한국지능정보시스템학회논문지 제9권 제1호 2003년 6월(pp.71~89)

Aggregating Prediction Outputs of Multiple Classification Techniques Using Mixed Integer Programming

Hongkyu Jo
Financial Engineering Research Center,
NICE Pricing Services
(hongkyujo@hotmail.com)

Ingoo Han
Graduate School of Management, Korea
Advanced Institute of Science and Technology
(ighan@kgam.kaist.ac.kr)

Although many 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 in the classification problems. Alternatively, it has been suggested that a better approach to classification problem might be to integrate several different forecasting techniques.

This study proposes the linearly combining methodology of different classification techniques. The methodology is developed to find the optimal combining weight and compute the weighted-average of different techniques' outputs. The proposed methodology is represented as the form of mixed integer programming.

The objective function of proposed combining methodology is to minimize total misclassification cost, which is the weighted-sum of two types of misclassification. To simplify the problem solving process, cutoff value is fixed and threshold function is removed.

The form of mixed integer programming is solved with the branch and bound methods. The result showed that proposed methodology classified more accurately than any of techniques individually did. It is confirmed that proposed methodology predicts significantly better than individual techniques and the other combining methods.

Key words: Bankruptcy prediction, artificial intelligence, hybrid prediction, integer programming

논문접수일 : 2002년 10월 게재확정일 : 2003년 3월 교신저자 : 조홍규

1. INTRODUCTION

The phenomenon of corporate failure is not infrequent nor of minor economic consequence. In United States, during the period 1980 to 1989, the number of bankruptcy filings tripled, approaching quarters of a million by that decade's end. In Korea, during the same period, the number of bankruptcy filings increased ten times. Especially, the number of bankrupt companies increases extremely this year under the influence of IMF (International Monetary Fund)'s control. It is obvious that such a crucial issue in corporate finance warrants careful investigation (Cadden, 1991).

Bankruptcy prediction is one of the business classification problems. Business classification covers wide areas of managerial decision-making. It includes bankruptcy prediction, accounting method choice, audit opinion decision, bank loan classification, and bond rating, etc.

Classification, classifying objects or observations into distinct previously defined groups, has been an important domain of traditional business research. The ultimate goal of classification is to provide the relevant outcome or to replicate the expert's judgment. The algorithm to separate and allocate objects or observations is called a classification technique.

During past decades, many classification techniques have been developed. Until the first half of 1980s, statistical methods have been mainly used for classification problems. In the domain of business classification, univariate or multivariate discriminant analysis, probabilistic regression or logistic regression have been applied (Beaver, 1966; Altman, 1968).

From the last half of 1980s, artificial intelligence methods, especially inductive learning, have been employed for various domains of business classification such as stock market prediction, scholarship and fellowship grant cases, and bankruptcy prediction (Braun & Chandler, 1987; Garrison & Michalson, 1989; Lee & Oh, 1990).

Quite recently, neural networks have been applied to classification problems including bankruptcy prediction, management forecasting, bond rating, stock price prediction, and extraction of accounting knowledge (Surkan & Singleton, 1990; Trigueiros, 1994; Jhee & Lee, 1993). Many researches have proved the effectiveness of neural networks in business classification problems (Odom & Sharda, 1990; Raghupathi, Schkade, & Raju, 1993; Tam & Kiang, 1992; Park & Han, 2002; Shin & Lee, 2002).

Although many 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, a 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 (Jo & Han, 1996; Markham & Ragsdale, 1995).

A number of studies compared various classification techniques and demonstrated that one technique was a dominant solver for a specific problem. Recent many studies showed that artificial intelligence techniques were superior to the traditional statistical techniques, but lots of previous studies still reached the contradictory results that artificial intelligence techniques did not outperform even was inferior to the traditional techniques (Fanning & Cogger, 1994; Altman, Marco, & Varetto, 1994).

As an alternative of finding proper

classification methods for a given problem, several types of integrated methodology have been suggested. Same approaches that integrate several techniques also have been studied for various forecasting problems (Liang, Chandler, & Han, 1990; Jhee & Lee, 1993).

One of the interesting combining approaches is to aggregate the results of different techniques with the linear equation. The method of linearly combining the results of different techniques has been widely used for finding more accurate solutions in the various forecasting or classification areas. One of the most popular forecasting techniques, ARIMA is the linear combination of auto regressive model and moving average model. Troutt, Rai, & Tadisina (1997) suggested the linear equation model for aggregating the results of multiple experts and validated it with the small example of credit approval task. Mostaghimi (1997) also developed weighted linear combination method, which used rank order and applied for predicting U. S. hog prices. Almost previous studies using the linear combination methodology have shown that the performance of combined approaches have been higher than those of individual forecasting techniques.

This study proposes the combining methodology of different classification techniques. Basically, the methodology is developed to find the optimal combining weight and generate the weighted-average of different techniques' outputs for a given data set. The proposed methodology is converted to the form of mixed integer programming for convenient computation of methodology.

Also, suggested methodology is supported by the results of experiment. Experiments that are included in this paper compare the results of suggested methodology with the performances of individual techniques and other combining methodologies. In addition, the experimental results help to understand the logical deployment process of the methodology.

This paper is organized as follows. Chapter 2 introduces the development process of integrated methodology in detail. Basic concept of linear combining methods and some examples are introduced also. The optimal combining methodology suggested in this paper is displayed as the mathematical form.

Sample application of this methodology to predicting bankruptcy is followed in chapter 3. The mathematical programming is relaxed by some assumptions and solved by the traditional "branch and bound" method.

Contribution and limitation of research and future research issues are discussed in final chapter.

2. METHODOLOGY

Consider an opinion aggregation problem in which a group of experts express their opinions on the outcome of a random event, and one desires to combine their opinions into a single aggregated opinion. Suppose a situation of a decision-maker that receives various outputs from efficient problem-solving models, faces the conflict between results generated by models, and wants to make single decision fundamentally. Above two problems are very resembling, but some differences are still exist between two problems.

The identical aspect of two problems is to extract final single solution by combining the various alternative solutions. The underlying idea here is that the consensus of numerous experts results in a superior forecast or prediction than any of the experts considered individually (Myung, Ramamoorti, & Bailey, 1996). Many studies have tried to find the best way of aggregating or combining the outputs of various techniques.

Techniques such as Delphi and Nominal Groups are often used to arrive at a consensus in the group decision-making situations. Such processes are suitable for highly unstructured decisions, because they converge on final reasonable solution by repeating discussion or interaction. However, it is hard to adopt such iterative process to the model of combing multiple forecasting methods. The main difference between aggregating human expertise and combining forecasting techniques is from the possibility of employing iterative procedure.

The examples of aggregating problems are credit applicant approval decision (Troutt, Rai, & Tadisina, 1997), binary predictions (Myung, Ramamoorty, & Bailey, 1996), and forecasting U.S. hog price (Mostaghimi, 1996). Some previous researches have paid attention to the area of combining various forecasting techniques. The problems are classifying products into five types (Kumar & Ormeda, 1997), the task of name pronunciation (Golding & Rosenbloom, 1996), and bankruptcy prediction (Jo & Han, 1996; McKee & Lensberg 2002). This chapter develops and implements the integrated model of various classification methods.

A number of studies showed that the combined forecasts turned out to be a linear weighted average, when the numerous factors are considered (Mostaghimi, 1996). One of the most popular forecasting techniques, ARIMA is the linear combination of auto regressive model and moving average model. Mathematical programming approaches to classification problems mostly use the weighted-average structure.

The types of linear combining methods are determined and classed by two main elements of prediction performance and prediction certainty. Prediction performance is measured by the historical results in general. Prediction certainty is the measure to represent the degree of confidence. The data of table 1 is the small example of decision results performed by three multiple experts. Several types of combining policies are compared using the data of table 1.

The data of table 1 contain 'multi-way contradictory' data. For instance, in table 1, cases 5, 15, and 21 are identical. However the first two loan officers decided rejection, while the third decided acceptance. Other overlap cases are also possible, some of which are reflected in table 1. For instance, see cases 8, 14, and 16, cases 9, 11, and 22, and cases 7, 10, and 17, respectively.

The output value determines the decision results about each case. The example of table 1

<Table 1> Credit applicant acceptance-rejection data

Loan officer	Case	Decision	X1	X2	X3	Output	Accuracy
1 (78%)	1	Reject	0	0.78	0.84	0.40	Correct
	2	Accept	0.68	1.00	0.76	0.55	Correct
	3	Accept	0.64	0.71	0.74	0.94	Correct
	4	Accept	1.00	0.70	0.81	0.84	Correct
	5	Reject	0.57	0.49	0.59	0.40	Correct
	6	Accept	0.80	0.83	0.75	0.51	Correct
	7	Accept	0.65	0.70	0.80	0.80	Incorrect
	8	Accept	0.74	0.71	0	0.83	Correct
	9	Reject	0.80	0.56	1.00	0.04	Incorrect
2 (67%)	10	Accept	0.65	0.70	0.80	0.96	Incorrect
	11	Accept	0.80	0.56	1.00	0.90	Correct
	12	Accept	0.98	0.74	0.83	0.95	Correct
	13	Reject	0.36	0.59	0.64	0.18	Correct
	14	Reject	0.74	0.71	0	0.07	Incorrect
	15	Reject	0.57	0.49	0.59	0.02	Correct
3 (71%)	16	Accept	0.74	0.71	0	0.98	Correct
	17	Accept	0.65	0.70	0.80	0.68	Incorrect
	18	Reject	0.40	0	0.87	0.22	Correct
	19	Accept	0.79	0.69	0.81	0.97	Correct
	20	Accept	0.84	0.70	0.80	0.54	Correct
	21	Accept	0.57	0.49	0.59	0.67	Incorrect
	22	Accept	0.80	0.56	1.00	0.62	Correct

Source: Revised from "Troutt, M. D., A. Rai, & S. K. Tadisina. (1997). Aggregating multiple expert data for linear case valuation models using the MDE principle. Decision Support Systems, 21, 75-82."

assumes that the cutoff value to classify cases is 0.5, When the output is bigger than 0.5, the decision is *accept*. The output that is less then or equal to 0.5 represents the *reject* decision.

First type of combining policies is to transform the performances of individual experts or techniques into the weights of combining model. The performances of three loan officers are 78%, 67%, and 71% respectively. The combined results are computed as follows:

Combined output of cases 5, 15, and 21

$$= \frac{0\times 0.78+0\times 0.67+1\times 0.71}{0.78+0.67+0.71} = 0.318704 \quad \rightarrow \textit{Reject}$$

Combined output of cases 7, 10, and 17

$$= \frac{1\times 0.78+1\times 0.67+1\times 0.71}{0.78+0.67+0.71} = 1 \quad \rightarrow \textit{Accept}$$

Combined output of cases 8, 14, and 16

$$= \frac{1\times 0.78+0\times 0.67+1\times 0.71}{0.78+0.67+0.71} = 0.689815 \quad \rightarrow \textit{Accept}$$

Combined output of cases 9, 11, and 22

$$= \frac{0\times 0.78+1\times 0.67+1\times 0.71}{0.78+0.67+0.71} = 0.638889 \quad \rightarrow \textit{Accept}$$

Above computation process does not consider the continuous output of each loan officer but use the binary decision result. Troutt, Rai, & Tadisina (1997) developed *Maximum Decisional Efficiency Principle* that is based on the concept of this policy. They used the binary decision result and the number of decision making that each expert is involved.

The second combining policy is to average the value of outputs of individual experts or techniques. The combined results for overlapping cases are computed as follows:

Combined output of cases 5, 15, and 21

$$= \frac{1}{3} \times (0.4 + 0.02 + 0.67) = 0.363333 \quad \rightarrow \textit{Reject}$$

Combined output of cases 7, 10, and 17

$$= \frac{1}{3} \times (0.8 + 0.96 + 0.68) = 0.813333 \quad \rightarrow \textit{Accept}$$

Combined output of cases 8, 14, and 16

$$= \frac{1}{3} \times (0.83 + 0.07 + 0.98) = 0.626667 \quad \rightarrow \textit{Accept}$$

Combined output of cases 9, 11, and 22

$$= \frac{1}{3} \times (0.4 + 0.02 + 0.67) = 0.363333 \quad \rightarrow \textit{Accept}$$

Jo & Han (1996) integrated neural networks, discriminant analysis, and case-based reasoning with this policy. The result showed that the average of output is more accurate predictor than any of techniques predict individually. The method to select the output that has the maximum difference to cutoff in the outputs of various techniques is included in this type of combining policies.

Final combining policy is to use both prediction performance and certainty. The simplest example is to make the linear weighted function of outputs of various techniques with the transformed weight of prediction performance. The combined results are as follows.

Combined output of cases 5, 15, and 21

$$= \frac{0.4 \times 0.78 + 0.02 \times 0.67 + 0.67 \times 0.71}{0.78 + 0.67 + 0.71} = 0.37088 \quad \rightarrow \textit{Reject}$$

Combined output of cases 7, 10, and 17

$$= \frac{0.8 \times 0.78 + 0.96 \times 0.67 + 0.68 \times 0.71}{0.78 + 0.67 + 0.71} = 0.810185 \quad \rightarrow \textit{Accept}$$

Combined output of cases 8, 14, and 16

$$= \frac{0.83 \times 0.78 + 0.07 \times 0.67 + 0.98 \times 0.71}{0.78 + 0.67 + 0.71} = 0.643565 \quad \rightarrow \textit{Accept}$$

Combined output of cases 9, 11, and 22

$$= \frac{0.04 \times 0.78 + 0.9 \times 0.67 + 0.62 \times 0.71}{0.78 + 0.67 + 0.71} = 0497407 \quad \rightarrow \textit{Reject}$$

Most of integration studies hired this type of combining policies, regardless of what the problem is for aggregating experts or for integrating forecasting techniques (Myung, Ramamoorti, & Bailey, 1996; Mostaghimi, 1996; Han, Jo, & Shin, 1997). Table 2 demonstrates the results of three different policies comparing to original decisions of loan officers.

Our methodology is developed based on third policy, which is combining with performance and outputs. The methodology searches the weights of linear combing function with the mathematical programming model. The remaining part of this chapter explains the development procedure of our methodology in detail.

Suppose C is the combined output value when each classification techniques generate output value O. Then

<Table 2> Resolution of overlaps cases

Overlapping case sets	Loan officer 1	Loan officer 2	Loan officer 3	Combining with performance	Combining with outputs	Combining with performance and outputs	Actual decision
5, 15, 21	Reject	Reject	Accept	Reject	Reject	Reject	Reject
7, 10, 17	Accept	Accept	Accept	Accept	Accept	Accept	Reject
8, 14, 16	Accept	Reject	Accept	Accept	Accept	Accept	Accept
9, 11, 22	Reject	Accept	Accept	Accept	Accept	Reject	Accept

$$C_i = \sum_{j=1}^{k} W_j \times O_{ij}$$

where i = 1, 2, ..., n is the index of cases, and j = 1, 2, ..., k is the index of different methods used. Binary classification problems let the dependent variable have the value of 0 or 1 to represent the membership of a case in two different groups. If the output value of a case is in the range of (*cutoff*, 1], this case is to be classified into the group of membership "1". In the same manner, the value of range [0, *cutoff*) represents the membership of "0". As a result, output value O_{ij} is continuous in the range of [0,1].

Weight W_j is the coefficient of a linear equation to combine the outputs of k different methods, thus C_i becomes the continuous value of [0, 1] as a weighted average of O_{ij}. The combined output C_i and the actual membership of a case determine whether a case is classified correctly or not as table 3.

There are two types of misclassification as shown in table 3. The one is that the predicted membership is "1" and the actual membership is "0", the other is that the predicted and actual memberships are "0" and "1". In addition, two types of misclassification have the different level of risk. For instance, predicting a bankruptcy firm as a sound firm is more risky than the case that a sound firm is classified into the member of

<Table 3> Classification results by the combined output

```
IF      (the combined output value Ci of a case is in the range of [0, cutoff))   THEN
        IF      (the actual membership of a case is "0")   THEN Correct classification
        ELSE    (the actual membership of a case is "1")   Misclassification
ELSE    (the combined output value Ci of a case is in the range of (cutoff, 1])
        IF      (the actual membership of a case is "0")   THEN Misclassification
        ELSE    (the actual membership of a case is "1")   Correct classification
END IF
```

bankruptcy firms.

Suppose MC is the total misclassification cost, M_1 is the number of cases misclassified when the actual output is "1", and M_0 is the number of misclassification cases for actual membership of "0". Then

$$MC = r_0 M_0 + r_1 M_1$$

where r_0 is the risk of misclassification about group "0", and r_1 is the misclassification risk about group "1". The step function of f is assumed to count the number of misclassified cases. This function transforms the difference between the predicted value O_{ij} and *cutoff* value into the binary value of 0 or 1. The objective function is to minimize the misclassification cost MC. The problem is represented as follows.

Minimize $$MC = r_0 M_0 + r_1 M_1$$

s.t. $$C_i = \sum_{j=1}^{k} W_j \times O_{ij}$$

$$M_1 = \sum_{i \in \text{Group}_0} f(cutoff - C_i)$$

$$M_2 = \sum_{i \in \text{Group}_1} f(C_i - cutoff)$$

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

$$\sum_{j=1}^{k} W_j = 1$$

$$0 \le O_{ij}, C_i \le 1 \qquad \forall i,\ \forall j$$

$$i = 1,2,3,?,n \qquad j = 1,2,3,?,k$$

Some conditions are assumed to simplify the above problem. The one is that r_0 and r_1, which represent the different risk level of misclassification, have the same value. From this assumption, the terminology of r_0 and r_1 are removed, and the objective function is changed. The second is to fix the *cutoff* value to 0.5, and to use the error measure Y_i instead of threshold function f. The sum of error measure Y_i is same as the value of M_0 + M_1. Finally the problem is modified as following mixed integer programming.

Minimize $$MC = \sum_{i=1}^{n} Y_i$$

s.t. $$\sum_{j=1}^{k} Y_j O_{ij} + 0.5 Y_i \ge 0.5$$

Actual group "0"

$$\sum_{j=1}^{k} W_j O_{ij} - \left(\sum_{j=1}^{k} O_{ij} - 0.5\right) Y_i < 0.5$$

Actual group "1"

$$Y_i = 0 \text{ or } 1 \quad \forall i$$

$$\sum_{j=1}^{k} W_j = 1$$

Above methodology assumes that n cases are used, so n different constraints are needed to generate the value of decision variable Y_i. But for the case i, when the output values of k different methods are smaller than 0.5 without exception or larger than 0.5 in all, the weighted sum MC_i always becomes the value of the range in which the output O_{ij} is. Because this case i do not contribute to find the combining weight W_j in the mathematical programming, it is excluded from the constraints set. This exclusion is beneficial to reduce the computational complexity of the problem, as the example of next chapter shows that the number of constraints is extremely reduced. After all, this methodology is to find the linear combining weights when different classification methods suggest contradictory output for same case.

The methodology is applied to the bankruptcy prediction case in the next chapters. Next chapter demonstrates the application of integrated methodology.

3. APPLICATION

The application of this chapter is about the integration of discrimimant analysis, logit, and neural networks with mixed integer programming. We compare the performances of integrated methodology and other individual classification techniques, and apply "branch and bound" method and genetic algorithms to solve the mixed integer programming.

To simplify the original mathematical formulation suggested in chapter 2, some conditions are assumed. The one is that r_0 and r_1, which represent the different risk level of misclassification, have the same value. From this assumption, the terminology of r_0 and r_1 are removed, and the objective function is changed. The second is to fix the *cutoff* value to 0.5, and to use the error measure Y_i instead of threshold function f. The sum of error measure Y_i is same as the value of $M_0 + M_1$.

For the case i, when the output values of k different methods are smaller than 0.5 without exception or larger than 0.5 in all, the case i is excluded from the constraints set. This exclusion is beneficial to reduce the computational complexity of the problem, as the table 4 shows that the number of constraints is extremely reduced. After all, the experiment of this chapter is to find the linear combining weights when different classification methods suggest contradictory output for same case.

The sample consists of the same number of bankrupt and non-bankrupt companies in Korea. The data set contains 901 companies that filed for bankruptcy during the period 1993-1995. The same number of non-bankrupt companies is selected in the database of "Korea Credit Guarantee Fund". To select the non-bankruptcy firms, the pair matching method by industry is used. There are 196, 240, 191, 148, 64, and 62 bankrupt companies in the heavy manufacturing 1, heavy manufacturing 2, light manufacturing, construction, wholesale/retail, and other services industries. The frequency structure of non-bankrupt companies by industries is same as the bankrupt companies' structure.

Financial ratios are used as independent variables to solve the binary classification problem of bankruptcy prediction. There are 10 financial ratios; growth rate of total asset, ordinary income to total asset, owner's equity to total asset, quick ratio, total borrowings to sales, time interest earned, receivable turnover, value added to sales, cash flow to total liabilities, and net financial expense to sales. These financial ratios are selected

<Table 4> Frequency of contradictory cases

Training sample (1622)												
Case result of three methods			Experimental set									
DA	Logit	NN	1	2	3	4	5	6	7	8	9	10
Correct	Correct	Correct	1136	1156	1130	1142	1136	1168	1141	1124	1125	1091
Correct	Correct	Incorrect	66	32	54	52	61	33	51	61	64	112
Correct	Incorrect	Correct	18	16	22	20	16	18	18	17	19	17
Correct	Incorrect	Incorrect	5	7	12	10	12	8	11	18	13	13
Incorrect	Correct	Correct	31	31	28	36	37	24	31	38	41	31
Incorrect	Correct	Incorrect	25	33	28	32	25	34	24	24	19	17
Incorrect	Incorrect	Correct	72	37	69	59	74	40	49	62	74	72
Incorrect	Incorrect	Incorrect	269	310	279	271	261	297	297	278	267	269
Contradictory cases			217	156	213	209	225	157	184	220	230	262
Percentage of contradictory cases to total cases			13.4%	9.62%	13.1%	12.9%	13.9%	9.7%	11.3%	13.6%	14.2%	16.2%
Holdout sample (180)												
Case result of three methods			Experimental set									
DA	Logit	NN	1	2	3	4	5	6	7	8	9	10
Correct	Correct	Correct	124	136	125	118	124	122	129	135	124	120
Correct	Correct	Incorrect	7	2	8	7	5	2	11	3	4	5
Correct	Incorrect	Correct	4	0	3	1	3	4	1	1	4	0
Correct	Incorrect	Incorrect	1	2	1	0	4	3	1	2	0	3
Incorrect	Correct	Correct	6	5	2	3	3	4	1	3	1	4
Incorrect	Correct	Incorrect	1	1	1	4	2	3	2	5	4	4
Incorrect	Incorrect	Correct	5	3	7	12	10	1	9	5	7	12
Incorrect	Incorrect	Incorrect	32	31	33	35	29	41	26	26	36	32
Contradictory cases			24	13	22	27	27	17	25	19	20	28
Percentage of contradictory cases to total cases			13.3%	7.22%	12.2%	15.0%	15.0%	9.4%	13.9%	10.6%	11.1%	15.6%

among 67 ratios, which are popular in the financial ratio analysis.

The whole sample is divided into 10 subsets to verify the generalized performance of methodology. Samples consist of two parts: 1622 training samples and 180 holdout samples. The experiments are repeated 10 times, so the results of 1800 holdout samples are achieved. All the results that are shown in the next section are average value of ten different experiments.

Two statistical methods of discriminant analysis, logistic regression, and one artificial intelligence method of neural network are used as three different classification methods. Three methods generate output O_{ij} for case i, and each O_{ij} makes constraints. The combined output C_i is the weight average of outputs which three different methods suggest.

To solve the mathematical programming, the software of "Cplex Mixed Integer Programming" UNIX version is used. The program is solved by the "branch and bound methods", and the maximum value of branch and bound nodes is fixed to 20000. Table 5 shows the small part of programming module for our example. Also, this study compares the performance of "branch and bound" method with that of genetic algorithms.

The programming module is formulized with cases for which k different methods generate the contradictory outputs. Because our example is to combine three different methods of discriminant analysis, logistic regression, and neural networks, when the three outputs for one case are not consistent, that case is included in the mathematical programming module.

As table 4 shows the frequency of contradictory cases for each experimental sample, the ratio of contradictory cases in the total cases is about 10 ~ 15%. Because the number of cases to be used for solving the programming decreases a lot, the complexity of problem, which is highly related to the number of constraints, extremely decreased. In table 5, the maximum index of dummy variable Y_i has the same value with the number of cases in the problem. The remainder of this chapter analyzes the results and performances

<Table 5> Example of mathematical programming module

```
Minimize Y1 + Y2 + Y3 + ... + Y216 + Y217

Subject to
0.454 W1 + 0.504 W2 + 0.492 W3 - 0.9504 Y1 < 0.5
0.661 W1 + 0.582 W2 + 0.347 W3 - 1.0896 Y2 < 0.5
0.343 W1 + 0.449 W2 + 0.532 W3 - 0.8239 Y3 < 0.5
...
0.305 W1 + 0.218 W2 + 0.570 W3 - 0.5931 Y216 < 0.5
0.449 W1 + 0.432 W2 + 0.563 W3 + 0.5  Y217 >= 0.5

W1 + W2 + W3 = 1

Bounds
0 <= W1 <= 1   0 <= W2 <= 1      0 <= W3 <= 1
0 <= Y1 <= 1   0 <= Y2 <= 1      0 <= Y3 <= 1
...
0 <= Y216 <= 1 0 <= Y217 <= 1

Integers
Y1   Y2       Y3
...
Y216  Y217

End
```

of integrated methodology.

Table 6 and 7 show the performances of discriminant analysis, logit, neural networks, and combined methodology. We also use two simple combining methods. The one is to average the outputs of discriminant analysis, logit, and neural networks. The other combining methods is to select the output that has the maximum difference from cutoff value in the outputs of discriminant analysis, logit, and neural networks. The comparison results with these two combining methods are included in table 6 and 7. It is obvious that the combined methodology is superior to the other individual techniques in the training sample. Because the objective function of methodology is to minimize the number of misclassification cases, the mixed integer programming searches the weight to reduce the misclassification error. The hit ratio, which is used as a performance measure in this experiment, is directly computed from the misclassification

<Table 6> Performance comparison (10 folds, whole cases, %)

Training sample							
	DA	Logit	NN	Combining (Max)	Combining (Average)	Genetic algorithms	Combined result
Fold 1	75.51	77.54	77.48	77.17	76.93	78.29	78.35
Fold 2	74.65	77.17	76.43	76.19	76.19	77.24	77.30
Fold 3	75.08	76.43	76.99	76.13	76.62	77.24	77.30
Fold 4	75.45	77.79	77.48	77.54	77.54	78.96	79.03
Fold 5	75.51	77.61	77.85	77.17	77.54	78.29	78.35
Fold 6	75.63	77.61	77.05	76.68	76.74	77.67	77.73
Fold 7	75.26	76.87	76.37	76.06	76.13	77.05	77.11
Fold 8	75.20	76.87	76.50	76.31	76.50	77.17	77.17
Fold 9	75.26	76.99	77.61	77.05	77.24	78.04	78.10
Fold 10	76.00	77.11	74.65	76.80	77.11	77.42	77.48
Average	75.35	77.20	76.84	76.71	76.85	77.74	77.79
Holdout sample							
	DA	Logit	NN	Combining (Max)	Combining (Average)	Genetic algorithms	Combined result
Fold 1	75.56	76.67	77.22	77.22	78.89	77.78	77.78
Fold 2	77.78	80.00	80.00	81.11	81.11	80	80.00
Fold 3	76.11	75.56	76.11	75.56	76.11	76.11	76.11
Fold 4	70.00	73.33	74.44	72.78	72.78	74.44	75.00
Fold 5	75.56	74.44	77.78	76.67	76.67	77.78	78.89
Fold 6	72.78	72.78	72.78	73.89	73.89	73.89	72.78
Fold 7	78.89	79.44	77.78	78.33	78.89	79.44	79.44
Fold 8	78.33	81.11	80.00	80.56	80.56	80	80.56
Fold 9	73.33	73.89	75.56	75.00	75.56	76.11	76.11
Fold 10	71.11	73.89	75.56	73.89	72.78	73.89	75.56
Average	74.94	76.11	76.72	76.50	76.72	76.94	77.22

<Table 7> Performance comparison (10 folds, contradictory cases, %)

Training sample (1622)								
	Number of cases used	DA	Logit	NN	Combining (Max)	Combining (Average)	Genetic algorithms	Combined result
Fold 1	217	41.01	56.22	55.76	53.46	51.61	61.79	62.21
Fold 2	156	35.26	61.54	53.85	51.28	51.28	62.22	62.82
Fold 3	213	41.31	51.64	55.87	49.30	53.05	57.77	58.22
Fold 4	209	39.23	57.42	55.02	55.50	55.50	66.48	66.99
Fold 5	225	39.56	54.67	56.44	51.56	54.22	59.59	60.00
Fold 6	157	37.58	57.96	52.23	48.41	49.04	58.62	59.24
Fold 7	184	43.48	57.61	53.26	50.54	51.09	59.23	59.78
Fold 8	220	43.64	55.91	53.18	51.82	53.18	58.15	58.18
Fold 9	230	41.74	53.91	58.26	54.35	55.65	61.32	61.74
Fold 10	262	54.20	61.07	45.80	59.16	61.07	62.97	63.36
Average		41.70	56.79	53.97	52.54	53.57	60.81	61.25
Holdout sample (180)								
		DA	Logit	NN	Combining (Max)	Combining (Average)	Genetic algorithms	Combined result
Fold 1	24	50.00	58.33	62.50	62.50	75.00	66.68	66.67
Fold 2	13	30.77	61.54	61.54	76.92	76.92	61.54	61.54
Fold 3	22	54.55	50.00	54.55	50.00	54.55	54.54	54.55
Fold 4	27	29.63	51.85	59.26	48.15	48.15	59.23	62.96
Fold 5	27	44.44	37.04	59.26	51.85	51.85	59.27	66.67
Fold 6	17	52.94	52.94	52.94	64.71	64.71	64.72	52.94
Fold 7	25	52.00	56.00	44.00	48.00	52.00	55.97	56.00
Fold 8	19	31.58	57.89	47.37	52.63	52.63	47.37	52.63
Fold 9	20	40.00	45.00	60.00	55.00	60.00	64.99	65.00
Fold 10	28	28.57	46.43	57.14	46.43	39.29	46.44	57.14
Average		41.45	51.70	55.86	55.62	57.51	58.07	59.61

error when the problem does not consider different cost level about misclassification results.

Among the 7 alternative methods, the suggested methodology has the highest level of accuracy (77.22%) in the holdout sample, followed by the genetic algorithms (76.945), averaging method and neural networks (76.72%), maximum difference method (76.72%), and logit (76.11%).

We use the paired t-tests to examine whether the predictive performance of integrated methodology is significantly higher than that of other techniques. As shown in table 8 and 9, the suggested combining methodology performs significantly better than the individual methods of discriminant analysis, logit, and neural network. Also, the performance of suggested methodology is higher than that of maximum difference method at the 5 % significance level.

Table 7 shows the results of six methods about the contradictory cases. The accuracy of linearly combining model is still highest than that of any other techniques. Because contradictory cases are not stable but volatile basically, overall hit ratios are lower than the performances when

<Table 8> Paired t-test with hit ratios (10 folds, whole cases, holdout sample)

	Logit	NN	Max	Average	Genetic algorithm	Combined
DA	-2.400 **(0.040)	-3.105 **(0.013)	-3.698 ***(0.005)	-4.598 ***(0.001)	-5.013 ***(0.001)	-4.172 ***(0.002)
Logit		-1.337 (0.214)	-1.210 (0.257)	-1.557 (0.154)	-2.085 *(0.067)	-2.372 **(0.042)
NN			0.667 (0.522)	0.000 (1.000)	-0.801 (0.444)	-2.862 **(0.019)
Max				-1.000 (0.343)	-1.633 (0.137)	-1.901 *(0.090)
Average					-0.738 (0.479)	-1.077 (0.310)
Genetic						-1.170 (0.272)

* significant at 0.1 level
** significant at 0.05 level
*** significant at 0.01 level

<Table 9> Paired t-test with hit ratios (10 folds, contradictory cases, holdout sample)

	Logit	NN	Max	Average	Genetic algorithm	Combined
DA	-2.442 **(0.037)	-3.361 ***(0.008)	-3.136 **(0.012)	-3.729 ***(0.005)	-5.245 ***(0.001)	-4.554 ***(0.001)
Logit		-1.240 (0.246)	-1.442 (0.183)	-1.847 *(0.098)	-2.035 *(0.072)	-2.362 **(0.043)
NN			0.082 (0.937)	-0.475 (0.646)	-1.066 (0.314)	-2.992 **(0.015)
Max				-1.174 (0.270)	-0.964 (0.360)	-1.215 (0.256)
Average					-0.219 (0.832)	-0.575 (0.579)
Genetic						-0.807 (0.441)

* significant at 0.1 level
** significant at 0.05 level
*** significant at 0.01 level

whole cases are used. However, the hit ratio of optimal combining model for holdout sample is 59.61%, and differences between the performances of our method and the other benchmarks increase. The results of statistical significance test represented in table 9, explained that optimal combining model is superior to three individual models.

4. CONCLUSION

This paper studies the integration of discrimimant analysis, logit, and neural networks with mixed integer programming. It compares the performances of integrated methodology and other individual classification techniques, and applies "branch and bound" method and GA to solve the mixed integer programming. The results of this application show that the suggested methodology outperforms any of techniques applied individually.

The main contribution of this paper is the development of a new integration methodology using the linear combining function of different techniques' outputs. The main idea of suggested methodology is to find the optimal weight structure to be able to minimize the misclassification errors of a given data set. Threshold function of f and binary error measure Y_i are suggested in order to make the linkage between optimal weight structure and minimum misclassification cost.

Traditional statistical methods of discriminant analysis and logistic regression, and artificial intelligence methods of neural networks are applied to the domain of bankruptcy prediction. The integrated methodology hires the joint solution schema among the various integration schemas in order to generate a final single solution with the outputs of individual techniques.

Branch and bound methods are applied to solve the mixed integer programming model, and genetic algorithms are also applied to find the optimal weights structure. Two methods play the same role in the problem solving process, and it is confirmed that two techniques are both useful.

The limitation of this research is not to expand the integrated model into the n group classification problem. Binary classification problem may be solved easily with the threshold function, but more consideration about various parameters has to be followed in order to solve the multiple group problems. Although we assumed that the risk of misclassification is always same for binary classification problem, the computational process of misclassification errors has to be changed for multiple group problems.

REFERENCE

1. Altman, E., "Financial ratios, discriminant analysis, the prediction of corporate bankruptcy", *Journal of Finance,* Vol. 23, No. 3 (1968), 589-609.
2. Altman, E., G. Marco, & F. Varetto, "Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)", *Journal of Banking and Finance*, Vol. 18 (1994), 505-529.

3. Beaver, W, "Financial ratios as predictors of failure. Empirical research in accounting: Selected studies", *Journal of Accounting Research*, Vol. 5 (Supp.) (1967), 71-111.

4. Braun, H., & J. Chandler, "Predicting stock market behavior through rule-induction: An application of the learning-from-example approach", *Decision Science*, Vol. 18 (1987), 415-429.

5. Cadden, D, "Neural networks and the mathematics of chaos - An investigation of these methodologies as accurate predictors of corporate bankruptcy", *Proceedings of the 1st International Conference on Artificial Intelligence Applications on Wall Street*, IEEE Press. (1991).

6. Fanning, K. M., & K. O. Cogger, "A comparative analysis of artificial neural networks using financial distress prediction , *Intelligent Systems in Accounting, Finance, & Management*, Vol. 3 (1994), 241-252.

7. Garrison, L., & R. Michaelson, "Symbolic concept acquisition: A new approach to determining underlying tax law construct", *Journal of American Taxation Association*, (1989), 77-91.

8. Jhee, W. C., & J. K. Lee, "Performance of neural networks in managerial forecasting", *Intelligent Systems in Accounting, Finance and Management*, Vol. 2 (1993), 55-71.

9. Jo, H. & I. Han, "Integration of case-based forecasting, neural network, and discriminant analysis for bankruptcy prediction", *Expert Systems With Applications*, Vol. 11, No. 4 (1996), 415-422.

10. Jo, H., I. Han, & H. Lee, "Bankruptcy prediction using case-base reasoning, neural networks, and discriminant analysis", *Expert Systems With Applications*, Vol. 13, No. 2 (1997), 97-108.

11. Kumar, A., & I. Ormeda, *Hybrid classifiers for knowledge discovery*, Working Paper, Colorado, 1997.

12. Lee, S. B., & S. H. Oh, "A comparative study of recursive partitioning algorithm and analog concept learning system", *Expert Systems With Applications*, Vol. 1 (1990), 403-416.

13. Liang, T. P., J. S. Chandler, & I. Han, "Integrating statistical and inductive learning methods for knowledge acquisition", *Expert Systems With Applications*, Vol. 1 (1990), 391-401.

14. McKee, T. E. & T. Lensberg, "Genetic programming and rough sets: A hybrid approach to bankruptcy classification", *European Journal of Operational Research*, Vol. 138 (2002), 436-451.

15. Markham, I. S., & C. T. Ragsdale, "Combining neural networks and statistical predictions to solve the classification problem in discriminant analysis", *Decision Sciences*, Vol. 26 (1995), 229-242.

16. Odam, M., & R. Sharda, "A neural network model for bankruptcy prediction", In R. Trippi, & E. Turban (eds.), *Neural Networks in Finance and Investing* (pp. 177-185), Chicago: Probus Publishing Company, 1993.

17. Park, C. & I. Han, "A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction", *Expert Systems With Applications*, Vol. 23, No. 3 (2002), 255-264.

18. Raghupathi, W., L. Schkade, & B. Raju, "A neural network approach to bankruptcy prediction", In R. Trippi, & E. Turban (eds.), *Neural Networks in Finance and Investing* (pp. 141-158). Chicago: Probus Publishing Company, 1993.

19. Shin, K. & Y. Lee, "A genetic algorithm application in bankruptcy prediction modeling", *Expert Systems With Applications*, Vol. 23, No. 3 (2002), 321-428.

20. Surkan, A., & J. Singleton, "Neural networks for bond rating improved by multiple hidden layers", *Proceedings of the IEEE Conference on Neural Networks*. (II pp. 157-162), San Diego, CA., 1990.

21. Tam, K. & M. Kiang, "Managerial applications of neural networks: The case of bank failure predictions", *Management Science*, Vol. 38 (1992), 926-947.

22. Trigueiros, D., "Incorporating complementary ratios in the analysis of financial statements", *Accounting, Management, & Information Technology*, Vol. 4, No. 3 (1994), 149-162.

23. Troutt, M. D., A. Rai, & S. K. Tadisina, "Aggregating multiple expert data for linear case valuation models using the MDE principle", *Decision Support Systems*, Vol. 21 (1997), 75-82.

요약

다수의 분류 기법의 예측 결과를 결합하기 위한 혼합 정수 계획법의 사용

조홍규*
한인구**

경영 분류 문제에 대한 많은 연구들은 여러가지 기법들간의 성과 비교에 대한 것이었지만, 각각의 연구들마다 가장 좋은 기법이 어떤 것인가에 대해서는 상이한 결론을 내고 있다. 다수의 분류 기법 중에서 가장 좋은 것을 사용하는 방법에 대한 대안으로, 분류 기법을 통합하여 성과를 향상시키는 방법이 있다.

본 연구에서는 개별 분류 기법의 결과를 선형 결합하여 예측력을 높이는 방법을 제시하였다. 최적 선형 결합 가중치를 계산하기 위해 혼합 정수 계획법을 사용하였다. 목적함수로 사용한 오분류 비용의 최소화에서 오분류 비용은 부도 기업을 모형에서 정상으로 예측한 오류와 정상 기업을 모형에서 부도 기업으로 예측한 오류의 합으로 정의하였다. 문제 풀이 과정을 단순화하기 위하여 본 논문에서는 절사점 (cutoff value)을 고정하였고, 경계 함수 (threshold function)를 배제하였다. 정수계획법의 계산을 위해 branch & bound 방법을 사용하였다.

선형 결합에 의한 모형의 예측력이 개별 기법에 의해 구축된 모형의 예측력을 상회하였고, 그 차이가 통계적으로도 유의하였다.

주제어: 도산 예측, 인공 지능, 결합 예측, 선형 계획 모형

* 나이스채권평가 금융공학연구소
** 한국과학기술원 테크노경영대학원