Analysis of Insurance Attrition Clients Using Decision Tree and Neural Network Model

Su Teak Kim · Jong Hoo Choi.

Department of Informational Statistics Korea University.

208, Jochiwon-up, Yongi-gun.

Chungnam, Korea.

stkim@apstat.korea.ac.kr

jchoi@tiger.korea.ac.kr

1. Introduction

The purpose of this work is an analysis of the attrition clients of insurance company. Also, we want to find a recommendable model for evaluating clients through attrition score of clients.

Nowadays, it is acknowledged that the client behavior scores are elementary information for evaluation of clients in the sense of marketing decision making. Specially, in the view of database marketing, it offers us valuable information about strategy of clients segmentation for target marketing in behalf of business enterprise. There are so many types of client behavior scores in the world of business enterprise, that is, the attrition scores in insurance company, the credit scores of clients in banks and so on.

In this work, we use the decision tree model and neural network model those are well known and widely used in data mining world as analytical tool.
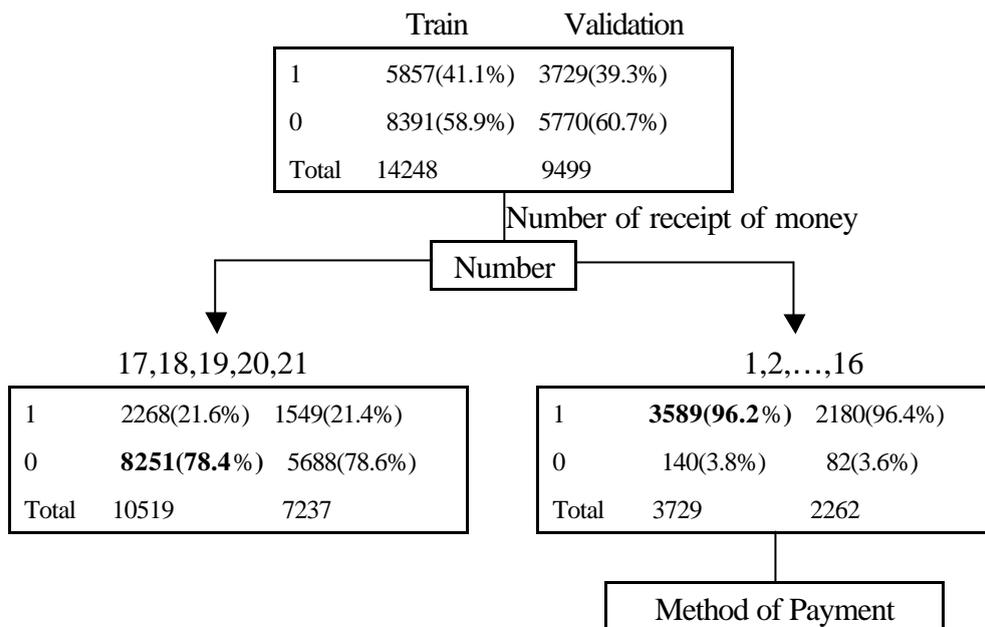
2. Data Description

We use a real world data, which is from L insurance company in Korea. The data contains 118,733 clients information. We use the variable, 'situation(1 : attrition , 0 : regularity)', as target variable. And the other 15 variables those are representative for client attributes are explanatory variables. Those are gender, age, residential area, method of payment, and so on. Among the analytical data mart, it is revealed that the attrition rate of clients is 40.2%.

First of all, we select a sample for analysis from the data mart. The sample size is 23,747, which is 20% of the analytical data mart. Secondly, we separate the sample into two parts, one training data set and the other validation data set.

3. Analysis

We apply to decision tree model and neural network model in classifying of clients whether he/she is on attrition or not. In order to obtain decision tree, CHAID(CHi-squared Automatic Interaction Detection) algorithm is applied[1]. The obtained decision tree is as follows;

Among several neural network models, MLP(Multi-Layer Perceptron) that contains one hidden layer and three hidden units is adapted[2].

	Train	Validation
1	5857(41.1%)	3729(39.3%)
0	8391(58.9%)	5770(60.7%)
Total	14248	9499

Number of receipt of money

Number

17,18,19,20,21

	Train	Validation
1	2268(21.6%)	1549(21.4%)
0	**8251(78.4%)**	5688(78.6%)
Total	10519	7237

1,2,…,16

	Train	Validation
1	**3589(96.2%)**	2180(96.4%)
0	140(3.8%)	82(3.6%)
Total	3729	2262

Method of Payment

<Figure 1. Decision Tree>

4. Assessment of Model

We use a lift chart, ROC (Receiver Operating Characteristic) curve and classification table for assessment of the models. As a result, we can find out that decision tree model is preferable to neural network model as Table1.

<Table1>

	Error rate	Accuracy	Sensitivity	Specificity
Decision Tree Model(p=0.2)	0.0913	0.9087	0.7970	0.9809
Neural Network Model	0.0988	0.9011	0.7908	0.9724

5. Discussion

After we obtain a recommendable model, we can calculate attrition score of clients. Logit model is widely used to do this. It is considered that the attrition score of clients is an elementary information for clients segmentation in insurance company.

REFERENCE

[1] Bigg, D., B. de Ville and E. Suen (1991). A Method of choosing multiway partitions for
classification and decision trees, *Journal of Applied Statistics*, 18, 49-62
[2] Murray Smith (1996). *Neural Networks for Statistical Modeling*, International Thomson
Computer Press.