

Research on E-commerce Customer Loyalty under Big Data

Kokula Krishna Hari Kunasekaran

*Association of Scientists, Developers and Faculties (ASDF), United Kingdom
prm@kokulakrishnaharik.in*

Abstract

Based on the characteristics of the e-commerce industry, this paper conducts a comprehensive analysis of the factors affecting the e-commerce customer loyalty. Based on the classic RFM customer loyalty model, two important attributes that affect customer loyalty in e-commerce are introduced: satisfaction and attention, thereby establishing a RFMSA e-commerce customer loyalty partition model. In this model, the basis for dividing customer loyalty in e-commerce transactions is analyzed more comprehensively from multiple perspectives. Based on the establishment of a customer loyalty model for e-commerce, the customer loyalty is divided by a cluster analysis algorithm. Based on the classic clustering algorithm K-means, an improved algorithm to determine the initial clustering center by segment was proposed to divide customer loyalty. This method can effectively reduce the time to determine the initial cluster center and the problem of local optimal solution caused by iterative calculation. In order to verify the effectiveness of the algorithm proposed in this paper, the customer loyalty is divided by analyzing the transaction data of an online mall member. The experimental results show that, compared with the K-means-based clustering algorithm, the improved method proposed in this paper has higher accuracy in the classification of e-commerce customer loyalty and can better segment the e-commerce customer loyalty.

Keywords: *Customer loyalty, RFMSA, K-means*

1. Introduction

In the e-commerce environment, companies can use the Internet to provide customers with products and services. In the face of fierce competition, how to maintain market competitiveness is the most important issue for e-commerce companies. This requires improving customer satisfaction through various channels, providing customers with satisfactory services, and maintaining customer loyalty. We can achieve this goal through customer segmentation in customer relationship management. Segmenting customers according to loyalty can identify unique customer groups, then analyze the different needs of each customer group and their stakes with related companies, and finally provide different customers with satisfactory products and services on demand [1][2]. At present, most e-commerce companies have established corresponding information systems based on their own characteristics, which are used to store valuable information such as customer personal information, historical purchase records, product attention, and satisfaction of purchased products. After analysis, we finally get the information about the customer's consumer

Article history:

Received (January 27, 2020), Review Result (March 2, 2020), Accepted (April 11, 2020)

psychology, purchasing behavior and ability, and living habits that the company needs. With this information, it is convenient for the company to analyze customer loyalty [3]. In order to solve the problem of dividing customer loyalty in massive data, after a lot of research and experiments, it was found that the introduction of data mining technology can better solve this problem. Data mining is to analyze each piece of data, understand the source of the data, and obtain relevant knowledge and technology, and finally integrate and check the data. Data mining uses association analysis, cluster analysis, classification analysis and other methods to extract potentially useful information from fuzzy data. Data mining technology can perform data processing and mining on the transaction data of e-commerce enterprises, so as to distinguish the level of customer loyalty, find potential customers, and conduct effective marketing based on high-loyal customers and potential customers [1].

The theory of customer loyalty can be traced back to 1947. The “self-involvement” proposed by American scholars Sherif and Cantril is mainly used in the decision-making process of consumers. It can be divided into brand inertia and brand loyalty. Inertia refers to the passive and passive search for information by low-stakeholders, which is susceptible to various factors to change their attitudes. Brand loyalty refers to the active and positive attitude of high-stakeholders in searching for product-related information, and considers and compares various Differences between brands to make the right decision [4]. When segmenting customers according to the value of the customer to the enterprise, a reasonable division index and division method need to be selected. Among the many analysis models, RFM models are the most common. The three indicators of the RFM model are the recent purchase time (Recency, R), purchase frequency (Frequency, F), and purchase amount (Monetary, M). In recent years, there have been studies on customer loyalty in various industries, and there are currently a large number of research results. For example, Carvajal [5] proposed a bank through the study of banking customer image, service quality, and customer satisfaction and loyalty. Structural equation model of industry development; Tung Feng-Cheng [6] studies the customer satisfaction model of China’s mobile service industry by analyzing the influencing factors such as the perceived value of customer satisfaction. With continuous breakthroughs and innovations in customer relationship management, both at home and abroad, clustering algorithms have been widely used in customer segmentation research. The K-means algorithm that is concise, efficient, and easy to implement is widely used in cluster analysis. However, the effect of the K-means algorithm is greatly affected by factors such as the number of clusters and the initial cluster center [7]. Rokham [8] proposed a cluster analysis of network users, established a group of users with similar browsing patterns, provided useful knowledge, and personalized network services; Morteza pour [9] proposed an artificial neural network, genetically based Customer Credit Evaluation Based on Algorithm and Bayesian Probability Clustering. This article will evaluate customer credit and obtain a proper result by combining classification using artificial neural network, genetic algorithm and Bayesian probability method.

This article will analyze the classic RFM model, increase customer satisfaction and attention, and build a new model. Based on the research of partitioning clustering method-K-means algorithm, and analyze its advantages and disadvantages, the K-means algorithm is optimized by using the segmentation idea.

2. Research on e-commerce customer loyalty model

Customer loyalty refers to a behavior in which customers have a preference for a particular brand or enterprise product and service, that is, customers who are not willing to change their

consumption habits for products of the same price and similar functional quality. Customers and enterprise voluntarily establish a relationship of mutual trust and dependence. Enterprise customers are mainly divided into two groups: old customers and new customers. Loyal customers are the key to a company's survival and the main source of profit. For this reason, many companies use defensive marketing strategies to maximize customer retention to increase their market share and profitability. Although traditionally, more energy is used by companies to implement offensive marketing strategies to develop new customers, but this does not affect the retention of old customers. Studies have shown that only by retaining the resources of old customers of an enterprise can it brings more benefits to the enterprise.

In the e-commerce industry, the most commonly used method for analyzing customer loyalty is the RFM model. RFM data is the real transaction data of customers, so the data is also more accurate. Generally, it does not involve the customer's personal privacy and is easy to obtain. The RFM model is a model widely used in customer loyalty research. RFM is based on three important customer behavior indicators. These three variables are the time interval, frequency and amount of customer purchases. The detailed definition of the RFM model is as follows:

(1) Time (R, Recency) from the last purchase. That is, within the observation window, the time difference between the last time that the user made the purchase behavior and the observation point. In most cases, the observation point is the end of the time window. Therefore, the smaller the time difference, the greater the probability that the user may continue to make purchases.

(2) Purchase frequency (F, Frequency), that is, the number of times customers make purchases within the observation window. Therefore, the greater the purchase frequency, the stronger the user's dependence on the mall and the higher the user loyalty. Therefore, the frequency of purchases is directly proportional to customer loyalty.

(3) Cumulative consumption amount (M, Monetary), that is, the total amount of consumption by the user within the observation window. Cumulative consumption amount is another attribute that affects user loyalty indicators in addition to the latest purchase time and purchase frequency. The user may not have recently purchased, and the frequency of purchase is not high, but the amount of consumption that occurs is large, and enterprise decision makers also need to analyze this type of customer and believe that user loyalty is high. Using the RFM attribute to analyze the user's loyalty, it has been proved to be a practical and effective analysis model after the test of time.

The disadvantage of the RFM model is that there are too many customer groups after classification, and there is a collinearity defect between the purchase frequency and the purchase amount. Customers in the e-commerce industry have greater randomness in purchasing products, and customers can more easily obtain other customers' satisfaction evaluations of a certain product. Therefore, the method of analyzing customer loyalty using only three elements of RFM has greater limitations. In the process of customers purchasing products, customer satisfaction or product evaluation is also an important indicator of customer loyalty. In addition, some customers have never purchased a certain product, but have been paying attention to the product through collections, adding shopping carts and other e-commerce technology links. For this type of customers, we cannot obtain their RFM data, but because they have the intention to purchase the product, this design proposes a loyalty analysis model based on RFMSA. At the most recent purchase time R, purchase frequency F, and purchase amount M. Based on this, Satisfaction (S) and Attention (A) are added as standards to measure customer loyalty.

The RFMSA model, as an important indicator of customer behavior in e-commerce, is defined in detail as follows:

(1) R (Recency), the last transaction time, the number of days since the user spent the last time in the mall at the observation time point, the smaller the R, the smaller the time interval, the more likely the customer is to buy again.

(2) F (Frequency), the frequency of transactions, the number of times a user has traded in the observation time interval, that is, the number of successful transactions in the mall.

(3) M (Monetary), the transaction amount, the total amount of the user's transaction within the observation time interval. The larger the M value, the higher the user's consumption in the mall, so the more likely the user will buy again.

(4) S (Satisfaction), product evaluation. Users who have successfully traded orders within the observation time period have evaluation data for each product. The product evaluation index indicates the average value of the evaluation of all purchased products in this interval. The higher the evaluation index, the higher the user's recognition of the products in the mall, so the higher the possibility of buying again.

(5) A (Attention), Attention, the degree of user's attention to the products in the mall within the observation time interval, including information such as the number of views. The higher the user's attention, the higher the likelihood that the user will purchase the product.

In the RFMSA model, the importance of each attribute dimension is slightly different, that is, the weights of the five attribute dimensions are different. The five types of indicators have different impacts on user loyalty, and the RFMSA model is defined according to the Customer Lift-time Value (CLV) customer lifetime value indicator that comprehensively reflects customer loyalty. As shown in Equation 1:

$$CLV_i = -\omega_R R_i + \omega_F F_i + \omega_M M_i + \omega_S S_i + \omega_A A_i \quad (1)$$

Among them, i represents the i customer among n customers, CLV_i represents the customer's loyalty index R_i , F_i , M_i , S_i , A_i respectively represent the customer's RFMSA indicator, and ω_R , ω_F , ω_M , ω_S , ω_A respectively represent the weight coefficients of the five indicators. Among the various indicators, the latest purchase time is inversely proportional to CLV , and the other four indicators are directly proportional to CLV . For the establishment of the weighting coefficient, the existing analytic hierarchy process is used to analyze, establish a judgment matrix of five indicators, compare these indicators in pairs and obtain the relative importance to generate a $5 * 5$ judgment matrix, and find the judgment matrix. Finally, the final weight value is determined by checking the consistency of the judgment matrix.

3. Research on cluster analysis methods

3.1. Data Mining Technology

Data mining can also be called data analysis. It is to extract patterns and knowledge from a large amount of data without extracting the data itself. Now it seems that it is also a buzzword. It is generally applicable to any form of large-scale data or information processing including collection. Extraction, warehousing, analysis and statistics, and any application computer decision support system.

Data mining is an interdisciplinary branch of computer science. It is a cross-cutting approach involving artificial intelligence, machine learning, statistics, and database systems in the computational processing of large data sets. The overall goal of data mining is to extract information from a set of data and transform it into an understandable structure for

continued use. In addition to the original analysis steps, it also involves database and data management aspects, data pre-processing, model and reasoning thinking, interest measure, complexity consideration, discovery structure, visual post-processing, and online update. Data mining is the analysis step of knowledge discovery in the database.

Data warehouse is an evolving technology that is part of an organization's long-term strategic investment. In this role, it is supporting the organization's customer relationship management and business intelligence strategies. Data warehouses are huge, time-consuming development and implementation projects. Data mining is an emerging interdisciplinary subject that integrates the latest advances in artificial intelligence, mathematical statistics, databases, and many other fields such as finance, insurance, and telecommunications. It has broad prospects for development. Defining data mining from a business perspective is a new technology for processing business information. It extracts, transforms, analyzes, and converts other models to process business data in large commercial databases and extracts key data that supports business decisions. Data mining technology can be applied in the decision analysis and management stage of customer-centric enterprises. In customer relationship management, data mining is mainly applied to the division of customer groups, data mining to maintain customers, cross marketing and analysis of corporate social responsibility.

In the e-commerce environment, customer relationship management is also the focus of e-commerce companies, because customer relationship management plays a vital role in the development of the enterprise. A large amount of data is generated in the process of e-commerce transactions. Data mining is to obtain meaningful information files and customer transaction data from the server, and analyze these data effectively to provide personalized information services and targeted e-commerce companies. Favorable decision-making support activities enable enterprises to achieve better customer retention and customer satisfaction, and improve their competitiveness. Data mining technology can provide corporate decision support (CSR) in areas such as collection, customer and consumer satisfaction research analysis and maintenance, and improve market competitiveness. The data generated in e-commerce has its own characteristics. Combining data mining technology with e-commerce and choosing a data mining method suitable for e-commerce can improve the efficiency of data mining and make data mining better serve e-commerce.

3.2. Cluster analysis algorithm

Commonly used data mining techniques include association analysis, cluster analysis, principal component analysis, classification analysis, web page mining, time series analysis, and so on. In fact, when using data mining technology to solve real-world problems, two or more technologies are often used. According to different practical requirements, the appropriate data mining method is selected. For example, cluster analysis algorithms are mainly used for classification, and data is divided into different categories according to requirements. It is a commonly used mathematical statistical method. It is generally used to classify the data to be analyzed for similarity, and then combine the knowledge in different fields for analysis to provide a basis for decision makers to make decisions.

Due to the overlapping categories of clustering methods, it is difficult to propose a clear classification. At present, distance-based clustering analysis is the main research direction of clustering algorithms. The main clustering algorithms can be divided into 5 categories: Partitioning Method; Hierarchical Method; Density-Based Method; Grid-Based Method);

Model-Based Method. The main application of this paper is the K-means algorithm in the partition-based method.

3.3. K-means algorithm

K-means algorithm is the most commonly used clustering algorithm. K in the name refers to the number of classes that the algorithm needs to find. The user needs to set the value of K in advance. For the convenience of explanation, this paper uses a two-dimensional graph with two-dimensional variables to visually show the algorithm. It is worth noting that in general, the input variables are mostly more than two-dimensional. However, the input variable is limited to two dimensions, and discrete points in two dimensions can be used to simply and vividly illustrate the clustering process. This process is also the same for high-dimensional variable data.

Each record is treated as a discrete point, which also means that the input variables must be numeric. The distribution of data on the scatter plot is like a cloud. The goal of the clustering algorithm is to find K points as cluster centers. The center of the class defines the class, and each record is assigned to the class where the cluster center closest to it is located. Obviously, good cluster centers should be in the densest areas of the data cloud. The optimal allocation method of the clustering center can be defined as: the distance (or the square of the distance) and the smallest of all data points to their nearest clustering center. Finding these optimal solutions is difficult, and the K-means algorithm does not intend to do so. Instead, it randomly selects some points as the cluster center points at the beginning, and then adopts a series of steps to improve it.

After selecting the initial clustering center point, the algorithm performs two steps alternately: assignment and update. Three points were randomly selected as the center of the class. The next step is to assign each record to the class where the cluster center closest to it is formed to form the initial class, as shown in [Figure 1]. The figure also shows the lines that separate the classes. The dividing line of the classes is “Y”. The entire data set is divided into three categories: the left one, the right one, and the upper one. The dividing line between classes is useful for showing the process geometrically. But in practice, the algorithm only needs to calculate the distance from each record to each cluster center, and then assign the record to the class where the cluster center closest to it is located. The steps of replication do not need to determine such cluster boundaries.

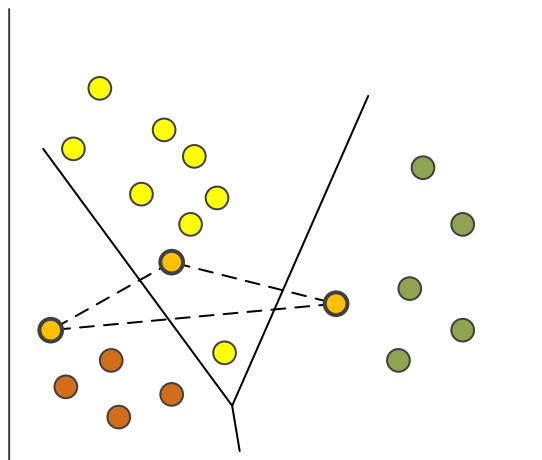


Figure 1. The initial class is formed by assigning each data point to the nearest cluster center

In the update step, the center of each class is calculated. The center is the average of all class members in each dimension. In the assignment step, each data point is assigned to the class to which the nearest cluster center belongs. In this example, the data points in the box are assigned to a new class. When the assignment step changes the class to which the data point belongs, it will cause the update step to be performed again in the next step. The algorithm alternates assignments and updates until no new assignments occur. Although an ill-conditioned data set that requires exponential processing time may be designed, the algorithm converges quickly in practical applications. [Figure 2] shows the final class and the calculated new cluster center after the data points in the box are assigned.

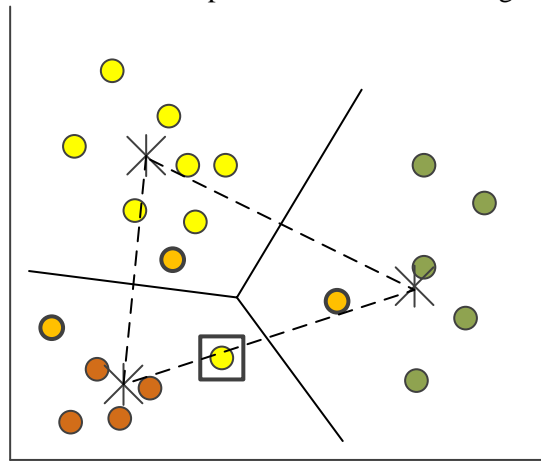


Figure 2. The K-means algorithm terminates when no records are reassigned after the cluster center has been reset recently

4. Cluster analysis of e-commerce customer loyalty

The RFMSA model is an important indicator of customer behavior in e-commerce. In the RFMSA model, the importance of each attribute dimension is slightly different, that is, the weights of the five attribute dimensions are different. The five types of indicators have different impacts on user loyalty, and the RFMSA model is defined according to the Customer Lift-time Value (CLV) customer lifetime value indicator that comprehensively reflects customer loyalty. As shown in Equation 2:

$$CLV_i = -\omega_R R_i + \omega_F F_i + \omega_M M_i + \omega_S S_i + \omega_A A_i \quad (2)$$

Among them, i represents the i customer among n customers, CLV_i represents the customer's loyalty index R_i , F_i , M_i , S_i , A_i respectively represent the customer's RFMSA indicator, and ω_R , ω_F , ω_M , ω_S , ω_A respectively represent the weight coefficients of the five indicators. Among the various indicators, the latest purchase time is inversely proportional to CLV, and the other four indicators are directly proportional to CLV. For the establishment of the weighting coefficient, the existing analytic hierarchy process is used to analyze, establish a judgment matrix of five indicators, compare these indicators in pairs and obtain the relative importance to generate a $5 * 5$ judgment matrix, and find the judgment matrix. Finally, the final weight value is determined by checking the consistency of the judgment matrix.

After the RFMSA model is established, the data needs to be pre-processed, converted into RFMSA data required by the system, and standardized. Then perform data analysis. In the process of data analysis, the system uses cluster analysis to complete the classification of the data set. After determining the number of initial classifications, the initial clustering center is determined using the segmentation method, and finally the clustering output is completed. The overall structure of the system is shown in [Figure 3].

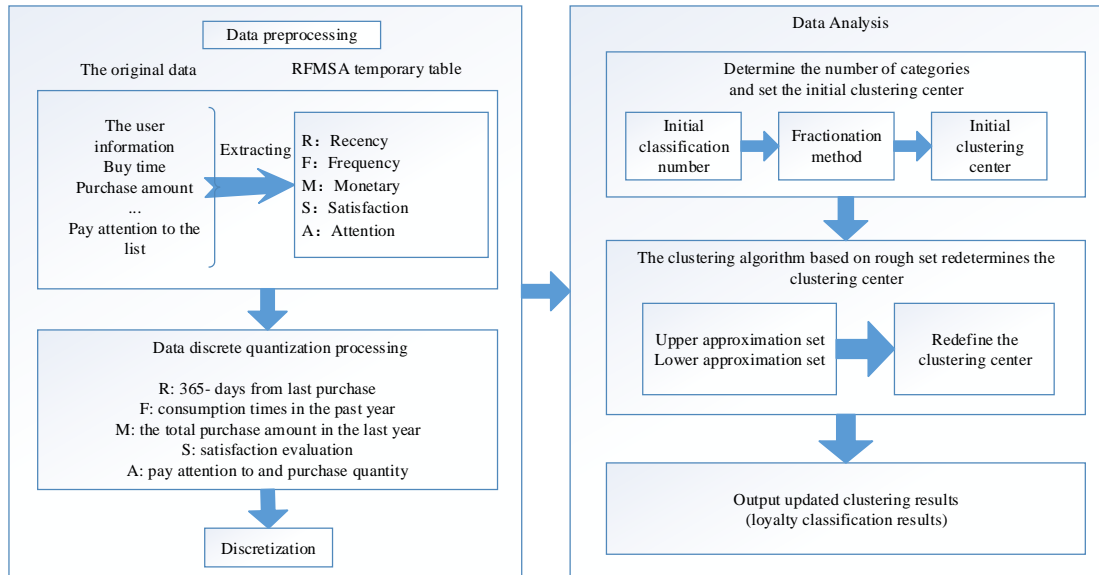


Figure 3. Overall system structure

4.1. Data organization and storage

The organization and storage structure of the data is shown in [Figure 4]. In user loyalty analysis, the data source is the transaction records of the online mall. In order to efficiently query and analyze a large number of transaction records, the system needs the data. It is organized and stored in the data warehouse in order to perform subsequent cluster analysis on the data more efficiently.

First, different types of data sources such as customer data, product data, transaction data, evaluation data, and product attention data are processed, and the data is imported into the data warehouse through steps such as integration, extraction, and loading. Use data warehouse management system to organize and store massive data. The data source is the annual sales data of a large online supermarket. The data source includes user information, product information, details of sales products, details of sales receipts, user browsing records, user evaluation records and other information. The user table has a total of 1082741 pieces of data.

The RFMSA model includes five key fields: the latest purchase time (R), purchase frequency (F), purchase amount (M), satisfaction (S), and attention (A). In addition, observation points and observation intervals need to be defined. Where the observation interval represents the time period for obtaining RFMSA data, also called the observation window, and the observation point is the end point of the observation window. If you can choose February 29, 2018 as the observation point, from March 1, 2017 to 2018 The observation window is February 29, 2014. In this way, one year of data is analyzed.

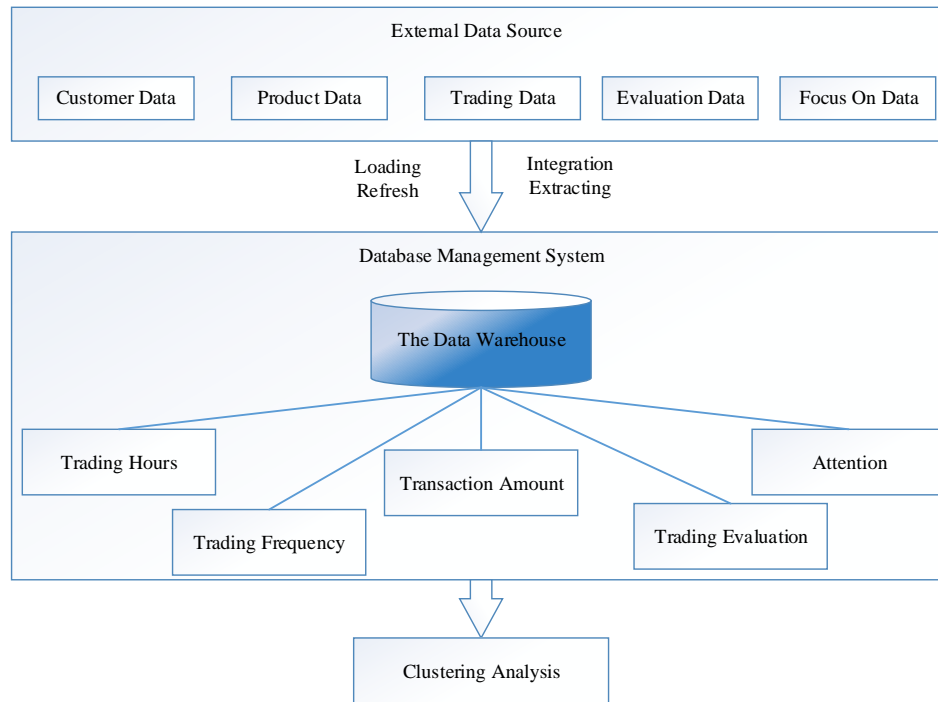


Figure 4. Organization and storage of data sources

In the specific input process of the key fields, you first need to determine the corresponding RFMSA attribute value, where:

R: the difference between the observation point and the user's latest purchase time;

F: number of user consumptions in the observation time interval;

M: the total amount of user consumption in the observation interval;

S: the average value of the user's evaluation of the purchased goods in the observation interval;

A: The time that users browse the website during the observation period.

Then through data preprocessing, the required data is organized and stored in the database, where the primary key is the username field.

First, obtain the initial RFMSA data. Relevant fields in the customer information table, product information table, transaction product information table, transaction ticket information table, transaction evaluation information table, and website browsing log table are extracted, and the four attributes in the model are calculated using various analysis functions of the database. Among them, the R attribute is the latest purchase time of the user, and the database can find the latest record of the consumption time in the observation window. The F attribute is the frequency of user consumption in the observation interval. This attribute calculates the number of user transactions in the observation interval according to the transaction ticket information table, and normalizes to obtain the user transaction frequency. The M attribute is the total amount consumed by the user in the observation interval. According to the transaction ticket information table, the transaction amount of each transaction of the user in the observation interval is cumulatively calculated to obtain the M attribute. S attribute is the user's satisfaction degree. This attribute uses the evaluation table and the transaction product information table to obtain information. The user's evaluation

information is divided into five levels in the data source. Each product of the user's each transaction has one evaluation information. Evaluation information including evaluation star rating and evaluation content, based on the operability, the evaluation star rating is selected as the evaluation standard of user satisfaction. The calculation method of the S attribute is the average of the star ratings of all the products purchased by the user in the observation interval. In the actual transaction process, some users fail to evaluate within the specified time. In this way, the website will automatically evaluate the star rating, and the default is the highest five stars. The attribute is the user's attention. This attribute uses the user's browsing log table to extract data.

Secondly, the five attribute values of RFMSA are standardized and standardized. Because the five attributes of R's most recent transaction time, F's transaction frequency, M's transaction amount, S's evaluation, and A's attention have different expressions, these different expressions need to be converted into standard data formats, which can also be removed through conversion. Some noise points. The RFMSA attribute normalization formula for member i is shown in Equation 3:

$$\begin{cases} R'_i = (R_i - \bar{R}) / \sigma_R \\ F'_i = (F_i - \bar{F}) / \sigma_F \\ M'_i = (M_i - \bar{M}) / \sigma_M \\ S'_i = (S_i - \bar{S}) / \sigma_S \\ A'_i = (A_i - \bar{A}) / \sigma_A \end{cases} \quad (3)$$

In the formula, the five attributes such as R_i are values in a unified format, R represents the average R attribute of all users, and σ_R represents the standard deviation of the R attribute.

K-mean clustering algorithm clustering is an important method in data mining algorithms. The basic idea is to divide the data set into multiple data subsets. Each data subset has the same or similar attribute information. A subset is also called a class or cluster in a cluster. If the two data points are in the same class, the two points must have the same or similar attributes; if the two data points are in different classes, the attributes of the two points must be significantly different. In the process of clustering, the attributes of objects are used to distinguish objects belonging to different classes. The K-Means cluster analysis algorithm is a typical algorithm in clustering data. K-means algorithm completes clustering based on Euclidean distance. The degree of similarity between objects is measured by Euclidean distance. K-means algorithm is currently a more effective clustering method. During the clustering analysis of the RFMSA model, the K-means clustering algorithm can effectively quantify the five attributes of the user and complete the clustering.

The basic process of K-means algorithm is shown in [Figure 5].

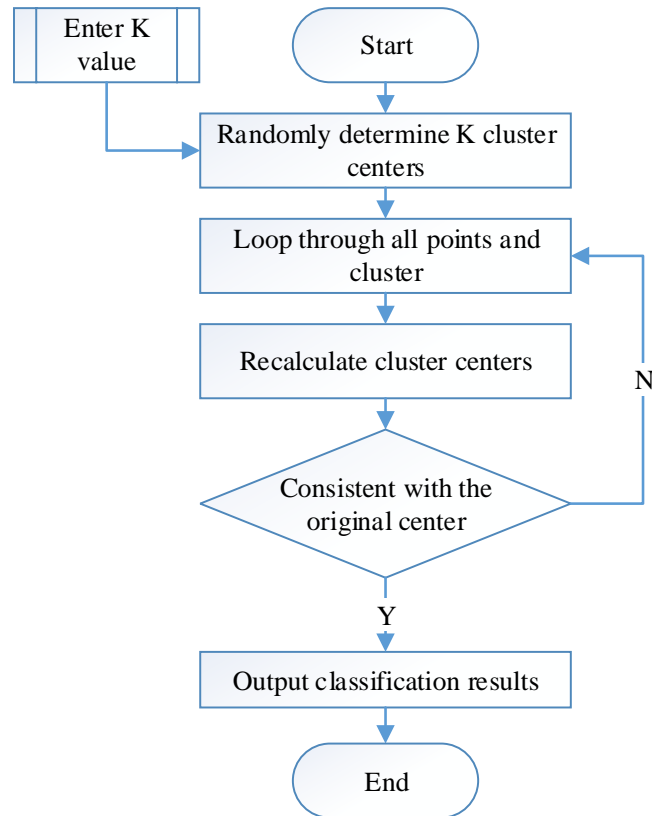


Figure 5. K-Means algorithm flowchart

4.2. Implementation of customer loyalty analysis

The process of customer loyalty analysis is shown in [Figure 6]. The real transaction data is used as the data source. The data is converted into a unified data format and stored in the database through operations such as data preprocessing, cleaning, and normalization. Cluster analysis is performed on the subsequent data. The cluster analysis includes determining the number of classifications, determining the initial cluster center, and using a rough set-based cluster center calculation method to iterate. The cluster results are output when the cluster center is no longer changed to complete the customer. Classification; complete the analysis of customer loyalty after generating clustering results.

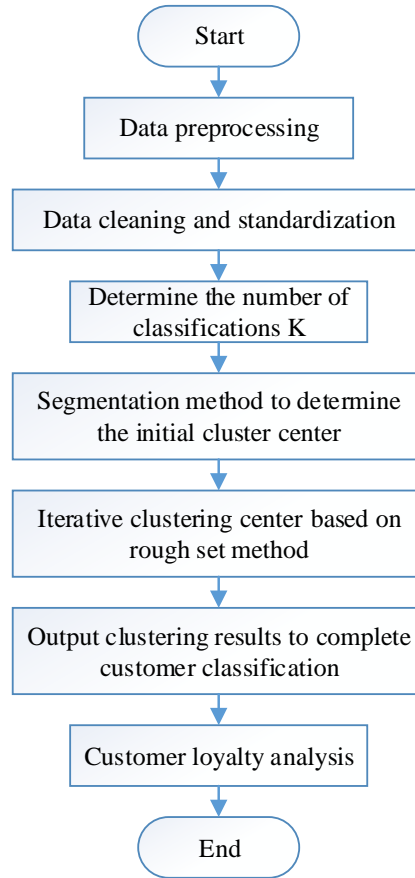


Figure 6. Customer Loyalty Analysis Implementation Process

4.2.1. Implementation of the algorithm

In the process of categorizing customer loyalty, customers are mainly divided into three categories, that is, high, medium, and low customer loyalty. The advantage of this classification method is that it can form an effective number of classifications, and the algorithm is completed in less than 15 minutes, and a balance is achieved between effective classification and algorithm efficiency. Therefore, the number of clusters is set to 3 during the implementation of the algorithm, that is, the loyalty is divided into 3 levels. The algorithm is implemented in C++. It reads the RFMSA data in the Oracle database and uses the segmentation method to determine the initial clustering center. The clustering center is determined as shown in [Table 1].

Table 1. Initial cluster centers generated after determining cluster centers by segments

	R	F	M	S	A
center 1	8.01	7.29	-0.51	-2.53	-1.91
center 2	3.65	23.89	21.09	-0.84	4.29
center 3	-0.72	40.49	169.49	0.85	10.49

The clustering algorithm formula based on rough set is used to iterate repeatedly until the conditional iteration is reached. The final clustering center is shown in [Table 2].

Table 2. The final cluster center after iteration

	R	F	M	S	A
center 1	4.56	-1.01	-4.98	2.85	-3.01
center 2	2.35	0.79	-0.03	-1.29	2.81
center 3	2.72	2.02	34.73	2.12	4.63

4.2.2. Customer loyalty analysis

According to the clustering results obtained by the clustering algorithm, customers are divided into three categories of high, medium and low according to loyalty, which are grades 1, 2, and 3, respectively. The customer loyalty index in the sample was matched by the customer loyalty index CLV. According to the formula.

$$CLV_K = -\omega_R \bar{R}_K + \omega_F \bar{F}_K + \omega_M \bar{M}_K + \omega_S \bar{S}_K + \omega_A \bar{A}_K \quad (4)$$

Where K represents the clustering category, which are levels 1, 2, and 3 respectively. $\omega_R, \omega_F, \omega_M, \omega_S, \omega_A$ respectively represent weight values representing the five attributes of RFMSA. R_K, F_K, M_K, S_K, A_K represents the normalized average of the attributes in the result set. The CLV_i corresponding to the three clustering result sets 1, 2, and 3 represents the i customer among the n customers. The CLV_i represents the customer's loyalty index R_i, F_i, M_i, S_i, A_i respectively, and the RFMSA indicators of the customer $\omega_R, \omega_F, \omega_M, \omega_S, \omega_A$ respectively represent the weight coefficients of the five indicators. The values are -2.51, 0.69, and 19.15.

The higher the CLV, the higher the customer loyalty. Level 3 customers are high-loyal customers, level 2 customers are medium-loyal customers, and level 1 customers are classified as low-loyal customers. The number of high-loyalty customers was the smallest, at 1,720; the number of medium-loyalty customers was 8,477; and the number of low-loyalty customers was 14,399. In the result set of level 3, the average recent transaction time is 4.27 days, the average transaction frequency is 31.37 times / month, the average cumulative purchase amount per customer is 2619.78 yuan, the average product evaluation is 4.65, and the average degree of attention (number of products viewed) is 129 times. Compared with the other two types of customers, the transaction interval of this type of customers is relatively short, and the indicators such as transaction frequency, accumulated amount, evaluation, and attention are significantly higher than other customers. High-loyalty customers spend a lot of money, which is of great value to the mall, so they need to be given corresponding attention. Among the medium-loyalty customers, the average recent transaction time is 8.18 days, the average transaction frequency is 11.04 times / month, and the average cumulative amount is 1036.22 yuan, the average product evaluation was 4.66, and the average attention was 87.23 times. This type of customer belongs to the level of ordinary consumption, but has a large potential value. The proportion of low-loyalty customers is the largest, the average recent transaction time is 23.15 days, the average transaction frequency is 1.37 times / month, the average cumulative amount is 179.25 yuan, the average The evaluation index is 4.24, and the average degree of attention is 42.28 times.

This type of customers pay less attention to the mall and spend less frequently, etc. In addition, it can be seen that the average evaluation value of this type of customers is relatively low, and this type of customers has a low contribution to the turnover and does not have great value. However, by collecting information about the needs and experiences of such customers,

and improving customer satisfaction, some customers can be converted to medium-loyal customers, so as to continue to improve efficiency.

5. Conclusion

In recent years, with the continuous development of e-commerce, competition between enterprises has become increasingly fierce, and customer loyalty to the enterprise has become increasingly important. E-commerce enterprises can analyze the massive e-commerce customer data to divide customer loyalty, and thus adopt different marketing methods for customers with different loyalty. Maximize profits by maintaining repeat purchase rates for high-loyal customers. For potential customers with low loyalty, implement targeted marketing methods, increase their attention to the website and satisfaction with shopping, and develop them into high-loyalty customers.

This article conducts loyalty analysis on the transaction data of members of an online mall, taking customer information, product information, sales details, user attention and user evaluation as research objects, combined with an improved RFM customer loyalty model, and using K-means based on rough sets. The cluster analysis algorithm analyzes the customers of e-commerce websites to achieve the division of customer loyalty.

The research contents and results of this article mainly include the following points:

Research the theory of customer loyalty. By analyzing the current status of domestic and foreign research on customer loyalty, it is found that companies are paying more and more attention to the customer's personalized service to continuously improve customer satisfaction, make them loyal users, and maximize corporate profits. With the more mature and development of data mining technology, aiming at the huge amount of data in e-commerce transactions, this article studies the basic method of analyzing customer loyalty using data mining technology.

Put forward RFMSA customer loyalty model. According to the characteristics of the e-commerce industry, this article analyzes the purchasing behavior of e-commerce users. Through analysis, it is found that based on the traditional customer loyalty model, customer satisfaction and attention in e-commerce transactions are also important factors affecting customer loyalty. Therefore, based on the traditional RFM model, this paper proposes an RFMSA customer loyalty model for e-commerce customers.

Improvement of K-means clustering algorithm. The K-means algorithm has a good effect when processing large amounts of data, so it is suitable for the analysis of customer loyalty in e-commerce. However, when determining the initial clustering center, due to the large amount of data, it takes a long time to determine the initial clustering center, and it is easy to fall into a local optimal solution. Avoid the above problems. With the continuous development of e-commerce, the establishment of a perfect e-commerce customer loyalty model requires multiple theoretical support. This article has only done basic research, and there are still some limitations. Further research can be done in the next research. The basis for establishing customer loyalty model in e-commerce transactions.

References

- [1] Liu Qingwei, "Research on the application and implementation strategy of precision marketing system in mobile value-added services," Master's Thesis of Beijing University of Posts and Telecommunications, pp.8-9, (2013)
- [2] Du Le., "Research on customer classification of B2C E-commerce enterprises," Master thesis of North China University of Technology, pp.8-9, (2014)

- [3] Field. "Data analysis of online shopping market questionnaire based on data mining technology," Jilin University Master Thesis, pp.10-12, **(2014)**
- [4] Sherif and Cantril, "The psychology of ego ~ involvements," Social Attitudes and Identifications (Book Review) Schermerhorn, R A. Social Forces 26, vol.1, pp.362, **(1947)**
- [5] Carvajal, Sara Arancibia, and Ruzzi, "Focusing resources for customer loyalty: An application to the Chilean banking industry," African Journal of Business Management, vol.1, pp.1100-1108, **(2012)**
- [6] Tung and Feng Cheng, "Customer satisfaction, received value and customer loyalty: the mobile services industry in China," African Journal of Business Management, vol.5, pp.1730-1737, **(2013)**
- [7] Wang Zhong, Liu Guiquan, and Chen Enhong, "A K ~ mkuoeans algorithm for optimizing the initial center point," Pattern Recognition and Artificial Intelligence, vol.22, no.2, pp.300-304, **(2009)**
- [8] Rokham Hooman and Falakshahi Hale, "Web users clustering analysis," International Journal of Computer Science and Information Security, vol.9, pp.49-52, **(2014)**
- [9] Morteza pour Reza, and Afzali Mehdi, "Assessment of customer credit through combined clustering of artificial neural network, genetics algorithm and Bayesian probabilities," International Journal of Computer Science and Information Security, **vol.12**, pp.53-57, **(2013)**

This page is empty by intention.