Communications of the IIMA

Volume 7 | Issue 4 Article 11

2007

Mobile Customer Clustering Analysis Based on Call Detail Records

Qining Lin

Economics and Management School Beijing University of Posts and Telecommunications

Mobile Customer Clustering Analysis Based on Call Detail Records

Qining Lin
Economics and Management School
Beijing University of Posts and Telecommunications

ABSTRACT

Competition in the mobile telecommunications industry is becoming more and more fierce. In order to improve mobile operator's competitiveness and customer value, several data mining technologies can be used. One of the most important data mining technologies is customer clustering analysis. This targeting practice has been proven manageable and effective for mobile telecommunications industry. Most telecommunications carriers cluster their mobile customers by billing system data. This paper discusses how to cluster mobile customers based on their call detail records and analyze their consumer behaviors. Finally, an application of a mobile customer clustering analysis is given in this paper.

INTRODUCTION

Competition in the mobile telecommunications industry is becoming more and more fierce. Mobile operator's profits and ARPU (Average Revenue Per User) are facing tremendous challenges. Customer's demand become diversified, differentiation and requirements of service quality become more rational and strict. In order to improve mobile operator's competitiveness and customer value, several data mining technologies can be used. One of most important data mining technologies is customer clustering analysis. The aim of clustering is to categorize prospective customers into distinct groups for distinctive contact strategies and proximal offerings (Adriaans & Zantinge, 1996; Berson & Smith, 1997; Mattison, 1997; Russell, 1996; Berry & Linoff, 1997; Russell & Lodwick, 1999). This targeting practice has been proven manageable and effective for mobile telecommunications industry.

Most telecommunications carriers cluster their mobile customers by billing system data. Billing system data describe customer subscribe, spend and payment behavior. Call detail records describe customer utilization behavior. They have more information to describe customer behavior than billing system data. Therefore, this paper discusses how to cluster mobile customers based on their call detail records and analyze their consumer behaviors. Finally, an application of a mobile customer clustering analysis is given in this paper.

K-MEANS CLUSTER METHOD

There are many clustering method, for example, fuzzy clustering method, system clustering method, dynamic clustering method and K-means clustering method. But the K-means method of cluster detection is most commonly used in practice. It has many variations, but the basic form is unchanged.

Suppose that the number of mobile customers is P. Each mobile customer is described by an n-element vector $X_i = (x_{i1}, x_{i2}, \ldots, x_{in},)$, i =1, 2, …, P. K-means clustering method can be briefly stated as follows:

1) Select K data points or K mobile customers to be the seeds. The seeds are described by n-element vectors $E_j = (e_{j1}, e_{j2}, \ldots, e_{jn},)$, j = 1, 2, …, K. MacQueen's algorithm simply takes the first K mobile customers records. That is, let $E_j = (e_{j1}, e_{j2}, \ldots, e_{jn},) = X_j = (x_{j1}, x_{j2}, \ldots, x_{jn},)$, j = 1, 2, …, K.

2) Calculate the distance of between each seed and every mobile customer by the following formula:

$$d_{ij} = \left\{ \sum_{k=1}^{n} \left| x_{ik} - e_{jk} \right|^{q} \right\}^{1/q} \quad i = 1, 2, \ldots, P,\ j = 1, 2, \ldots, K$$

Here，q is a constant. Usually, let q = 2.

3) Every data point or mobile customer is assigned to one seed according the principle of minimum distance Calculated by step 2). In this way, we get P group data sets or P group mobile customers.

4) Calculate the centroids of P group data sets and form new K seeds by the following formula:

$$e_{jk} = \frac{1}{m_j} \sum_{i=1}^{m_j} x_{ik} \text{ , k = 1, 2, …, n, j = 1, 2, …, K}$$

Here, m_j denotes the number of j group mobile customers.

5) Return to step 2) or stop if the changes of cluster boundaries are small enough.

It is not difficult to program K-means clustering method mentioned above by software engineering. Some data mining software tools, for example, SPSS Clementine, SAS Enterprise Miner and IBM Intelligent Miner may be used in the mobile telecommunications industry.

PREPARE THE DATA FOR CLUSTERING

The data from billing system describe customer subscribe, spend and payment behavior. Usually, mobile billing system data include all kind of services billing data for customers to pay in every month. The data of call detail records describe customer utilization behavior. They record the data of every call detail records for each mobile customer. The data of some call detail records are shown in table 1.

Table 1: Mobile call detail records.

Customer_ID	Call_type	Start_Date	Start_time	Call_duration	Service_type
Service_code	Home_area_code	Visit_area_code	Called_code	Roam_type	Fee_type
common_fee/ Roam_fee	Long_fee	…			

It will take quite much time to prepare the data for clustering, especially to prepare the data of call detail records. Preparing the data of call detail records for clustering can be stated as follows:

1) Get the data of call detail records from data warehouse, usually takes several weeks;
2) Clean dirty or outlier data;
3) Define and calculate some index;
4) Form the data table for clustering. It is shown in table 2.

Table 2: Data for clustering which describe mobile customer utilization behavior.

Customer_ID	Local calling duration	Local called duration	Local calling number	Local called number
Local calling fee	Local called fee	Roam calling duration	Roam called duration	Roam calling number
Roam called number	Roam calling fee	Roam called fee	Long distance calling duration	Long distance called duration
Long distance	Long distance	Long distance	Long distance	IP calling

calling number	called number	calling fee	called fee	duration
Number of IP calling	IP calling fee	Idle period local call duration	Idle period local call number	Idle period roam call duration
Idle period roam call number	Idle period long distance call duration	Idle period long distance call number	Number of Short message	Number of multimedia message
GPRS duration	GPRS traffic volume	The number of calls to 10086	The duration of calls to 10086	The number of calls to 12530
The duration of calls to 12530	Music tone function fee	Music tone unload fee	Total number of calls	Total duration of calls
Call diameter	…			

Here, Idle period is 21：30—8：00. 10086 is the call center number of China mobile, and 12530 is the new service number of China mobile. Call diameter is defined as the number of different mobile customers called by or calling to one mobile customer in one month. This indicates the activeness of a mobile customer.

It is necessary, for most time, to define some new index, for example, the ratio of calling number and called number, the ratio of local calling duration and long distance calling duration, and so on.

CASE STUDY: A MOBILE OPERATOR’S CUSTOMERS CLUSTERING ANALYSIS

A mobile operator, located in north of China, has about 600,000 mobile customers. The operator wanted to cluster and analysis their all mobile customers because of competition last year. Therefore, a project team was formed to take charge of this project. The first thing need to do is to define an index set of elements which describes mobile customer utilization behavior, referring to Table 2. Then all mobile customers are clustered into 15 distinguishable groups using the mining software tools. The main characteristics of some groups are shown in Table 3.

Table 3: The main characteristics of some groups (mean value.

Items	Total customers	Group 1	Group 2	Group 3
ARPU	75.23	379.56	77.96	19.27
MOU	316.65	1122.58	418.25	81.34
Local call percentage	89.52	37.98	**96.04**	**97.58**
Long distance call percentage	4.98	**28.73**	1.67	0.96
IP call percentage	3.33	**9.45**	1.42	0.79
Roam percentage	2.17	**23.84**	0.86	0.67
Idle period local call percentage	10.92	7.89	**17.56**	**15.12**
Idle period long distance call percentage	9.85	6.78	**11.12**	**19.23**
Idle period roam call percentage	7.84	5.71	**10.98**	**18.34**
GPRS traffic volume	0.05	0.15	0.08	0.01
Number of short message	46.09	89.65	**226.34**	6.89
The number of calls to 12530	0.18	0.16	0.19	0.02
Call diameter	123.73	**287.89**	131.56	38.76

Here, MOU is Minutes Of Usage per user on the average.
Local call percentage = (Local call duration + Local called duration)/ MOU.
Long distance call percentage = Long distance call duration / MOU.
IP call percentage = IP calling duration / MOU.
Roam percentage = (Roam calling duration + Roam called duration)/ MOU.
Idle period local call percentage = Idle period local call duration / local call duration.
Idle period long distance call percentage = Idle period long distance call duration / long distance call duration.

Idle period roam call percentage = Idle period roam call duration / roam call duration.
From the data of table 3, the following facts can be deduced.

1) Group1, group 2 and group 3 are top, medial and low value mobile customers respectively.

2) The long distance duration percentage, IP duration percentage and Roam duration percentage of Group1 are obviously over the average index of total customers. Moreover, the call diameter of group 1 is also over the average index of total customers or other group customers.

3) The local call percentage, idle period local call percentage, idle period long distance call percentage, idle period roam call percentage and the number of short message number of group 2 are obviously over the average index of total customers.

4) The local duration percentage idle period local call percentage, idle period long distance call percentage and idle period roam call percentage of group 3 are obviously over the average index of total customers, and the all other index are less than the average index of total customers.

The marketing managements can use the facts above to design distinguishable marketing strategy in order to get better marketing results. For example, from their utilization behavior, we can see that the mobile customers of group1 often travel out, and marketing managements need take special care for their travel so as to improve their satisfaction. The mobile customers of group 2 have special concern about short message service and marketing managements can design suitable short message price policy for these customers. The mobile customers of group 3 have low consuming ability and marketing managements need to encourage these customers to use more mobile service.

CONCLUSION

The mobile telecommunication marketplace is highly competitive. The marketing managements of mobile telecommunication operators often need to design distinguishable marketing strategy based on different behavior of their mobile customers in order to improve their marketing result.

Call detail records describe customer utilization behavior. They have more information to describe customer behavior than billing system data. Clustering analysis based on call detail records can give more information than other clustering analysis for marketing managements. Defining some new index to describe mobile customer behavior is very useful for forming mobile customer clusters. It can help to achieve better distinguishable groups, so that marketing managements can design more suitable marketing strategy.

REFERENCES

Adriaans, P., & Zantinge, D., (1996). Data Mining. Addison Wesley, Harlowe, EngIand.

Berry, M. J.A., & Linoff, G. S., (1997). Data mining techniques: for marketing, sales, and customer. John Wiley & Sons, Inc.

Berson, A., & Smith, S. J., (1997). Data Warehousing, Data Mining, and OLAP. McGraw-Hill, New York, NY.

Mattison, R., (1997). Data Warehousing and Data Mining for Telecommunications. Artech House, Norwood, MA.

Russell, S., (1996). Neural Networks for Business Systems in Data Warehousing. McGraw-Hill, New York, NY.

Russell, S., & Lodwick, W., (1999). Fuzzy clustering in data mining for telecom database marketing campaigns. Fuzzy Information Processing Society, 1999. NAFIPS. 18th International Conference of North American. 10-12, 720 – 726.