Analysis of the Business Customer Churn Based on Decision Tree Method

Hongxia Ma Min Qin Jianxia Wang
College of Information Science and Engineering, Hebei University of Science and Technology
Shijiazhuang 050018, China
Email: *mhx_92@126.com*

Abstract–*Customer churn analysis is a predictive analysis applying data mining technology. In this paper decision tree method of data mining technology is applied to analyze business customer churn to identify the customer churn characteristics and some of rules of the customer churn. Thus businesses can in a targeted manner to improve customer relationships. The experiment of customer churn analysis on decision tree is illustrated deliberately. It provides a new research ideas and analysis methods for business customer churn and early warning analysis.*

Keywords– *Customer churn, data mining, decision tree, information entropy.*

I. INTRODUCTION

Customer churn refers to customers terminating the service contract with the company or turning to services provided by the other company. Because of various uncertainty factors and market growth as well as the existence of some of our competitors, a lot of customers continuously turn from one supplier to another supplier to seek lower prices and better services. At present, CRM (customer relationship management) has become a magic weapon for the business to win the fierce market competition. CRM emphasize the customer is core. It requires business completely understanding the entire life cycle of customers, and providing a unified, integrated platform and tools communicating with customers. All customer-related information is dealt with on this platform to enhance the efficiency of employees contacting with customers and customer feedback. At the same time, data mining technology is used on the customer information analysis to acquire valuable information and knowledge [1]. How to more efficiently find potential new customers as well as maintain and improve relations with customers has become a key factor for business to occupy market. Customer churn is one of the important questions of customer relationship management.

Data mining is a technology of finding valuable hidden events from a huge database. It identifies the knowledge by in-depth analysis of the data using statistics and artificial intelligence technology, and establishes different models in accordance with the question to provide a policy decision reference for business. Its powerful data analysis function can be used in extracting effective information from the mass customer data. Through dada mining, the customer churn characteristics can be found, thus businesses can take targeted measures to prevent those customers with similar characteristics from churning. And businesses through it can also find the characteristics of clients, so that businesses can sell their products to those customers who have the same characteristics but have not buy products. With the applications of data mining in such areas as business intelligence, decision tree method will play more and more powerful role in the future [2]. This paper provides a method of customer churn crisis based on decision tree. Also it describes how to apply the decision tree method analyzing business customer information to get the characteristics of the customer churn and customer churn rules. Based on this analysis, businesses can improve customer relations in a targeted manner and avoid customer churn.

978-1-4244-3864-8/09/$25.00 ©2009 IEEE

II. THE DECISION TREE METHOD

A. The concept of decision tree

Data mining technology includes case-based reasoning, neural network, decision tree, expert system, probability rule, stochastic model based on knowledge, fuzzy logic, evolutionary computation, and so on. In which decision tree method is an induction algorithm based on examples. It derives classification rules in the form of decision tree from a group of non-priority, non-rules examples. It uses of top-down recursive way to compare the attribute values in the decision tree internal node, according to different attribute values to determine the next node from the branch. In the decision tree leaf nodes the conclusion is acquired.

The basic principle of decision tree is recursively split the data into sub-variables which are predictable properties. Each split of the tree, it is necessary to evaluate the impact all the input attributes to predictable attributes. When this recursive is end a decision tree is created [3].

Decision tree method has these characteristics, such as less decision tree, high classification accuracy rate and simple rules generated, so it has a wide range of applications [4]. Since decision tree method is using information theory principle to make the information amount analysis of the substantial examples characteristics, computing the information entropy of those characteristics, to identify the important

characteristics that can reflect the categories, and therefore it grasps the nature of the problem. The error rate of decision tree algorithm is relatively small. The forecast effect of decision tree algorithm is better, but also it is very simple and practical.

B. Decision tree

Decision tree classification algorithm has CLS, ID3, C4.5, CART and other algorithms. The classification of decision tree includes two steps. The first step is to use the training set to establish and refine a decision tree. Thus decision tree model is set up. This process is in fact an acquisition of knowledge from the data and machine learning process. The second step is to use that decision tree to make classification of input record. About an input record, testing the value of the record's property from the root node until one of leaf nodes, thus the record's category is found.

This paper adopt ID3 algorithm. This algorithm introduces information gain of information theory as metric in selecting the essential property of entity. The biggest information gain property is selected to generate the decision tree node. Different values of this node are used to establish the branch, then recursive on the branch to establish the next node and branches. Using this method set up decision tree nodes and branches, up to the examples of a certain subset belonged to the same category.

Set E is a collection contains of samples. Category can take *m* different values correspond to different category C_i, *i* belong to (1, 2, 3, ..., m). If selected *A* property as the test property, *A* has *v* different values (a_1, a_2, a_3, ..., a_v).*A* will divided set *E* into v sub-sets (E_1, E_2, ..., E_v), set E_{ij} is the sample size for the subset E_j belong to C_i. When *A* is used to divide the current sample collection, the required information entropy is calculated as follows:

$$E(A)=\sum_{J=1}^{V}\frac{E_{1j}+\ldots+E_{mj}}{|E|}\ I(E_{1j},\ E_{2j},\ \ldots,\ E_{mj}) \tag{1}$$

Set *E*, its information entropy is

$$I(E_1,\ E_2,\ \ldots,\ E_m)=-\sum_{i=1}^{m}\frac{|E_i|}{|E|}\ log2\frac{|E_i|}{|E|} \tag{2}$$

Of a given subset E_j, its information entropy is

$$I(E_{1j},\ E_{2j},\ \ldots,\ E_{mj})=-\sum_{i=1}^{m}\frac{E_{ij}}{|E_j|}\ log2\frac{E_{ij}}{|E_j|} \tag{3}$$

$P_{ij} = E_{ij} / |E_j|$, P_{ij} is the probability of the samples of subset E_j belong to category C_i. The information gain obtained by using property *A* to divide the current branch node to get corresponding samples set is

$$Gain\ (A) = I\ (E_1,\ E_2,\ \ldots,\ E_m\ ) - E\ (A) \tag{4}$$

Firstly calculating the information entropy for each property, and then selecting the biggest gain property as a test property of given set *E*, and producing the corresponding branches and nodes.

III. CUSTOMER CHURN ANALYSIS EXAMPLE

A. Decision tree method example

In data mining the data need be pre-processed. In general, pre-processing includes data selection, transformation, purification, statute and so on. There are a lot of data attributes in the database. We use attribute-oriented induction method to remove irrelevant or weak relevant property. First of all those properties will be deleted which have too much value and can not be generalized, such as name, address, telephone and so on. Second, correlation analysis, including two aspects: on the one hand, to reduce redundancy between the input variables, on the other hand to ensure the correlation between that input and output variables. Some property can be accepted or rejected in accordance with intuitive judgment. Followed by the least square method is used to calculate the correlation coefficient of every property and customer churn. Then, these properties that have greater correlation coefficient are accessed in analysis of customer churn.

Through data pre-processing, we select the following property as analysis example. These properties include "after-sales service satisfaction", "North China areas", "large scale" and "transactions times>20". Data set is divided into positive example sets and negative example sets, that is churn and no churn. The training sample model is as shown in table1.

Table 1. The training samples

customer No.	North China areas	large scale	transactions times> 20	after-sales service satisfaction	churn or no churn
0001	True	True	False	≤5	Yes
0002	True	True	False	5-8	Yes
0003	True	False	True	>8	No
0004	False	True	False	>8	Yes
0005	True	False	True	5-8	No
0006	True	True	True	>8	No
0007	False	True	False	>8	No
0008	True	True	True	>8	No
0009	False	True	False	5-8	Yes
0010	True	False	False	5-8	No
0011	True	True	True	5-8	No
0012	False	True	False	≤5	Yes
0013	False	True	False	>8	Yes
0014	True	True	True	5-8	No
0015	False	True	False	≤5	Yes
0016	True	False	True	5-8	Yes
0017	False	False	False	>8	No
0018	True	False	False	5-8	No
0019	False	True	False	≤5	Yes
0020	True	False	True	5-8	No

Using ID3 algorithm to make data classification and construct decision tree. Specific calculation process is as follows.

(1) Information entropy calculation of the given samples set

From Table 1, we can see that examples belong to the not churn is 11 and examples belong to the churn is 9.

$P(u1)=\frac{11}{20}$, $p(u2)=\frac{9}{20}$

So Information entropy of the given samples set is

$I(11, 9)= -\frac{11}{20}\log 2\frac{11}{20} - \frac{9}{20}\log 2\frac{9}{20} =0.99277445$

(2) Calculating each property's information entropy

For example "after-sales service satisfaction" property:

After-sales service satisfaction>8 (total 7): 5 positive examples, 2 negative examples

$I(5, 2)= -\frac{5}{7}\log 2\frac{5}{7} - \frac{2}{7}\log 2\frac{2}{7} =0.86312057$

After-sales service satisfaction is between 5 and 8(total 9): 6 positive examples, 3 negative examples

$I(6, 3)= -\frac{6}{9}\log 2\frac{6}{9} - \frac{3}{9}\log 2\frac{3}{9} =0.91829583$

After-sales service satisfaction≤5 (total 4): 0 positive examples, 4 negative example

$I(0, 4)= 0$

The information entropy of "after-sales service satisfaction" property is

$E(after\text{-}sales\ service\ satisfaction)= \frac{7}{20} I(5, 2) + \frac{9}{20} I(6, 3) + \frac{4}{20} I(0, 4)=0.71532532$

(3) Calculating each property's information gain

$Gain(after\text{-}sales\ service\ satisfaction) = I(11,9) - E(after\text{-}sales\ service\ satisfaction) = 0.99277445 -0.71532532 =0.27744913$

Similarly, other property's information gain can be calculated.

$Gain(large\ scale) = I(11, 9) - E(large\ scale)=0.16088519$

$Gain(North\ China\ areas) = I(11, 9) - E(North\ China\ areas)=0.18149633$

$Gain(transactions\ times>20) = I(11, 9) - E(transactions\ times> 20)= 0.22437117$

We can see that the property "after-sales service satisfaction" has the largest information gain, that is, it provided the largest amount of information. So "after-sales service satisfaction" should be chosen as a classification property.

(4) Set up decision tree

Select "after-sales service satisfaction" property as the root node. This property has three value and leads to three branches. At this point, the training example set is divided into 3 sub-sets, generate a decision tree contains of three leaf nodes, as shown in Fig.1.

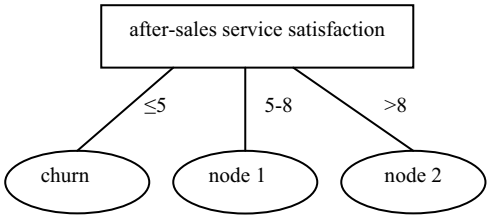

Fig. 1. Root node classification decision tree.

Continuous classification on node 1 and node 2 gained the final decision tree, as shown in Fig.2.

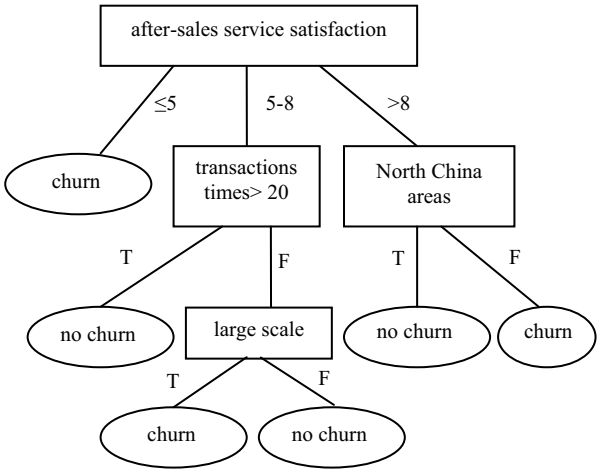

Fig. 2. Classification decision tree

B. Experimental result analysis

Through the classification of data and the construction of decision tree, business customer churn factors and the high churn customer groups can be found. Through analyzing the final decision tree, some rules can be draw.

(1)The first classification property of the decision tree is "after-sales service satisfaction" property. It shows that a product's after-sales service is the first one of the main influencing factors, which is in line with the results got by normal experience.

(2) When after-sales service in general, the customer churn has relation with transactions times and the size of the customer business.

(3) It will directly lead to customer churn when after-sales service can not meet the customer demand.

(4) When the customer is not located in the area of North China and after-sales service satisfaction is higher than 8 points, the customer normally does not select the business products. From here it can be seen that this business's customer has regional characteristics.

Ⅵ. CONCLUSIONS

This paper introduced how to use decision tree method of data mining to analyze business customer churn. It provides a new method for businesses to

analyze customer churn crisis. And a specific training sample set is used to make experiment of customer churn risks on decision tree. The experiment results show that the method is feasible and effective. Based on decision tree analysis, rules of customer churn can be draw. According to the characteristics of the churn customer groups, as well as relevant factors lead to customer churn, businesses can not only put forward corresponding countermeasures and improving measures from the perspective customer relationship management to retain customers, to prevent the churn of potential customers, but also by increasing the competitiveness to win the trust of customers, enhance their satisfaction and loyalty, so that they will re-become consumers.

In short, the easily understood rule information can be discovered by using the method of decision tree, which provides a new perspective and a new method for analysis and prevention of customer churn crisis.

REFERENCES

[1] Chen Sixing, Chen Baogang，"Analysis on the Loss of Clients Based on Manufacturing Industry CRM", Construction Machinery Technology and Management, No.3, pp.106-109, Mar. 2008.

[2] John Durkin, "Decision Tree Technique and its Current Research", Control Engineering of China, Vol.12, No.1, pp. 15-18, Jan. 2005.

[3] Deng Yong, Bao Jie. "The Application of Data Mining to Customer Churn Management", Design Technology of Post and Telecommunication, No.5, pp. 57-60, May.2008.

[4] Sheng Zhao-han, Liu Bing-xiang, "Research method of customer churn crisis based on decision tree", Journal of Management Sciences in China, Vol. 8, No.2, pp. 20-25,Apr. 2005.