International Journal of Management Science
Vol 3, No 1, May 1997

THE PERFORMANCE OF THE BINARY TREE CLASSIFIER AND DATA CHARACTERISTICS

JEONGSUN PARK
Department of Industrial Engineering Myongji University

ABSTRACT

This paper applies the binary tree classifier and discriminant analysis methods to predicting failures of banks and insurance companies. In this study, discriminant analysis is generally better than the binary tree classifier in the classification of bank defaults; the binary tree is generally better than discriminant analysis in the classification of insurance company defaults. This situation can be explained that the performance of a classifier depends on the characteristics of the data. If the data are dispersed appropriately for the classifier, the classifier will show a good performance. Otherwise, it may show a poor performance. The two data sets (bank and insurance) are analyzed to explain the better performance of the binary tree in insurance and the worse performance in bank; the better performance of discriminant analysis in bank and the worse performance in insurance.

1. INTRODUCTION

For the past few years, the number of banks/insurance companies that entered bankruptcy[1)] has increased dramatically. These failures have many serious side effects on the economy as a whole. The federal and state governments have mandated a number of regulatory agencies to monitor the operations of commercial banks/insurance companies. These agencies supervise their member banks/insurance companies through the quarterly/annual reports of banks, on-site examinations, and off-line surveillance systems. Though the most direct method for supervision is to send auditors to each bank on a regular basis, such a practice is not feasible given the tremendous resources involved and the large number of banks under supervision.

In order to identify banks/insurance companies at risk of failure, there have been numerous studies on bank/insurance company bankruptcy predictions.

1) Bankruptcy is made in US.

Most of the studies advocate the use of statistical methods called discriminant analysis. Typically, discriminant analysis methods assume that all variables are normally distributed. In the case of linear classifier, it also requires identical covariance matrices. In reality, however, these assumptions are rarely satisfied.

In view of these limitations, a binary tree classifier [4, 9] has been introduced as an alternative method for bank bankruptcy prediction. The binary tree classifier requires no distributional assumptions and no functional forms. Thus, it is argued as a more robust method than general statistical methods. In addition, the binary tree classifier takes into account economic concepts of prior probabilities and misclassification costs like discriminant analysis. To use the class proportions in the sample as the prior probabilities may generate biased results. The cost of classifying as viable a bank/insurance company which subsequently fails should be much higher than the cost of classifying as failing a bank/insurance company which is actually viable. Thus, the concepts of priors and misclassification costs are important in the prediction of bank/insurance company bankruptcy. Recently, neural network methods have been used and show good performances in prediction [5, 20].

However, this paper explores the performances of classifiers with the concepts of prior probability and misclassification cost. Thus, we compare the binary tree classifier[2)] with discriminant analysis for bank/insurance company failure prediction and analyze the results of them in various experiments. The rest of this paper is organized as follows: In section 2 and 3, the binary tree classifier and discriminant analysis are reviewed. In section 4, the bank and insurance company failure problems are described. The bank/insurance company data sample, and the results of the bank/insurance company classification are presented and discussed in section 5. Section 6 describes data characteristics of bank and insurance data. Section7 concludes the paper with discussion on the limitations of current approach and outlines future research directions.

2. A BINARY TREE CLASSIFIER

The binary tree classifier has a similar learning mechanism as other inductive learning methods [13, 17, 18]. To build a binary tree, the model (Figure 1) requires a set of training examples and the binary tree method. Each example in the training set has decision variables & a decision. The built binary classifier involves if-then rules and an inference engine. The classifier is then used to predict a decision for a new example which requires a decision.

---

2) The binary tree classifier is the classification and regression tree (CART).

2.1 Concepts of the Binary Tree Classifier

The binary tree classifier procedure requires four elements to create a tree: a set of questions involving a predicate of the form {Is x < A?} where x is a variable and A is a given threshold value, a goodness of split criterion that can be evaluated for any split condition at a node, stopping rules to terminate splitting, and a rule to assign a class to a terminal node.

The procedure employs a recursive splitting strategy that repeatly splits a node into 2 disjoint nodes (children). Observations contained in the original node are divided into 2 groups with each stored in one child. When a node decides to split, a set of questions is made by using the threshold values of each variable at the node. Each question is evaluated by the split criterion which yields an evaluation value. If all values obtained from evaluating the questions are less than some given minimum value for splitting, then the node stops splitting. Otherwise, the question form that yields the maximal evaluation value is selected as the splitting predicate and two new children nodes are created. Also, if all observations at a node belong to the same class, the node stops splitting. Such a node becomes a terminal or leaf node of the tree. The class assignment rule is then applied to determine the class tag of the node.

If the observations are exact and free of error, the largest tree which contains the most information will have the best classificatory accuracy. In reality, observations are recorded with error, important variables may be missed, and the chosen variables may be inappropriate for prediction. It is therefore necessary to prune a tree so constructed to make it more robust as reported in [15] and to improve the generality of the classifier. The minimal cost-complexity pruning method is directed towards this end. The minimal cost-

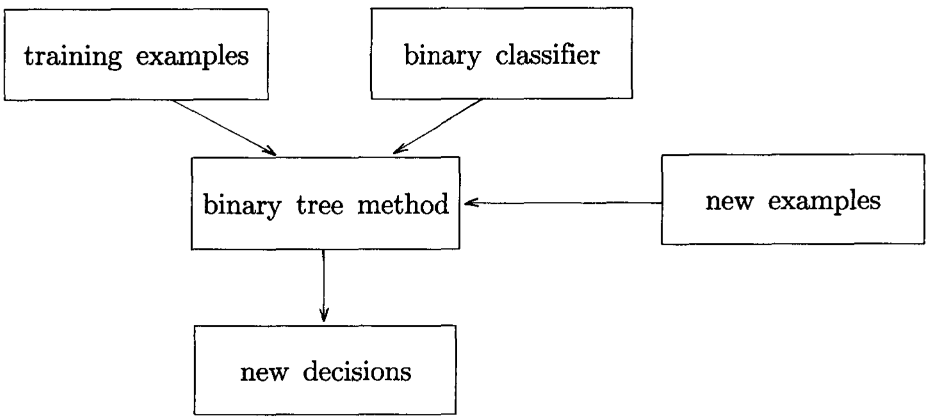

Figure 1. A Classification Model Using the Binary Tree.

complexity pruning method considers both misclassification costs and penalizes for additional complexity of the tree. In selecting a pruning node, the method finds a node with the weakest links to its branches. The pruning continues to the point where only one node remains. The pruning process creates a sequence of trees with decreasing order of tree size.

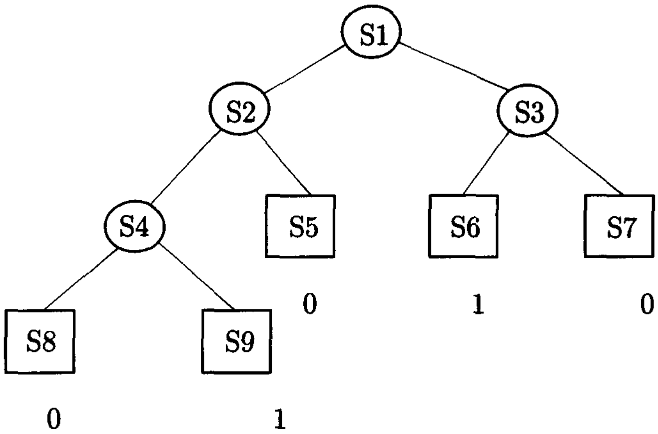

Figure 2. The Binary Tree Classifier

2.2 A Classification in the Binary Tree Classifier

The tree induction technique, or more correctly, binary tree classifier, is constructed by splitting of a node into two descendant children nodes. In Figure 2, S2 and S3 are disjoint with S1 = S2∪S3. Similarly S4 and S5 are disjoint with S4∪S5 = S2. The nodes which are not split, S5, S6, S7, S8, S9, are called terminal nodes. In Figure 2, terminal nodes are represented by rectangular boxes, non-terminal nodes by circles. Each terminal node is assigned a class label. There may be many terminal nodes with the same class label. In this example, terminal nodes, S5, S7, and S8 have the same label of class 0; terminals S6 and S9 have the same label of class 1. In our current study, 0 will denote solvent, and 1 will denote bankruptcy.

The splits are made by conditions on the variables (x1, x2, ...). For example, the split of S1 into S2 and S3 could be made by defining S2 = $\{S1;\ x4<7\}$, and S3 = $\{S1;\ x4\geq 7\}$. The split of S3 into S6 and S7 could be made by defining S6 = $\{S3;\ x5<-2\}$, and S7 = $\{S3;\ x5\geq -2\}$.

The tree classifier predicts a class by using a vector x = (x1, x2, x3 ...... xn) pertaining to an object . From the top of the tree, the classifier determines which way to take. For example, if x4 is less than 7, it goes into S2, otherwise into S3. When x arrives at a terminal subset, the classification is made using the terminal label.

3. DISCRIMINANT ANALYSIS

The discriminant analysis (DA) has a similar classifying mechanism as other statistical methods such as linear regression and logistic regression. To build a discriminant analysis equation, a set of training examples and discriminant analysis method are required (Figure 3). Each example in the training set has decision variables & a decision. The derived discriminant analysis equation is then used to predict a decision if there a new example which requires a decision.

Discriminant analysis method is a Bayesian method using classification rules to minimize the expected misclassification cost. Classify an observation with a vector of attribute variables x = (x1, x2, x3,...,xn) to class 1 if

$$\frac{f1(x)}{f0(x)} > \frac{\pi 0 C(1 \mid 0)}{\pi 1 C(0 \mid 1)} \text{ otherwise assign it to class 0.}$$

Here *f1* and *f0* represent the multivariate probability density functions of variables for classes 1 and 0. C(0 | 1) and C(1 | 0) are misclassification costs of predicting a class 1 observation as class 0 and vice versa. $\pi 1$ and $\pi 0$ are prior probabilities of class 1 and class 0 respectively.

DA makes the Bayesian rule operational by assuming that probability densities f1 and f0 are multivariate normal. If we also assume that covariance

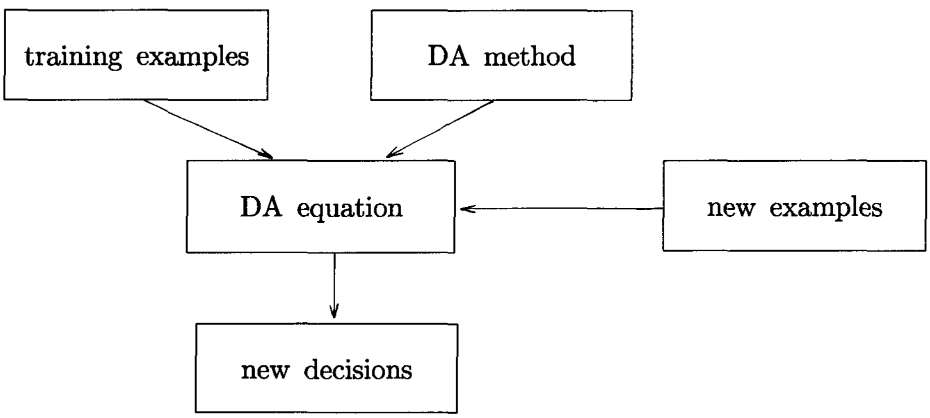

Figure 3. A Classification Model Using Discriminant Analysis.

matrices of the two densities are equal, Bayes rule reduces to the linear discriminant rule. Assign an observation x to class 1 if

$$x'u - v > \ln \frac{\pi 0 C(1 \mid 0)}{\pi 1 C(0 \mid 1)}$$

otherwise assign it to class 0, where u is a multiplication of an inverse covarance matrix and (u1 - u0), and $v = \frac{(u1 + u0)'u}{2}$ ($u1$, u0: class means.)

4. BANK/INSURANCE COMPANY FAILURE PROBLEM

4.1 Bank Failure Problem

Regulatory agencies rely on regularly filed reports from banks to assess their financial conditions. As a means to monitor member banks, on-site examinations are often conducted. The purposes of on-site examinations are determining the quality of assets, checking for compliance with laws and regulations, evaluating controls and management, and appraising capital adequacy.

However, these examinations are expensive and hard to schedule because a bank's operations may interrupt. If sound banks are examined frequently, resources will be wasted. If high risk banks are not examined as they should be, the possibility of bankruptcy will be increased. Thus, a regulatory agency needs some criteria to decide to send examination teams to individual banks or not.

To develop such criteria, many studies have been done [1, 8, 11, 14, 19, 20, 23]. Most studies employ statistical methods such as discriminant analysis. Yet they have limitations. In the case of discriminant analysis, it requires that all variables are normally distributed. In reality, however, this requirement is seldom met. In addition, most distribution-free and nonparametric methods do not take into account prior probabilities or misclassification costs. Most methods use the class proportions in the sample as the prior probabilities, which may generate biased results. The cost of classifying as viable a bank which subsequently fails should be much higher than the cost of classifying as failing a bank which is actually viable.

As an alternative to statistical methods, which does not require normal distributions of all variables, a binary tree classifier has been used to predict the bank bankruptcy prediction [9]. It uses prior probabilities as well as misclassification costs.

4.2 Insurance Company Failure Problem

Insolvency of insurance companies is important to policyholders as well as management. This study focuses on bankruptcy in the property/casualty industry because this industry has shown an increase in bankruptcies recently. The property/casualty insurance is regulated by state laws. Surveillance of solvency is a main responsibility of regulatory agencies, which use financial indices of the Insurance Regulatory Information System (IRIS) to predict the insolvency of insurance companies. IRIS suggests regulation of a company if four or more indices out of eleven show performance beyond the acceptable ranges. However, the IRIS method has been criticized because it does not consider the interdependence among the indices, the difference among companies which have different number of violent indices; it also does not show the statement date and the date the indices are available to regulators, accurate ranges for some of indices, a consideration of the lines of business, or an early warning of financial distress [12].

State regulators, insurance companies, and schools share the importance of solvency for the healthy growth of the insurance industry. To solve the insolvency problem, some studies have been done. Denenberg finds that the Best rates of failing companies show a downward trend prior to failure [7]. Trieschmann and Pinches use discriminant analysis and financial indices to discriminate viable companies from failing companies [21]. BarNiv and Smith compare the prediction ability between the logistic regression model and univariate model from the view of the expected cost of misclassification, and conclude that the logistic regression model has a better potential to predict failures [3].

5. EXPERIMENTS

5.1 Data Sample

a. Bank Data Sample

The data used in these experiments are taken from the FDIC annual reports from 1985 to 1987. The sample consists of failed banks in Texas during the period from January 1985 through December 1987 for which all necessary information is available. These banks constitute the failure class. A non-failure sample (sound banks) is needed. The banks in both samples should be similar in other aspects as a control measure; thus a matching procedure is necessary. The matching is conducted using the following criteria [20]: asset, charter status (state or federal), branch size, age of bank.

We consider two prediction periods: one year before failure and two years before failure. Information for each period is collected for each bank in the sample. One study suggests that prediction more than two years ahead is invalid because most factors which cause a bank to fail will not appear until two years preceding bankruptcy [1]. The data for bank failure prediction in one year are composed of 81 failed banks and 81 non-failed banks. The data sample for two-year period consists of 79 failed and 79 non-failed banks.

b. Insurance Company Data Sample

Any insurance company is required to file its financial statements annually. It must use the regular form made by the National Association of Insurance Commissioners (NAIC). The primary purpose of the annual statements is to make state regulators evaluate the solvency status of insurance companies.

In this study, a paired-sample method is used to compare bankrupt insurance companies with non-bankrupt insurance companies. All bankrupt insurance companies are included as a half of the sample and the same number of non-bankrupt insurance companies as the other half of the sample.

The number of bankrupt companies during the period of 1984 and 1990 is 154. However, the A. M. Best tapes[3)] contain information on only 142 companies [10]. We hope to see the effects of financial status on failures prior to one or two years. A study suggests that prediction more than two years ahead is invalid because most determinants which cause a corporate bankruptcy will not appear until two years preceding bankruptcy [1]. The tapes did not contain all one-year and two-year data of the bankrupt companies. The tapes missed one-year or two-year data for some bankrupt companies. Those companies were not considered to have consistency of the data. The collected bankrupt data involve only the 102 companies which have both one-year and two-year data. Additionally, some bankrupt companies had zero value for the variable of surplus which is used in the denominator of some financial indices. For better consistency of the data, those companies with zero surplus were eliminated from the data. Finally, the size of the data reached 92.

---

3) We appreciate Dr. Robert C. Witt for providing the A. M. Best tapes.

5.2 Used Variables

a. Bank Data Variables
There has been some controversy in the selection of financial ratios for default prediction. Some ratios are so highly correlated that the results overemphasize the importance of those ratios. On the other hand, some ratios may have no prediction power. This study adopts the same set of variables used in previous study [20].

b. Insurance Company Data Variables
The nine variables in Table 1. are calculated for each company in the data sets. Four indices (NPS, LR, LLA, and LEL) were selected first because they are recognized as good indicators which discriminate bankrupt companies from non-bankrupt companies [2]. The other variables are from [10], who investigated nineteen financial indices to verify the validity of each variable as a factor which discriminates bankrupt insurance companies from non-bankrupt insurance companies using the whole data of A. M. Best tapes. He concluded that eight variables (NPS, DPS, LRS, LR, NIS, COS, LLA, and RAS) show differences between bankrupt and non-bankrupt companies. Three variables (NPS, LR, LLA) are duplicated. Therefore, nine variables are used in this study where two variables (NPS and LLA) are IRIS's.

Table 1. A List of Variables

Name	Description
NPS(IRIS1)	Net Premiums Written / Surplus
DPS	Direct Premiums Written / Surplus
LRS	Loss Reserves / Surplus
LR	(Losses + Loss Adjustment Expenses) / Premiums
NIS	Net Income / Surplus
COS	Cash From Operations / Surplus (previous year)
LLA(IRIS7)	Liabilities/Liquid Assets(Bonds+Stocks+Cash+Interest)
RAS	Reinsurance Assumed / Surplus
LEL	Loss Adjustment Expenses / Losses Incurred

5.3 Experimental Design

We have constructed two prediction models, one-year ahead prediction (one-year period) and two-year ahead prediction (two-year period). The method for comparing the trees with DAs is the independent test sample method. The independent test sample method uses a separate test sample which is different from the training sample. This method is useful when the available data sample is large. The parameters chosen for our experiments are: There are two prior probabilities of failure, 0.01 and 0.02. For each probability, type I misclassification cost takes on one value from [1, 5, 15, 25, 50, 60, 75, and 100], while type II cost remains at 1. These sixteen tests are carried out using the one-year data, and are repeated with the two-year data. In each experiment that uses the one-year bank (insurance company) data, 59 (56) failed and 59 (56) non-failed banks (insurance companies) are used for training, with 22 (36) failed and 22 (36) non-failed for testing. In the case of the two-year data, 59 (56) failed and 59 (56) non-failed banks (insurance companies) are used for training, and 20 (36) failed and 20 (36) non-failed banks (insurance companies) are used for testing.

The expected misclassification cost is used as the performance measure. Type I and type II error rates are calculated first and the rates are weighted by their priors and costs. Namely, expected misclassification cost = type I error rate * prior probability of class 1 * type I error cost + type II error rate * prior probability of class 0 * type II error cost.

5.4 Empirical Results

To validate trees using the test sample method, 32 experimentations are performed using two bank (insurance company) failure prior probabilities, 0.01 and 0.02, type I error costs (i.e. 1, 5 ,15, 25, 50, 60, 75, and 100), in two prediction periods, the one-year period and two-year period. The experimental results[4)] are summarized in Tables 2 to 5. The same test procedures are used for discriminant analysis validations.

In Tables 2 through 5, as type I cost increases, the type I error rate using either classification method generally decreases. The type I error rate using the binary tree method decreases more drastically than that of discriminant analysis method. In contrast, as type I error cost increases, the type II error rate using the binary tree and discriminant analysis method generally increases. The type II error rate using the binary tree method increases more speedily than when using discriminant analysis method. The total cost (the sum of two expected misclassification costs) for both the binary tree and discriminant analysis methods increases, as type I cost increases. As shown, the total cost using the binary tree method is generally higher than that by discriminant analysis

4) The binary tree classifier program is written in Pascal and runs on an Encore Multiprocessor machine. The discriminant analysis procedure is run by SAS on an IBM machine.

method.

As known, the total cost using the binary tree method is generally higher than that by discriminant analysis method in bank data (28 cases out of 32) and is generally lower in insurance data (23 cases out of 32). The results show that the performance of a model may change depending on the characteristics of data. These results will be analyzes by the characteristics of data sets.

6. TASK CHARACTERISTICS

6.1 Task Characteristics of Insurance

When we consider task characteristics, we assume that sample data reflect the characteristics of the task. Insurance is one of the industries5) controlled strictly by federal or state agencies [10]. Because this regulation generates better data, financial data for these industries (insurance, utilities, and banking) indicate their financial status more vividly than do the data for other industries.

Table 2. Test1 (Test Sample; 0.01 prior; one-year)

type I cost type II cost	1 1	5 1	15 1	25 1	50 1	60 1	75 1	100 1
type I rate bt (banking) da	1.000 0.409	1.000 0.273	1.000 0.136	1.000 0.091	0.136 0.091	0.136 0.091	0.227 0.091	0.091 0.046
type II rate bt (banking) da	0.000 0.046	0.000 0.091	0.000 0.091	0.000 0.136	0.273 0.182	0.273 0.227	0.273 0.227	0.318 0.227
total cost bt (banking) da	*0.010 0.050	*0.050 0.104	0.150 0.110	0.250 0.157	0.338 0.226	0.352 0.279	0.440 0.293	0.406 0.271
type I rate bt (insurance) da	1.000 0.972	1.000 0.944	1.000 0.889	0.472 0.833	0.361 0.750	0.389 0.639	0.250 0.611	0.194 0.500
type II rate bt (insurance) da	0.000 0.056	0.000 0.056	0.000 0.083	0.083 0.083	0.222 0.083	0.222 0.083	0.333 0.083	0.361 0.194
total cost bt (insurance) da	*0.010 0.065	*0.050 0.102	*0.150 0.216	*0.201 0.291	*0.401 0.458	*0.453 0.466	*0.518 0.541	*0.552 0.693

*: Total cost of the binary tree is less than that of DA

5) Insurance, utilities, and banking are required to report their financial status in more detail than other industries.

Table 3. Test2 (Test Sample; 0.02 prior; one-year)

type I cost type II cost	1 1	5 1	15 1	25 1	50 1	60 1	75 1	100 1
type I rate bt (banking) da	1.000 0.273	1.000 0.227	1.000 0.091	0.136 0.091	0.091 0.046	0.091 0.046	0.136 0.046	0.000 0.046
type II rate bt (banking) da	0.000 0.091	0.000 0.091	0.000 0.136	0.273 0.182	0.318 0.227	0.318 0.227	0.364 0.227	0.545 0.227
total cost bt (banking) da	*0.020 0.095	*0.100 0.112	0.300 0.161	0.335 0.224	0.403 0.268	0.421 0.278	0.561 0.291	0.535 0.314
type I rate bt (insurance) da	1.000 0.944	1.000 0.917	0.472 0.833	0.389 0.750	0.194 0.500	0.194 0.472	0.194 0.389	0.139 0.167
type II rate bt (insurance) da	0.000 0.055	0.000 0.083	0.111 0.083	0.222 0.083	0.361 0.194	0.361 0.250	0.361 0.278	0.389 0.750
total cost bt (insurance) da	*0.020 0.073	*0.100 0.173	*0.251 0.332	*0.412 0.457	*0.548 0.691	*0.587 0.812	*0.646 1.019	*0.659 1.068

*: Total cost of the binary tree is less than that of DA

Table 4. Test3 (Test Sample; 0.01 prior; two-year)

type I cost type II cost	1 1	5 1	15 1	25 1	50 1	60 1	75 1	100 1
type I rate bt (banking) da	1.000 0.750	1.000 0.450	1.000 0.200	0.800 0.200	0.450 0.150	0.500 0.150	0.450 0.050	0.250 0.000
type II rate bt (banking) da	0.000 0.000	0.000 0.000	0.000 0.100	0.050 0.150	0.100 0.250	0.050 0.250	0.000 0.250	0.400 0.300
total cost bt (banking) da	0.010 0.008	0.050 0.023	0.150 0.129	0.250 0.199	0.324 0.323	0.350 0.338	0.338 0.285	0.646 0.297
type I rate bt (insurance) da	1.000 1.000	1.000 0.917	1.000 0.861	0.861 0.861	0.139 0.778	0.222 0.778	0.333 0.750	0.528 0.694
type II rate bt (insurance) da	0.000 0.000	0.000 0.000	0.000 0.028	0.028 0.028	0.389 0.028	0.333 0.028	0.278 0.028	0.222 0.056
total cost bt (insurance) da	0.010 0.010	0.050 0.046	*0.150 0.157	0.243 0.243	0.454 0.416	*0.463 0.494	*0.525 0.590	*0.748 0.749

*: Total cost of the binary tree is less than that of DA

Table 5. Test4 (Test Sample; 0.02 prior; two-year)

type I cost type II cost	1 1	5 1	15 1	25 1	50 1	60 1	75 1	100 1
type I rate bt (banking) da	1.000 0.600	1.000 0.350	0.800 0.200	0.450 0.150	0.250 0.000	0.400 0.000	0.150 0.000	0.150 0.000
type II rate bt (banking) da	0.000 0.000	0.000 0.000	0.050 0.150	0.100 0.250	0.350 0.300	0.100 0.350	0.300 0.400	0.350 0.450
total cost bt (banking) da	0.020 0.012	0.100 0.035	0.289 0.207	0.323 0.320	0.593 0.294	0.578 0.343	0.519 0.392	0.643 0.441
type I rate bt (insurance) da	1.000 0.972	1.000 0.889	0.861 0.861	0.167 0.778	1.000 0.694	0.528 0.639	0.389 0.472	0.528 0.361
type II rate bt (insurance) da	0.000 0.000	0.000 0.028	0.028 0.028	0.333 0.028	0.000 0.056	0.222 0.111	0.806 0.250	0.222 0.417
total cost bt (insurance) da	0.020 0.019	*0.100 0.116	0.286 0.286	*0.410 0.416	1.000 0.749	*0.851 0.876	1.373 0.953	1.273 1.131

*: Total cost of the binary tree is less than that of DA

As the task characteristics[6], we use dependence among the independent variables, linearity between the independent variables and the dependent variable, covariance equality of two classes, and normality of the independent variables. To test the dependence, correlation coefficients are used and P values show the significance of the coefficients. To test the linearity, linear regression is used and the F value shows the significance. Covariance equality and normality are tested by using an option of discriminant analysis and the Kolomogorov test respectively.

There are 36 correlation coefficients among 9 variables. 16 cases out of 36 (44%) show linear dependencies at the significance level of 0.1 using the one-year training data; 6 cases out of 36 (17%) using the one-year test data. Linear dependencies are shown for the 16 cases out of 36 (44%) at significance level of 0.1 using the two-year training data; 7 out of 36 (19%) using the two-year test data. Therefore, the data sets have a characteristic of weak linear dependence.

The one-year training data set has a linear relationship between independent variables and the dependent variable at the significance level of 0.1; while the one-year test data set does not. Both the training and test data sets of two-year, however, do have a linear relationship of significance level 0.1.

6) To test the task characteristics, SAS package is used.

Table 6. Characteristics of Insurance Data Sets

characteristics	training(one-year)	test(one-year)	training(two-year)	test(two-year)
dependence	44%	17%	44%	19%
linearity	yes	no	yes	yes
normality	0%	0%	0%	0%
equal covariance	no	-	no	-

Therefore, the data sets have a characteristics of good linearity between the independent variables and the dependent variable.

If a value from the Kolomogorov test is greater than zero and less than or equal to one, the variable is generally classified as having a non-normal distribution. The test shows that all variables of the four data sets have non-normal distributions at the significance level of 0.1. The equality of covariances is tested by a chi-square statistic, with the result that the two covariances of each year are different at the significant level of 0.1. The above results are summarized in Table 6.

From the above results, we can see that dependence is low and linearity is high. Two assumptions of DA (normality and equal covariance) are violated, resulting the data have characteristics of non-normality and non-equal covariance. These characteristics may explain the weak performance of DA.

6.2 Relative Task Characteristics

To see the performance of a model changes depending on a task, we introduce bank industry. The same data sets were used in the paper [20]. We extract the task characteristics from these bank data sets. Finally, we compare the characteristics of insurance with those of banking. We show that the performance of a model changes depending on the change of data characteristics. For example, DA shows a better performance in a domain which fits the assumptions of DA.

We use the same statistical methods. There are 171 correlation coefficients among 19 variables. 99 cases out of 171 (58%) show linear dependencies at the significance level of 0.1 using the one-year training data; 85 cases out of 171 (50%) using the one-year test data. Linear dependencies are shown for 91 cases out of 171 (53%) at significance level of 0.1 using the two-year training data; 78 out of 171 (46%) using the two-year test data. Therefore, the data sets are characterized as having a medium dependence.

Table 7. Characteristics of Bank Data Sets

characteristics	training(one-year)	test(one-year)	training(two-year)	test(two-year)
dependence	58%	50%	53%	46%
linearity	yes	yes	yes	yes
normality	11%	42%	5%	26%
equal covariance	no	-	no	-

The training and test data sets of both years have linear relationships between independent variables and the dependent variable at the significance level of 0.1. Therefore, the data sets have a characteristic of excellent linearity between the independent variables and the dependent variable.

The test shows that most variables of the four data sets have non-normal distributions at the significance level of 0.1. Two variables (11%) in the one-year training data set, eight variables (42%) in the one-year test data set, one variable (5%) in the two-year training data set, and five variables (26%) in the two-year test data set do not violate the assumption of normal distribution. The equality of covariance is tested by a chi-square statistic, with the result that the two covariances of each year are different at the significant level of 0.1. Table 7 summarizes these results.

In the bank domain, we can see that the dependence is medium and linearity is very high. In the insurance domain, the dependence is weak and linearity is high. The normality assumption of DA is violated more strongly in insurance than in banking. This implies that DA may perform better in the banking domain.

7. CONCLUSIONS

This paper applies the binary tree classifier and discriminant analysis methods to predicting failures of banks and insurance companies. While discriminant analysis requires the assumptions that all variables have normal distributions and all classes have the same covariance, the binary tree classifier does not require any assumptions on the sample distribution or its class dispersions. Either method incorporates both prior probabilities and misclassification costs. All these conditions are considered both in trainings and evaluations.

In most experiments using bank data, discriminant analysis method shows better performance in terms of the type I error rate and the total cost. Only four experiments out of thirty two show less total cost of the binary tree than discriminant analysis: test sample, 0.01 prior, one-year data, type I cost of (1 or 5); test sample, 0.02 prior, one-year data, type I cost of (1 or 5).

In contrast, most experiments using insurance data show that the binary tree method performs better than discriminant analysis. Only 6 cases out of 32 show less total cost of discriminant analysis than the binary tree: test sample, 0.01 prior, two-year data, type I cost of (5 or 50); test sample, 0.02 prior, two-year data, type I cost of (1, 50, 75, or 100).

The performances of both the binary tree classifier and discriminant analysis depend on the parameters: failure prior probability, data used, type I error cost, type II error cost, and validation method. In this study, discriminant analysis is generally better than the binary tree classifier in the classification of bank defaults; the binary tree is generally better than discriminant analysis in the classification of insurance company defaults.

Default predictions of insurance companies show contradictory results to those of banks, confirming the previous results [9, 22] that the binary tree classifier shows better performances than discriminant analysis. This situation can be explained that the performance of a classifier depends on the characteristics of the data. If the data are dispersed appropriately for the classifier, the classifier will show a good performance. Otherwise, it may show a poor performance.

The two data sets (bank and insurance) are analyzed using four characteristics: dependence among the independent variables, linearity between the independent variables and the dependent variable, covariance equality of two classes, and normality of the independent variables. The results show that insurance data have more independence, more non-linearity, and less normality compared with bank data. These results imply that the binary tree classifier performs better and discriminant analysis worse in insurance data. As explained before, the binary tree method builds a tree recursively and terminal nodes show classes. The terminal nodes are independent and they have non-linear relationships each other. Thus, the binary tree classifier may show better results in data with more independence and more non-linearity. One of the two assumptions of discriminant analysis is normality. Discriminant analysis may perform worse in data with less normality. Therefore, the characteristics of the two data sets explain the better performance of the binary tree in insurance and the worse performance in bank; the better performance of discriminant analysis in bank and the worse performance in insurance.

The limitation of this study is not to show statistical significant tests in performance differences between the two classifiers (Tables 2-5) and in differences between the two domain data characteristics (Tables 6-7). The performance of a classifier is thought to depend on the distributions of the variables used. Then, the research question becomes what is the relationship between the dispersions of the variables used and a classifier not between the domain and a classifier. If data for the relationships are accumulated, we may choose the best classifier based on the given data by analyzing the data characteristics.

REFERENCES

[1] Altman E. I., "Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy," Journal of Finance, Vol. 23 (1968), pp. 589-609.

[2] Ambrose, J. M., "Bests Ratings, Financial Ratios and Prior Probabilities in Insolvency Prediction," Journal of Risk and Insurance, Vol. 55, No. 5 (1988), pp. 229-243.

[3] BarNiv, R. and Smith, M. L., "Understanding Investment and Solvency," *Journal of Insurance Regulation*, Vol. 5, No. 4, June (1987), pp. 409-428.

[4] Breiman, L., Frieman, J., Olshen, J., and Stone, C., *Classification and Regression Trees,* California, Wadsworth & Brooks, 1984.

[5] Chung, H. M. and Silver, M. S., "A Comparative Analysis of Inductive-Learning Algorithm, Intelligent Systems in Accounting," *Finance and Management,* Vol. 2, (1992), pp. 3-18.

[6] Chung, H. M. and Silver, M. S., "Rule-Based Expert Systems and Linear Models: An Empirical Comparison of Learning-By-Examples Method," *Decision Sciences,* Vol. 23, No. 3, (1992), pp. 687-707.

[7] Denenberg, H. S., "Is A-Plus Really a Passing Grade? Insurer Risk Capacity and Financial Ratings: The Use of Financial Ratings of Insurance Reporting Services to Avoid Potentially Insolvent or Otherwise Delinquent Insurers," *Journal of Risk and Insurance,* Vol. 34, No. 3, September, (1967), pp. 371-384.

[8] Eisenbeis, R. A., "Pitfalls in the Application of Discriminant Analysis in Business, Finance, and Economics," *Journal of Finance,* Vol. 32, (1977), pp. 875-900.

[9] Frydman, H., Altman, E. I., Kao, D., "Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress," *Journal of Finance,* Vol. 11, No. 1, March, (1985), pp. 269-291.

[10] Koo, B., *Solvency Surveillance in the Property/Casualty Insurance Industry: A Financial Analysis and Statistical Evaluation of Sampling Biases,* Unpublished Ph. D. Dissertation, The University of Texas at Austin, 1992.

[11] Korobow, L. and Stuhr, D., "Performance Measurement of Early Warning Models," *Journal of Banking and Finance,* Vol. 9, (1985), pp. 267-273.

[12] Lann-Tennant, J., Starks, L., and Stokes, L., "Solvency Surveillance: An Empirical Evaluation of the Property-Liability Insurance Industry," *Proc. Third Int. Conf. on Ins. Fin. and Solvency,* 1990.

[13] Liang, T., "A Composite Approach to Inducing Knowledge for Expert Systems Design," *Management Science,* Vol. 38, (1992), pp. 1-17.

[14] Meyer, P. A. and Pifer, H., "Prediction of Bank Failures," *Journal of Finance,* Vol. 25, (1970), pp. 853-868.

[15] Mingers, J., "An Empirical Comparison of Pruning Methods for Decision Tree Induction," *Machine Learning,* Vol. 4, (1989), pp. 227-243.

[16] Mooney, R., Shavlik, J., Towell, G., Gove, A., "An Experimental Comparison of Symbolic and Connectionist Learning Algorithms," *Proceedings of Eleventh IJCAI,* 1989.

[17] Quinlan, J. R., "Induction of Decision Trees," *Machine Learning,* Vol. 1, (1986), pp. 81-106.

[18] Rumelhart, D. E., Hinton, G., and Williams, R., "Learning Representation by Back-propagation Errors," *Nature,* Vol. 323, No. 9, (1986), pp. 533-536.

[19] Sinkey, J. F., "A Multivariate Statistical Analysis of the Characteristics of Problem Banks," *Journal of Finance,* Vol. 30, (1975), pp. 21-36.

[20] Tam, K.Y. and Kiang, M., "Neural Net Applications in Bank Failure Predictions," *Management Science,* Vol. 38, (1992), pp. 926-947.

[21] Trieschmann, J. S., and Pinches, G. E., "A Multivariate Model for Predicting Financially Distressed Property-Liability Insurance Companies," *Journal of Risk and Insurance,* Vol. 40, No. 3, September, (1973), pp. 327-338.

[22] Weiss, S. M. and Kapouleas, I., "An Empirical Comparison of Pattern Recognition, Neural Nets, and Machine Learning Classification Methods," *Proceedings of the Eleventh IJCAI,* 1989.

[23] West, R. C., "A Factor-Analytic Approach to Bank Condition," *Journal of Banking and Finance,* Vol. 9, (1985), pp. 253-266.