Case analysis of data mining technology in customer relationship management

Huimin Xing

Inner Mongolia Normal University, Inner Mongolia, China

Abstract. With the development of digital economy, customer relationship management has been paid more and more attention in life. Many enterprises not only face changeable market demand, but also face more fierce competition between peers, so the survival and development of enterprises need more stable customer groups. As customers will produce a large amount of valuable data in the management process, which can be used to tap the potential value of customers. It may bring more business opportunities, and to deepen the use of information can also detect the potential needs of many customers. Based on the data of customer transaction behavior, this study uses the gradient improvement algorithm to mine the customer loyalty rules and improve the customer management level.

Keywords: Data mining; Customer management; Loyalty.

1. Case analysis and introduction of data mining technology in customer relationship management

Customer relationship management is a common business strategy for the purpose of meeting the broad needs of customers, by communication and using correspond information technology have a deeper understanding of customer needs, and the enterprises can adjust products, services and sales strategy according to the needs of customers. In order to improve the market competitiveness and expand customer groups, enterprises may implement a series of coordination, management and sales which are based on customer information to provide customer-centered convenience and quality service so as to help enterprises achieve the rising goal of market share. Therefore, customer relationship management is one of many enterprise management methods, which has great significance to enterprise marketing management.

Data mining, also known as Knowledge Discover in Database (KDD), is a hot spot in the field of artificial intelligence and database research. The so-called data mining refers to the process of revealing hidden, unknown and potentially valuable information from a large amount of data in the database. Data mining is a kind of decision support process. It is mainly based on artificial intelligence, machine learning, pattern recognition, statistics, database, visualization technology, and etc.It can automated analyse enterprise data, make inductive reasoning, and dig out the potential model from useful data.It is a highly effective approach to help decision makers to adjust the market strategy, reduce risk and make the right decisions.

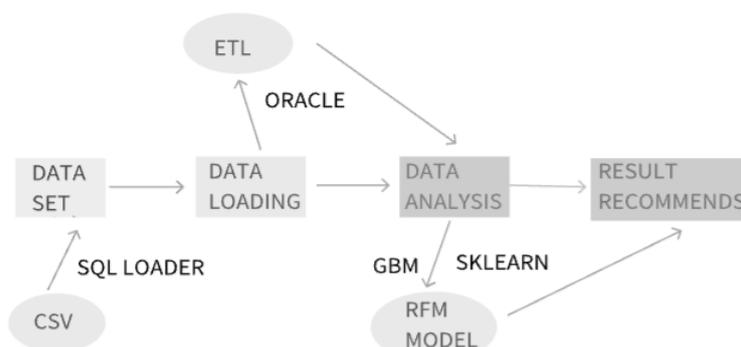## 2. Implementation mode and introduction of data mining technology

Figure 1. Process summary diagram

By studying the user consumption behavior data of a payment company, the regularity characteristic of customer loyalty is found and provide a basis for personalized services. The main implementation process is divided into four steps: (1) load the file data into the database, (2) process the data, (3) analysis data, and (4) conduct the analysis of result and feedback on data.

2.1 Implementation methods and tools of data mining

Data uses the SQL * LOADER to import the data into the ORACLE database by csv form. Mainly writing SQL statements process data by Oracle for preparation of data analysis, and using Scikitlearn to analyse data.

2.1.1 Oracle

The ORACLE database system is characterized by fine portability, which also have the feature of convenient use and strong function. It is suitable for various large, medium and small microcomputer environments. Therefore, it is a widely popular relational database. It is a high-efficiency, reliable and high-throughput database scheme.

SQL * LOADER is a data loading tool for ORACLE, commonly used to migrate operating system files to a ORACLE database.

2.1.2 Scikitlearn

Scikit-learn (also known as sklearn) is a freeware machine learning library for the Python programming language. It has a variety of classification, regression, and clustering algorithms that including support vector machines, random forests, gradient lifting, k-mean, and DBSCAN. Using sklearn requires the support of other libraries, such as Numpy and Scipy. Sklearn is more conservative compared to other open-source projects.

Using the GBM regression algorithm train the model. This algorithm is one kind of Boosting promotion algorithm, which is a machine learning method that uses a series of learners to learn. It uses certain rules to integrate various learning results to obtain better learning results than a single learner. The main idea is based on the previously established base learners of the loss function gradient descent direction to establish the next new base learners, it uses different weights to the base learners for linear combination, make the good performance learners reuse. Overall, by integrating these learners make the loss function of the model decline and the model may improve continuously.

2.1.3 RFM model

The specific analysis mainly adopts the RFM model, the RFM model is an important tool and means to measure customer value and customer profit creation ability. When enterprises or commercial tenant analyze numerous customer relationship management model models, RFM models are mainly widely used in customer measurement of life-cycle value, customer segmentation and behavior analysis. The value of the customer is described through the three indicators of the recent purchase behavior, the overall frequency of purchase and the amount of consumption.

The RFM represents Recency, Frequency and Monetary. R represents the time interval-the time interval between the last purchase time and the end time of the statistical period. The shorter the time interval, the larger the number of customer purchases during the statistical period. F represents the frequency of purchases,the higher of the F value, the higher of the customer loyalty, and the stronger the willingness to buy again. M indicates the total amount spent during the purchase, the higher the total purchase amount, the more loyal the customer.

The Customer value score is as follows:

$$score = wR \times R + wF \times F + wM \times M \tag{1}$$

(The wR, wF and wM indicate the weights of the three indicators R, F and M respectively)

3.   Specific process of data mining

Explore the user consumption behavior data, mainly to analyze customers' consumption frequency and consumption amount.

3.1 Get the data and load it into the database

Obtain the data file tables such as the historical transaction behavior data and the training data of the open source data, analyze the data business meaning through the data dictionary, analysis the effective indicators and conduct the feature extraction.

Load the training data files and the historical transaction files into the database.

3.2 ETL data handling

3.2.1 Determine the scope of data analysis

Check the time span of historical transaction behavior data. Due to the data table has a large data time span, the consumption data of the cardholder within the past 3 months is selected as the sample data.

3.2.2 Processing of the data

(1) Data processing of the historical transaction behavior table

Explore the sample situation of the historical transaction behavior data and check the data value of the database table. If the sample value exists empty, then turning the empty value to 0. After processing the sample data, exploring the consumption situation of different customers in different regions, states, merchant category groups, merchant category and merchants.

(2) Generate a new customer basic information data table

Table 1, historical transaction behavior of database, was merged with the training data table. Consolidated data showed a large amount of null data in some samples. Therefore, After judging the data of the same consumer user of sample, by adding the association behavior table 2 generate the new customer basic information data table. Through the above operation, the availability of sample is improved.

Export the new customer base information data table as a csv file.

3.2.3 Data analysis

(1) Analyze the characteristic variables of the customer basic information data sheet itself

Observing the three variables features in the basic information of the customer data table: feature_1,2,3. The distribution of the target variables is basically the same, and the mean value is close to 0. Therefore, these three features have insufficient predictive power for the target variable, other features need to be extracted.

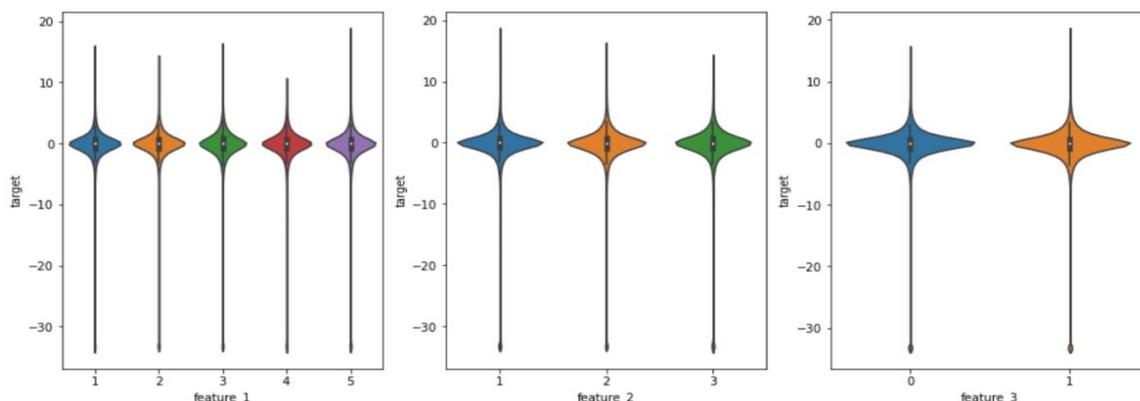The visualization method extracts the sample features.

Figure 2. The sample characteristics

(2) Analysis of other variables

In addition to the existing feature variables in the training data table, data information of the historical transaction behavior tables can be used for feature extraction. The new customer basic information data table has been processed and combined with the basic data,the following is the analysis of the sample characteristics.

Table 1. Selecting customers randomly and observing the relationship between consumption and loyalty in the past three months.

sample range	Average consumption amount	Average loyalty score
Random selected, 70,000 customers	-1.35	-0.36
Random selected, 70,000 customers	-1.67	-0.39
Random selected, 70,000 customers	-2.88	-0.42

Table 2. Selecting customers randomly, and observing the relationship between consumer merchant type, merchant type group, merchant number, consumption area, and customer approval and loyalty in the past three months.

sample range	Average Number of Merchant Type Groups	Average Merchant Type Number	Average number of merchants	Average number of cities	The average number of states	Whether to approve	Average loyalty score
Random selected, 70,000 customers	2.5180	1.9755	3.48	0.94	0.577	0.0096	-0.36
Random selected, 70,000 customers	2.5182	1.9789	3.47	0.95	0.60	0.0094	-0.39
Random selected, 70,000 customers	2.5430	1.998	3.52	0.96	0.58	0.0090	-0.42

Table 3. Observing the relationship between consumption times and loyalty in the past three months.

The number of consumption	Average loyalty score
Consumption once	-0.54
Consumption 2 times	-0.66
Consumption 3 times	-0.72
Consumption 4 times	-0.79
.....	....
Consumption 999 times	0.20

According to the summary data, the data was roughly meet the RFM model conclusions and the model training was started.

3.2.4 The model was trained using the GBM algorithm regression algorithm

Reading the newly synthesized feature dataset of Scikitlearn: a new customer basic information data is prepared, the missing value of merged data is set to 0, and using train_test_split() function

divide-the proportion is 80:20 (test_size = 0.20), random_state is 42,then the model was trained by the GBM regression algorithm.

Extracting features of consumption behavior nearly three months.In the past three months, the proportion of purchase times, purchase amount and consumption business categories is relatively large, it conforms to the judgment of the cardholder loyalty from the business meaning.

3.2.5 Draw the feature importance distribution map

By plotting the distribution of feature importance, Features are much more clear.Extracting features of consumption behavior nearly three months from the historical transaction behavior data table, these features show that hist_amount_new (total of purchase amount in recent 3 months), the hist_merchant_category_id_amount (total of consumer merchant categories in recent 3 months), the hist_merchant_id_amount (total of consumer merchants in recent 3 months) and hist_subsector_id_amount (total of consumer merchant category groups in recent 3 months) have a relatively large porproration of the model.
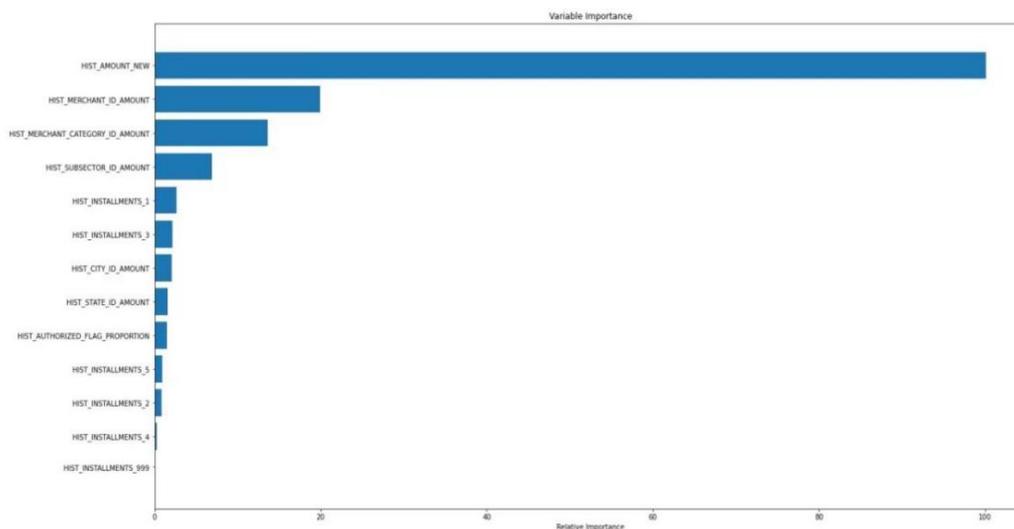

Figure 3. Importance distribution map

3.2.6 Loyalty correlation analysis

In the process of consumption, the consumption of customers is large, and the more times there are, customers may have sufficient consumption power and have a tendency to repeated purchase,so the method of increasing customer loyalty can start from the amount of customer expenditure. The expenditure determines the consumption ability of the customer, and the store can push products of goods in different amounts according to the customer's consumption ability. Another method is rising the advertising volume of merchants in the same category and increasing the discount intensity of merchants in the same category. On the other hand, for unstable consumer groups, the discounts and promotions of merchant can also be carried out to achieve the purpose of stabilizing customer and increasing customer loyalty.

4. Results feedback and suggestions

If customer relationship management may only be reflected in the management level, it can not mobilize the enthusiasm of customers and lose the meaningful of customer relationship management. Good CRM relationship can promote the management and serve in the management. Combined with this study, put forward the customer management suggestions:

1)Through the loyalty prediction model, timely identify customers with decreased loyalty, and take customer maintenance measures in advance to ensure user retention.

2)Hierarchical and personalized management are based on customer loyalty and customer consumption and habits.

3)Timely adjust the commercial operation positioning according to customers' consumption capacity and consumption areas.

References

[1] Jun Wu, Li Shi, Wen-Pin Lin, Sang-Bing Tsai, Yuanyuan Li, Liping Yang, and Guangshu Xu. "An Empirical Study on Customer Segmentation by Purchase Behaviors Using a RFM Model and K-Means Algorithm".

[2] Ma Qian. "Machine learning-based e-commerce platforms repeat purchase customer forecasts". Master of Engineering (professional degree). Lanzhou University. Issue # 03, 2018. F416.61; TP181.

[3] M. S. Chen, J.Han, and P. S. Yu ."Data mining: an overview from a database perspective," IEEE Transactions on Knowledge and Data Engineering, vol.8, no.6, pp.866–883, 1996.

[4] Alex Szalay, Ani R. Thakar, Jim Gray. "The sqlLoader Data-Loading Pipeline". Computing in Science & Engineering (IF 2.152) Pub Date: 2008-01-01 DOI: 10.1109/mcse.2008.18.

[5] Fan Ying. "Technical Analysis of Customer Forecasting Based on Python Data Mining". Information Record Material, September 2021, Volume 22, Issue 9, DOI:10.16009/j.cnki.cn13-1295/tq.2021.09.081

[6] Long bingting. "Customer Value Based RFM Model". Journal of Xichang University (Natural Science Edition),Volume 36, Issue 1, March 2022,DOI: 10.16104 /j.issn.1673-1891.2022.01.010.

[7] Xu Xiangbin, Wang Huaqiang, Tu Huan, and et al. "E-commerce customer segmentation based on improved RFM models [J]". Computer Application, 2012,32 (5): 1439-1442.

[8] Xiong LAN, Gao Bing. "Customer Segment Research Based on the RFM Multi-Level Customer Value Model [J]". Business Economy Research, 2017 (5): 55-57.

[9] Rao Shuzhen. "Analysis of several data migration methods of Oracle database", Volume C omputer Knowledge and Technology 14, Issue 21 (July 2018).