The Application of Data Mining in Telecommunication Churn Customer

Hai-fei Qin

Computer Science Department, Chuxiong Normal University, Chuxiong, China

Abstract: This study analyzes telecom owe churn customers. Firstly, use the technology of OLAP to analyze the whole theme of the loss of arrears. Then, adopting cluster analysis analyzed consumer behavior (traffic, consumption amount, owed amount) before churn of the churn customer, age and in the net time and explaining the cluster. Finally, using the decision tree analyze the discrete variable and draws the conclusion.

Key words: Cluster analysis, customer churn, data mining, decision tree, OLAP

INTRODUCTION

Telecom market competition becomes increasingly, how provided high grade serves attracts and retains the customer, expands the market share, reduce the cost and increase the profit, has become a common goal (Mattison, 2005; Richter et al., 2009). In the highly competitive telecommunications market, of course, difficult to get a customer, but after how to get to keep, how to effectively manage customers, how to improve customer satisfaction and loyalty, how to prevent loss of customers, it is important content of telecommunication operation business. The customer churn already becomes the major issue which each competitive enterprise had to face, has become the hot spot and the difficult question which the theorists and the enterprise. Dass and Jain (2011) presents a comprehensive list of factors extracted by reviewing existing literature on various studies conducted for identifying determinants of customer churn in telecom industry. Yang et al. (2007) describes a comparative study investigating the influence of different reference group on consumer purchasing behavior between the mobile phone users of USA and China. Kuo et al. (2009) constructs an instrument to evaluate service quality of mobile value-added services and have a further discussion of the relationships among service quality, perceived value, customer satisfaction, and post-purchase intention. As the leading information industry, telecommunications has a huge data, how discovered the knowledge in the magnanimous data, found the rule, will be the matter which each operation business longs for even in dreams. The data mining took the cross many specialized new disciplines, processing data and discovery the knowledge have effectiveness unique.

The papers base on the item of Kunming telecom the customer churn analysis, It mainly analysis churn customer, according to analysis the behavior of the churn before churn, apply the technology of data warehouse and data mining to analysis data, analysis the trend of consumption, the aim is to find the character of the customer churn and papers for the further work analysis.

Fig. 1: Owed churn user income

BASED ON DATA WAREHOUSE DATA MINING

The development of data warehouse for the implementation of the data mining provides convenient, OLAP technology is part of the data mining, OLAP analyzes the data mainly comes from the multi-dimensional cube of the data warehouse, the data already passed through DTS (Data Transformation Services) and ETL (Extract, Transform, Load) transformation. So, this paper applied to OLAP analysis the whole churn theme on the multi-level analysis firstly. Such chart 1 and chart 2 as follows.

Through Fig. 1 and 2 analysis to the product dimension, obtained the most owed churn serious are PAS customer, the residence telephone, and LAN. From the owed amount, we can see the owed amount fluctuate seriously. Therefore, the characteristic of PAS owed churns may have very big difference with other product. So, we will track analyze to the question of the owe churn...
APPLICATION DATA MINING TECHNOLOGY ANALYSIS

Data preparation: As constructed the data warehouse, has already carried on Data Transformation Services, the data has been Extracted, Transformed, Loaded. So the work of data preparation has a basic. The data sample not only comes from data warehouse but also from the database primary data. From above, data warehouse is only analyzing the whole, after we may understand the question of the whole, also need understand stepped abstracts question, such as the traffic, the consumption amount and the owe amount among the owe. So we will analysis the owe churn customer churn previous 5 month partial expenses behavior data. So the work of data cleaning up and integrated will continue.

The data cleaning up also may be called the data clean. The work of the data clean is solved the data wrong, inconsistent object recognition process. Data clean including vacancy value processing, noise data processing and inconsistent data processing and so on. The data inconsistencies reduce the result of data mining credibility. The data mining needs to integrate the data and merge many data from difficult data source, after solving the semantic fuzziness saved in unified data storage. In order to the further cluster analysis, the numeric data, we needs replace null to zero.

Cluster analysis: Cluster analysis to call the group analysis, it is a method of multivariate statistical analysis which research the classification of sample and attribute. Cluster, popularly speaking, it is a set of the similar element. The classification is assigns an observation to some cluster (group). The function of cluster analysis is establishes a method of classified, according to the alienation and the similar in nature classify, we applying the method of sum of squares of deviations (the method of Ward) to churn customer age, in net time and previous five month expense behavior (traffic, consumption amount, owed amount) cluster. The method of Ward based on the thought of variance analysis, if classification correctly, then the sum of squares of deviations is small in same cluster, the sum of squares of deviations is big in different cluster. This paper applies the software of SAS (Statistic Analysis System) to age, in net time and previous five month expense behavior in churn customer. The results of cluster tree as follows Fig. 3.

From the Fig. 3, we can gain the different cluster according to different professional and different industry. According to the character of telecom and conveniently explained. We can classify the churn sample as eight clusters, but there are two clusters are unusual, save it to as unusual point analysis. There are six clusters are obvious, we will explain this cluster.

Remove unusual cluster data, analysis the data in cluster, The result of cluster as follows Fig. 4, 5 and 6.
From chart 4, chart 5, chart 6, the result of cluster, we can get the sixth cluster average consume and average owe fluctuate bigger, in this cluster have small average traffic. So we can gain this cluster is larger customer, this part customers might still linger before churn, they are not very satisfaction. So when the better products or better service appear, they are leave. The forth cluster average consume and average traffic present reduce tendency in previous churn, the average owe is not exist, explaining this cluster customer owed small, it can be omit. This part customer may become a major custom, should be encouraged consume, they are voice product. The third cluster average consumes and average owe present escalation trend, explaining this cluster customers are able to leave for owed. These customers have no traffic, explaining this cluster is non-voice products, is wideband and narrowband. The fifth cluster average consumes and average traffic is the highest in last month, but have no changed before two month, Illustrate some customers may leave with the purpose of intentionally owe. So should be to prevent this part customer fraud. The first cluster customer exist the certain proportion in all income and all owed, but in average is small, explaining these customer are relatively stable. The second cluster is the most stable among average consume, average owe, but average traffic. This part customers appears is not obvious, need analysis further.

In order to understand the behavior of consume each cluster, we compared the five months preceding the churn of data, the result as Table 1.

From the Table 1 we can see, regardless the minus between three monthly means and five monthly means, or the minus of last month consume and previous 5 months average consume, in the third cluster and the fifth cluster are higher and closer churn. So we analysis the third cluster and the fifth cluster fatherly, we can obtain the third cluster is ADSL (wideband and narrowband) customer, the fifth cluster is the PAS and the residence customer, others haven obvious characteristics. These two clusters previous churn are all raise, have the possibility to belong fraud user, we will analyzed it fatherly.

**Apply decision tree analysis:** The decision tree is a method of classification. Cluster not equal to classify. Cluster is Unconscious, classify is conscious. Classify mainly using in the attribute variable or the discrete data. Cluster mainly using in numeric or Continuous variable. According to numeric variable analysis above, each class was discredited. So we will adopt segments method discrete the numeric data, and as the training set to forecast other attribute variable.

The decision tree has based on the theory of information gain, when gain information is uncertain will be transformed. Therefore, the information content is uncertain or is the small probability event. According to small probability event have more information than big probability event. The decision tree classifies from the root node to some leaf node, the leaf node namely respective classification. Each leaf node showed some attribute test, and this leaf node each successor branch corresponds is the attribute possibility value.

The basic algorithm of decision tree is the greedy algorithm, it start from the top, build the decision tree. This paper see the whole set as root. Layer by layer down to end, the end is leaf node, as Fig. 7 shows, it adopt decision tree verification the result cluster.
From the analysis, we may get a result that if cluster not equal 3 and age is not null, the values is between 80 to 100, this part customer are mainly PAS user and officer telephone. Adopt Cluster equal 3, all these customers are ADSL user. This result confirmed the conclusion of cluster analysis. Step by step, we can draw the similar conclusion.

CONCLUSION

This article mainly uses the OLAP technology in the data mining, cluster analysis analyzes to the owe churn customer. Firstly, we using the OLAP technology to analyze the owe churn users as a whole, obtains the residence telephone, PAS and the wideband churn biggest. Then we using cluster analysis to analyze the expenses behavior before they churn (traffic, consume amount, owed amount), in net time and the customer’s age, obtains some cluster. Finally, we use the knowledge of professional and industry explaining the cluster. According to these analysis, we have a more comprehensive understanding to the owe churn customers, but also for the follow-up to lay the foundation.

REFERENCES


