, I.S. and C. T. Ragsdale, "Combining neural networks and statistical predictions to

solve the classification problem in discriminant analysis," ***Decision Science***, Vol. 26 (1995), pp. 229-242.

[12] Nelson M.M and W. T. Illigworth, ***A practical guide to neural networks,*** Addison Wesley, Reading, Massachusetts, 1990.