Integration Methodology of Multiple Techniques Using Genetic Algorithms : A Case of Corporate Failure Prediction

Kyung-shik Shin, Hongkyu Jo and Ingoo Han

Graduate School of Management, Korea Advanced Institute of Science and Technology,
207-43 Cheongryangri-Dong, Dongdaemoon-Gu, Seoul, 130-012, Korea.
Telephone : 02-958-3673, Fax: 02-958-3604, e-mail: ingoohan@msd.kaist.ac.kr

ABSTRACT

Although numerous studies demonstrate that one technique outperforms the others for a given data set, there is often no way to tell a priori which of these techniques will be most effective to solve a specific problem. Alternatively, it has been suggested that a 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 prediction performance. This paper proposes the post-model integration method, which means integration is performed after individual techniques produce their own outputs, by finding the best combination of the results of each method. 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 applied three individual classification techniques (Discriminant analysis, Logit and Neural Networks) as base models to the corporate failure prediction context. Results of composite prediction were compared to the individual models. Preliminary results suggests that the use of integrated methods will offer improved performance in business classification problems.

* This paper was supported by NON DIRECTED RESEARCH FUND, Korea Research Foundation.

INTRODUCTION

A number of ways of determining the potential for corporate failure using past financial data is a well-documented topic. The solution to this problem is a discriminant function from the variable space in which observations are defined into a binary set.

Collins and Green[4], Altman[1], Jones [8] and Boritz[3] provide comparisons of major methods and the empirical test results of these methodologies. In particular, a number of recent studies have demonstrated 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, there is often no way to tell a priori which of these techniques will be most effective to solve a specific classification problem. Thus, an user might 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 a 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 basic concepts and models of three classification methods (Discriminant Analysis, Logit, Neural Networks). 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 results of experiments and the final section discusses further research issues.

PROBLEM DESCRIPTION

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-bankruptcy firms, pair matching method by industry is used. Table 1 shows the organization of data set.

Table 1 Industry classification of data set

	1993		1994		1995	
IND	B	N-B	B	N-B	B	N-B
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					

(B:Bankrupt / N-B:Non-bankrupt)

The financial ratios are used as the indicators (independent/input variables). The whole set of independent variables consists of 67 ratios in six categories; stability, profitability, growth, cash flow, activity, and productivity. This study used two methods for input variable selection, the one is to select the input variables which have the larger relationship to 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 size of firm.

Discriminant Analysis

Discriminant analysis 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.

$$D = B_0 + B_1X_1 + B_2X_2 + \cdots + B_pX_p \quad (1)$$

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.

$$P(G_i|D) = \frac{P(D|G_i)P(G_i)}{\sum_{i=1}^{s} P(D|G_i)P(G_i)} \quad (2)$$

The prior probability, represented by P(Gi) 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.

Logistic regression (Logit)

The regression model implicitly assumes that the dependent variable is continuous. The general regression model is represented by

$$Y_i = \beta_0 + \beta_1X_1 + \cdots + \beta_nX_n \quad (3)$$

Y_i is continuous dependent variable, X_i is independent variable, and β_i is coefficient estimate. But this equation is not useful for classification problems, because classification has categorical dependent variable. Logistic regression is a kind of method to solve the discrete dependent variable problem. It is shown below.

$$P_i = E(Y = 1|X_i) = \frac{1}{1 + \exp^{-(\beta_0 + \beta_1X_1 + \cdots + \beta_nX_n)}} \quad (4)$$

In the above equation, when the dependent variable of problem is binary, P_i or $E(Y = 1|X_i)$ 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.*

From the above equation, interesting terminology L_i is generated, which is similar to general regression equation. L_i is called as the *logit* and represented as following.

$$L_i = \ln\left(\frac{P_i}{1 - P_i}\right) = Z_i = \beta_0 + \beta_1X_1 + \cdots + \beta_nX_n \quad (5)$$

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 in classification problem was proven.

Finding an appropriate neural network model which can reflect problem characteristics is an art because there are numerous NN architecture, learning methods, and parameters. Multi-layered feedforward network has been frequently used. This study employed the 3-layered, fully connected, feedforward network which has 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 used the common asymmetric sigmoid, and the common quadratic error function. The value of raw input data are ranged continuously with different scales. The values of input node are standardized to be from 0 to 1.

INTEGRATED APPROACH

Some studies have suggested new 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 optimal combining weights of predicted output. Predicted outputs of each classification method are used as input, in this study DA, Logit, NN are used to generate the predicted output. Weight adjusting rule is represented as following nonlinear programming.

Minimize $$\mathrm{CP} = \sum_{i=1}^{n} (T_i - O_i)^2 \quad (6)$$

s.t. $$O_i = f\left(\sum_{j=1}^{m} w_j V_{ij}\right)$$

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

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

$$0 \le w_j, V_{ij} \le 1, \quad \text{real value}$$

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 Vij have the same range [0,1] in order to apply the above nonlinear programming to a problem. NN produce 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 [0,1].

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 of solving methods of nonlinear programming use derivatives of objective function which includes variables of problem. The problem of this study is solved for the values of the variables w_j. It is hard to transform the step function f to continuous equation, then we apply the genetic algorithm to this nonlinear programming. The constraints are used as the conditions of genetic algorithm.

GENETIC ALGORITHMS METHODOLOGY

Genetic Algorithms are stochastic search 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 application [6]. A GA is 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 in member from the high scores by a "roulette wheel" approach and choosing the 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 present in the population and, secondly, it introduces instances of new schema into the population. A schema is a subset of strings with similarities at certain positions. The crossover operation is repeated on the parents until at least one offspring is feasible.

Mutation is a GA mechanism where 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]

RESULTS

To study the effectiveness of integrated approach for the corporate failure prediction, the results obtained by applying genetic algorithms are compared with results from individual classification techniques. Table 2 presents the comparison of the results of the these models.

Table 2 Classification accuracy

Experiment Set		DA	Lo-git	NN	Composite Prediction	
						Weights (D,L,N)
Set 1	Tr.	76.4	77.1	79.0	79.4	(.053,.013, .934)
	Ts.	73.2	73.6	75.6	76.0	
Set 2	Tr.	76.3	77.1	79.3	79.8	(.208,.177, .615)
	Ts.	72.0	73.0	74.0	74.0	
Set 3	Tr.	75.8	76.5	78.7	79.1	(.181,.025, .794)
	Ts.	72.8	74.4	74.8	77.2	

(Tr:Training set / Ts:Test set)

For each set of data set(bankrupt and non-bankrupt), a training subset and testing subset, consisting of 90% and 10% of the data, respectively, are randomly selected. We replicate three times(Set1,2 and 3) of data set selection, estimation and testing process to reduce the impact of random variation in data set composition. Among individual techniques, the neural networks have the highest level of accuracy in the given data sets, followed by Logit and Discriminant analysis.

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

CONCLUSION

This study proposes the post-model integration method by finding the best combination of the results of each method. 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 significant instability due to sampling error.

REFERENCE

[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., Kennedy D.B. and Albuquerque A., "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 Account-ants, Toronto,1991.

[4] Collins, R.A. and Green, R.D., "Statistical Methods for Bankruptcy Forecasting," Journal of Economics and Business, vol.32, 1982, pp.349-354.

[5] Deboeck, G.J. ed, Trading On The Edge, 1994, pp.245-247.

[6] Goldberg, D.E., "Genetic Algorithms in Search, Optimization and Machine Learning," Reading, MA:Addison-Wesley, 1989.

[7] Jo. H. and Han, I., "Integration of Case-based forecasting, Neural Network and Discriminant Analysis for Bankruptcy Prediction," Proc. Of the 3rd WCES, 1996.

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

[9] Lee, K.C., Kim, H. and Kim, M., "An Inductive Learning-Assisted Neural Network Approach to Bankruptcy Prediction: Comparison with MDA, Inductive Learning, and Neural Network Models," Korean Management Study, 1995.

[10] Markham, I.S and Rafsdale, C.T., "Combining Neural Networks and Statistical Predictions to Solve the Classification Problem in Discriminant Analysis", Decision Science, vol 26, 1995.

[11] Nelson M.M and Illigworth W.T., A Practical Guide to Neural Neworks, 1990.