

# **Classifying the segmentation of Telecom Customer Value Using Decision Tree Model**

**Lu ShuiXiu , Han ShuiHua**

(Department of Management Science, Xiamen University, 361005,Xiamen, P.R.China)

Abstract: Customer life-cycle prediction is the key to achieve long-term customer value. In terms of customer lifecycle, customer value is defined as customers' net cash flow and prospective profit, which is perceived from crucial deciders, and five decision models including current value, historic value, the prediction of long-term value, credit and loyalty are proposed. Considered that the difficult quantitative computation of long-term value, credit and loyalty, we apply data mining to extract important parameters of long-term value, credit and loyalty, our models present judgment matrix that is based on characteristics data and the experience of business expert, then a simplicity and utility practical appraisal system of customer value is built. This model is applied for telecom operator in some province in China and good accuracy is achieved.

Keywords: customer value, customer lifecycle, loyalty, credit, decision tree, AHP

## **1. Customer value**

With the development of telecom recombination, the business charged by China Mobile and China Unicom is now charged by the three telecom operator, China Mobile, China Unicom and China Telecom. The competition in respect to mobile

commerce is more intense. For China Telecom, the forceful competition is weakened by China Mobile which can do business in fixed telephone and broadband. Therefore, it is more and more important to keep up the customers with high value. The definition of the customers of high value is inseparable to compute customer value, however current methods for customer segmentation based on experience or ARPU (Average Revenue Per User) consider neither future revenue and customer lifecycle, nor the service cost of customers.

From a business point of view, the customer value can be the potential profit from customer along the customer's lifecycle. As far as we know, practical appraisal system of customer based on customer value has not been designed nowadays, and appraisal system of customer value is a still unsolved problem. In terms of customer lifecycle, customer value perceived from crucial deciders is defined as customers' net cash flow and prospective profit. According to Achim Walter etc[1], we classify customer value into two parts: one is direct value which is used to scale monetary effect, the other is indirect value which is used to scale the non-monetary effect. For direct value, we predict the lifecycle of in-net customer base on characteristics data, then compute customer long-term value, we consider all profit not only the earnings when computing the customer's current value. For indirect value, positive and negative sample are used to extract the characteristics data which may affect loyalty, then based on characteristics data and experience, business experts present judgment matrix needed by AHP method. Our purpose is to build a simplicity and utility practical appraisal system of customer value. The appraisal system is used to evaluate

customer's contribute and support the enterprise's decision.

## **2. The model of customer value computation**

We divide customer value into two parts: direct value and indirect value. Direct value refers to monetary value, that is, crucial deciders perceive the customers' net cash flow from the beginning of the lifecycle to the end of the lifecycle. It includes historic value, current value and long-term value. Long-term value is the prospective profit from customers, so how to predict customer's lifecycle is a critical issue. Indirect value refers to non-monetary value, which can't be quantitatively computed in general. Here we focus on the crucial determinations of indirect value, indirect value is computed indirectly based on loyalty and credit,. However, the computation of loyalty and credit is quite difficult and can't quantitatively be calculated, so it is necessary to structure the influence index.

The customer value computation includes five models: the computation of historic value, the computation of current value, the prediction of long-term value, the computation of loyalty, the computation of credit. The results of the five models must be dimensionless, then experts present the weight on customer value from historic value, current value, long-term value, loyalty, and credit, after then we get the ranking of custom value.

According to the ranking of customer value, the rank of customers is presented, then the key customers are obtained. According to the classification of customers, customers' manifest is obtained. Experts can adjust the customers' manifest based on

their experience. On the one hand, the adjusted manifest is fed back to experts. On the other hand, it is imported into data warehouse in order to reevaluate the customer value.

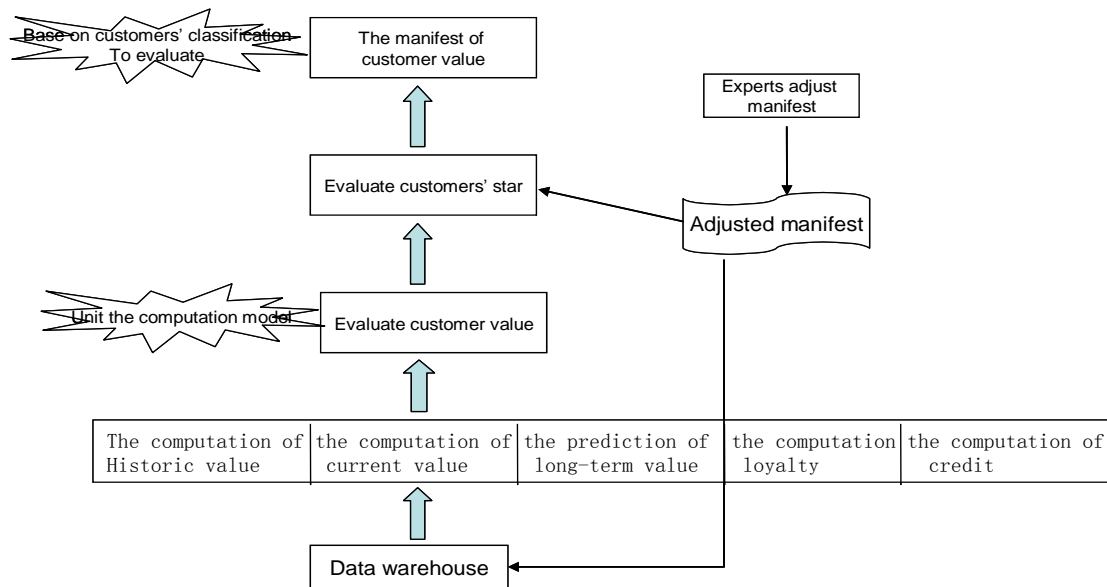


Fig1: the model of customer value computation

### 3. Empirical finding

#### 3.1 Data source

It is troublesome to compute customer value in some province in the telecom operator of China. The model we proposed in the paper solves the problem. The data is from the data warehouse the telecom operator of the some province, including about 220,000 customers. There is only the data of demographic statistic in the data warehouse.

#### 3.2 The application of the model

- (1) Historic value and current value

Historic value includes the accumulative charge and the accumulative cost. The accumulative charge is the actually charge and the charge shared with another operator. The accumulative cost is the sum of the cost shared with other operator, discount cost and channel reward. We use the latest the average net cast flow of the latest 3 months to compute current value.

## (2) Long-term value

We classify customer type into five kinds such as family customer , personal customer, enterprise or government customer , and key customer so on. Then Clementine12.0 is used to draw a bar graph of off-net customer based on the pre-processing data. See fig.2.

There are 3 high points. So the customer lifecycle is divided into 3 groups. Group 1 is the length of in-net time not more than 60 months but more than 36 months. Group 2 is the length of in-net time not more than 36 months but more than 18 months. Group 3 is the length of in-net time not more than 18 months.

We import the data directly from data warehouse and build the CART model. Fig.3 shows the model. We use percentage to set stopping criterion, where minimum records in parent branch is 5% and minimum records in child branch is 70%.

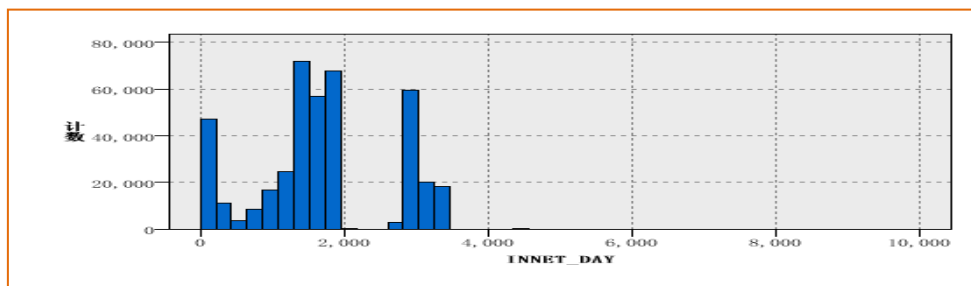


Fig.2: The bar graph of off-net customer

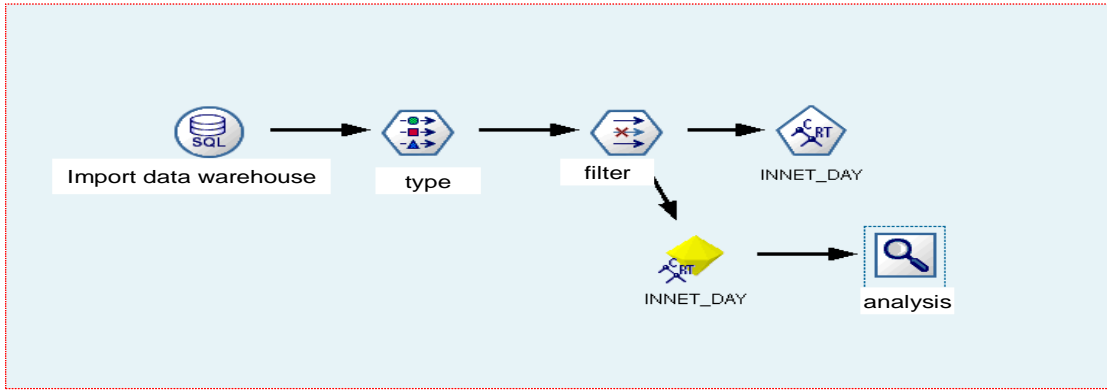


Fig.3: CART model

When 222829 customers is classified, the CART model can correctly predict about 149600 customers. Fig. 4 shows the result of the analysis. Table 1-4 show the accuracy, total amount and percentage (namely, the percent of customers who belong to the node in the 222829 customers) in the node 1,5,13 and 14.

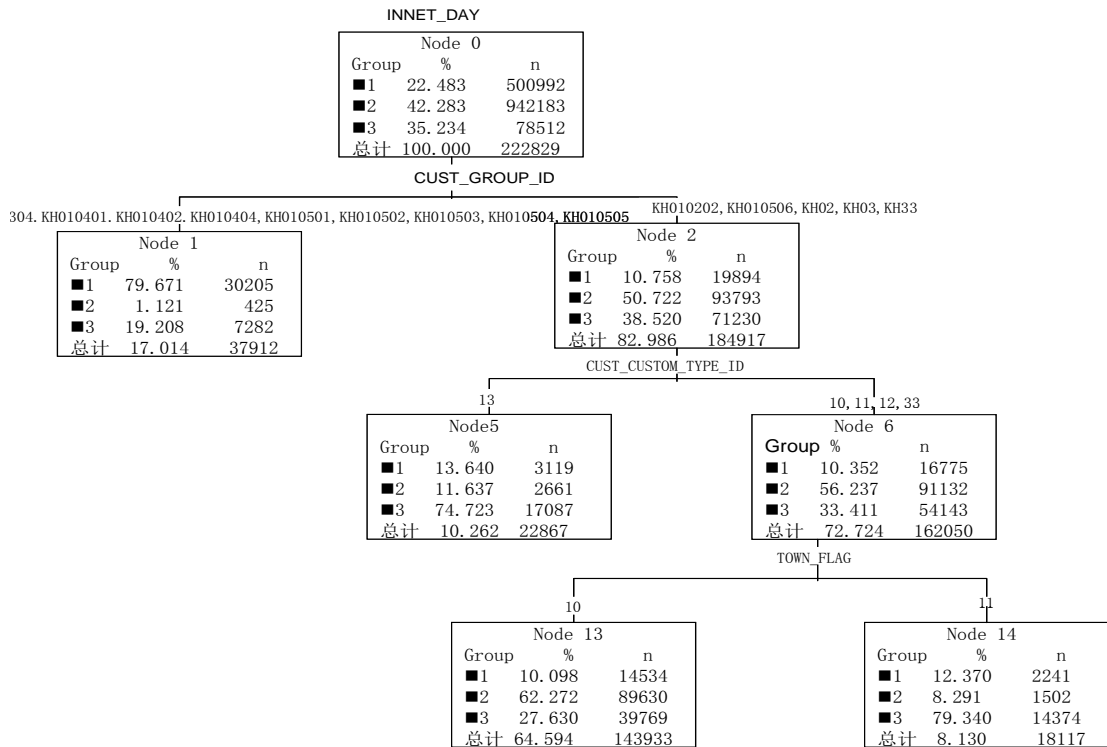


Fig. 4: The result of CART model

Based on the extracted characteristics of variable to classify customer lifecycle, we can predict the lifecycle of in-net customer. When the characteristics of business

of in-net customer is in accordance with the demographics of the determinants of the classification of customer lifecycle, we predict the customer belong to the classification and stop the prediction of customer lifecycle. For example, the customer classified type is familial type, customer type is familial customer and the flag is city, then we predict the customer belong to group 2, namely, we predict the length of the customer lifecycle is not more than 36 months but more than 18 months.

### (3) Loyalty

The determination of loyalty given by experts is the length of customer's in-net time. According to the length of customer's in-net time, the positive sample is the customers whose in-net time is between 8 years and 15 years. The total of the positive sample is 200,000. The negative sample is the off-net customers who have been off-net in the latest 6 months and the length of whose in-net time is short (less than 1 years). The total of negative sample is 100,000.

According to the positive and negative sample, based on C5 decision tree, we get the determinations of loyalty. They are the behavior in the net( the weight is 0.38), the charge of communication(the weight is 0.29), the abnormal consumption behavior(the weight is 0.25), the amount of scomb( the weight is 0.03), the kind and the amount of the product(the weight is 0.02).

Based on the judgment matrix given by the experts, we use AHP to get the weight the variable. Then we compute the hit rate of loyalty, and test the model. Table 1 shows the result of the verification.

Table 1 : the hit ratio of loyalty

Time	The in-net time is more than 8 years	The order is before	Hit ratio
2009.01	1192713	1076045	90.22%
2009.02	1195947	1062780	88.87%
2009.03	1165379	1017797	87.34%
2009.04	1174831	1035261	88.12%

(4) Credit

The determination of credit given by experts is that if the customers have the behavior fallen into arrears. According to the arrears, the negative sample is the customers who have the behavior fallen into arrears in the last 6 months. The total of the negative sample is 100,000. The data is chosen at random. The positive sample is the customers who have never fallen into arrears since they enter into the net. The total of positive sample is 200,000. The data is chosen at random.

According to the positive and negative sample, based on C5 decision tree, we get the determinations of loyalty. They are the frequency of delayed payment (the weight is 0.55), the payment type(the weight is 0.13), the product structure(the weight is 0.12), the remaining balance( the weight is 0.11), the customer type(the weight is 0.03).

Based on the judgment matrix given by the experts, we use AHP to get the weight the variable. Simultaneously, the variable must be nondimensionalization and the credit is the weighted mean. Then we compute the hit rate of loyalty, and test the

model. Table 2 shows the result of the verification.

Table 2: the hit ratio of credit

Time	The in-net time is more than 8 years	The order is before	Hit ratio
1	1530089	1423429	93.03%
2	2054504	1929116	93.90%
3	2220098	2048152	92.26%
4	2453267	2029405	82.72%

#### (5) Customer value

We utilize hit ratio to test the accuracy of the computation. The hit ratio refers to the percentage of the customers whose rank is before in the customers whose receivable earning after privilege is more than 50(ARPU).

The amount of the customers whose receive earning after privilege is more than 50(M), refers to the total of the customers that the earning available from who is more than 50 yuan in current month. The amount of customers whose rank is before (N) refers to the total of the customers whose order is before M-th, and whose receive earning after privilege is more than 50.

Hit ratio=the amount of the customers whose rank is before/ the amount of the customers whose receive earning after privilege is more than 50 =M/N

#### (6) Assessment of the model

In order to test the accuracy of the model, we utilize hit ratio to test the accuracy of the model. Here we apply two methods, in the view of ARPU and current value.

In the view of ARPU, the hit ratio refers to the amount of the customers whose rank is before/ the amount of the customers whose receive earning after privilege is more than 50.

In the view of current value, the hit ratio refers to the amount of the customers whose rank is before/ the amount of the customers whose current value is more than 50.

Table 3(a) shows the result of the hit ratio in the view of ARPU. Table 3(b) shows the results of the ratio in the view of current value.

Table 3(a): the hit ratio of customer value(ARPU)

Date	the amount of the customers whose receive earning after privilege is more than 50.	the amount of the customers whose rank is before	Hit ratio
2009/01	1622345	1397701	86.15%
2009/02	1514837	1317870	87.00%
2009/03	1482836	1283909	86.58%
2009/04	1482617	1299837	87.67%

Table 3(b): the hit ratio of customer value(current value)

Date	the amount of the customers whose current value is more than 50	he amount of the customers whose rank is before	Hit ratio
2009/01	1600449	1527924	95.47%

2009/02	1701695	1624843	95.48%
2009/03	1716000	1640139	95.58%
2009/04	1654274	1580533	95.54%

#### **4. Conclusion**

With the development of customer relationship management and the research, customer value is more and more important to support the decision of the enterprise. How to apply customer value in evaluating customer is the focal point for many enterprises. But there isn't applied appraisal system based on customer value as far as we know. It is still a hard work to structure the index of appraisal system and to compute customer value.

Here we structure a brief and applied appraisal system base on customer value. The customer's contribution is evaluated and quantized to support the decision of the enterprise. The model to predict customer lifecycle can solve when there is only demographics in the data warehouse. The models to compute loyalty and credit use AHP to get loyalty and credit. But when the experts score, they not only base on their experience, but also base on the characteristics data of the data warehouse. It overcomes the subjective of the experts' scoring in some degree.

Due to the actually condition constraint, there is some limit. We don't pay attention to the advantage of the models proposed in the paper compared with another models. We will focus on it in future. When we compute the long-term value, considered the utility of the computation, the computation of the monthly average

long-term value is relatively simple. Further research on it is needed.

### Reference

- [1] Hsin Hui Lin, Yi Shun Wang. An examination of the determinants of customer loyalty in mobile commerce contexts[J]. *Information & management*,2006,43:271-282.
- [2] Nattakarn Eakuru, Nik Kamariah Nik Mat. The Application of Structural Equation Modeling (SEM) in Determining the Antecedents of Customer Loyalty in Banks in South Thailand. *The Business Review*[J],2008,10(2)
- [3] Duen-Ren Liu, Ya-Yueh Shiha. Integrating AHP and recommendation based on customer lifetime value[J]. *Information & Management*,2005,314(42):387–400.
- [4] Ching-Hsue Cheng, You-Shyang Chen. Classifying the segmentation of customer value via RFM model and RS theory[J]. *Expert Systems with Applications*,2009,36(3).
- [5] Yang Wenjie. A customer segmentation model research based on customer life time value: a case study on the wireless telecom industry[D]. Shanghai Jiao Tong University,2008.
- [6] Tian-Shyug Leea, Chih-Chou Chiub, et al., Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics and Data analysis*[J],2006,50: 1113–1130 .
- [7] John T. Bowen, Shiang-Lin Chen. The relationship between customer loyalty and customer satisfaction[J]. *International Journal of Contemporary Hospitality Management*,2001,13 (5),213–217.