

THE STUDY ON CONSTRUCTION OF CUSTOMER DEVELOPMENT MODEL – AN EXAMPLE OF HEALTH CLUB IN TAIWAN

Chen Hung-Yen, Fu Jen Catholic University, Taiwan, 035404@mail.fju.edu.tw

Chen Chi-Wen, National College of Physical Education and Sports, Taiwan

Lee Shih-Yuan, China Institute of Technology, Taiwan

INTRODUCTION

It is very difficult to choose the target customers correctly, touch them further, and reduce the churn rate of them in potential customers. Especially in customer development field, it is a dilemma to face huge potential customers and want to reduce the cost of customer development.

There had been very examples to use data mining technology to analysis customer data, product data and trade data in order to understand the customer information of preference and habit, and increase sales and cross sale that built the long-term relationship between enterprises and customers. In view of this, the objective of the proposed study is to explore the performance of customer development using discriminate analysis, logistic regression as well as artificial neural networks (ANNs) and multivariate adaptive regression splines (MARS), two new classification techniques of data mining. The customer development model was constructed by members' data of Taiwanese sports club (the following substituted as G club) and relevant information.

METHODS

1. Selection of the Subjects: The subjects of this study were from the membership database provided by G club in Taiwan. There are total of 397 subjects who had tried to experience the G club one time before January 2005. After eliminating subjects with incomplete and illegitimate information, there are total of 306 subjects.
2. Instrumentation and Procedures: To evaluate the customer development model, the structure of this study was divided into two stages: model training and model testing. The first stage is the model training that is to establish a customer development model. The next stage is the model testing, which was to apply the first stage model to predict the correct rate that whether the customer would become the membership or not. The number of subjects in training and testing stage was decided as 80:20 random samples to avoid errors in research result. The ratio was based on the majority of the studies in the field of data mining.
3. Plans for Research Design and Data Analysis: There are 17 variables in the member database provided. A random sample of 245 subjects was examined in the training stage and a random sample of 61 subjects was examined in the testing stage.

RESULTS

1. The whole correct classification rate of discriminant analysis was 78.69%.
2. The whole correct classification rate of logistic regression analysis was 80.33%.
3. The whole correct classification rate of ANNs was 88.52%.
4. The whole correct classification rate of MARS was 93.44%.

DISCUSSION

For G club, the whole correct classification rate of four classification tools was above 70%, it was meant that the important variables found by this study were extreme importance for G club. According to these important

variables, G club should formulate every kind of plan in order to design the customized marketing projects of need-satisfaction. To do so, G club would increase the number of membership and lower the cost in acquiring new members.

Future studies should aim at collecting more important independent variables that may further increase the classification accuracy. These variables, such as life styles, consumer behaviours, satisfaction, and even company imaginations are important factors that this study cannot collect to transform them into an effective function. Otherwise, using other classification techniques, such like the classification and regression tree (CART), bagging and boosting, fuzzy discriminant analysis, and support vector machines (SVM) in evaluating their applicability to customer development models is also recommended for further study.