Research on Precision Marketing Strategy Based on Clustering Algorithm

Zhengmin Fu

767483891@qq.com

University of Chinese Academy of Social Sciences, Beijing, China

Abstract: With the development of personalized services, the use of big data technology to guide precision marketing has become a trend in the future development of e-commerce. However, data mining algorithms such as clustering algorithms to parse precision marketing patterns have not been widely studied. In this paper, a python crawler and a relevant public dataset were used to collect 25,000 data from a shoe store in Taobao, and the data composition came from two aspects: user attribute data and transaction data. The data were pre-processed using SPSS software, and the RFM model was established and standardized, then the relevant value weight coefficients were derived using Matlab software, and the data were analyzed by clustering algorithm using SPSS, and the total customer value was used to verify the clustering results. Combining the results of text analysis and clustering algorithm analysis, we finally propose a precise marketing strategy.

Keywords: Clustering Algorithm, RFM Model, Total Customer Value, Precision Marketing.

1 INTRODUCTION

From the previous literature, it can be seen that the concept of precision marketing and the concept of data mining appeared relatively early, and most scholars have a wide range of research on the impact of data mining on precision marketing, but there is still relatively little research on the application of data mining technology to e-commerce precision marketing, and most scholars have only studied what kind of impact precision marketing in the context of big data can bring to the development of e-commerce, but for how to Most scholars have only studied the impact of precision marketing on the development of e-commerce in the context of big data. [1]. The rapid popularity of the Internet has brought new opportunities for the development of e-commerce industry, and the rapid rise of big data in recent years has also provided new impetus for the development of e-commerce industry, and it can be seen from the development in recent years that big data technology has played a huge role in many fields. How to take advantage of the big data background in the field of e-commerce, through data mining technology to help the e-commerce industry to better identify target customers and achieve accurate marketing is the main topic of this paper focus on research. With the development of society, the traditional seller's market has been gradually replaced by the buyer's market, and consumers are more inclined to personalization when purchasing products. This demand for personalization now extends to the marketing field, where consumers want to get the marketing ads they want to see. The e-commerce industry uses data mining techniques to analyze user

BDEDM 2023, January 06-08, Changsha, People's Republic of China

DOI 10.4108/eai.6-1-2023.2330341

behavior data, and this information will help e-commerce platforms to implement more accurate marketing and advertising strategies. [2] This paper provides practical reference value for e-commerce companies to use big data technology to analyze the precise marketing model through case study analysis. [3]

2 MODEL DESIGN AND DATA ANALYSIS

2.1 Model Framework Design

(1) Collect user-related transaction data of e-commerce industry, pre-process the data, remove the abnormal data in the data set, and then standardize the pre-processed data.

(2) The values of R, F and M are calculated from the data and constitute the value matrix of users.

(3) Apply the clustering algorithm to cluster the matrix and derive the corresponding user categories.

(4) Apply principal component analysis to calculate the importance of each indicator in the matrix as the weight value of the indicator.

(5) Use the weight values calculated in step 4 to calculate the value of each category and each user.

(6) The value of users is used to verify the effect of clustering.

(7) For each category, propose a specific marketing strategy based on the characteristics of the user to achieve the purpose of precision marketing.

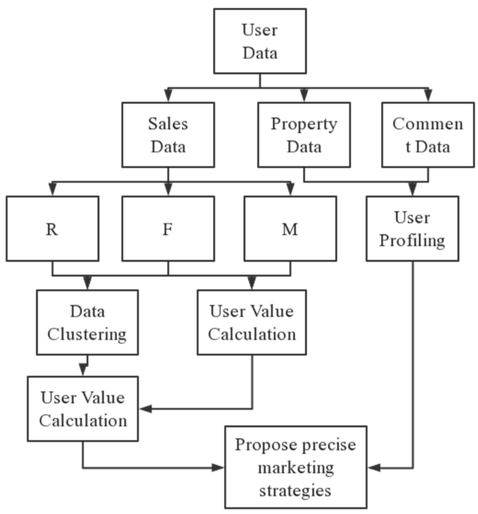

Figure 1：Marketing Model Design

2.2 Data Processing

The first point is that there are a large number of missing values in some indicators in the user attribute data and sales data; the second point is that there are refunded users in the sales data, which is not meaningful for user value calculation and will have a great impact on the model results if this part of data is kept. When further analyzing the user attribute data, we can see that the missing values of consumption level and city level indicators account for about 53.9% and 32.6% respectively. Since the city level indicator is a categorical indicator, this paper uses the plural filling method for filling. [4]

After cleaning and sorting the data set, the indicators of RFM model need to be constructed. According to the introduction of Chapter 2, it is known that: R indicator is the time when the customer purchased the goods in the previous time; F indicator is the total number of times the customer's consumption occurred from January 2021 to June 2021; M indicator is the average consumption amount of the customer from January 2021 to June 2021; using the user consumption dataset to calculate according to the definition of the three indicators. [5]

2.3 Calculation of RFM model indicators

In this paper, principal component analysis is used to calculate the weight value of each indicator. The principal component analysis method is a method to simplify the data set. It mainly uses orthogonal transformation to linearly transform all possible relevant variables to obtain a series of linearly uncorrelated variables. These uncorrelated variables derived from the original variables by linear transformation from are called principal components. [6] The specific implementation steps of principal component analysis are as follows. [7]

(1) Centering the original data. Centering is to have each data point subtracted from the mean of the category to which it belongs.

(2) Derive the covariance matrix of the features. If the original data has n features, the covariance matrix is a matrix of order n.

(3) The covariance matrix is decomposed by the eigenvalues. The eigenvalues and eigenvectors of the above covariance matrix are calculated. In this paper, if there are three indicators R, F and M, then 3 eigenvalues will be obtained, and each eigenvalue corresponds to one indicator. Expressed in the formula as follows.

$$\varsigma\alpha = \gamma\alpha \tag{1}$$

where α denotes the matrix composed of eigenvectors and γ denotes the column vector composed of eigenvalues.

(4) The eigenvalues corresponding to each indicator are used as the weights of each indicator. We calculate the value of each user, which is calculated by the formula

$$V = \alpha_1 \times R + \alpha_2 \times F + \alpha_3 \times M \tag{2}$$

where （α_1，α_2，α_3） is the eigenvalue derived by principal component analysis and V is the user's value matrix. After clustering, the value matrix of each category of users is then calculated according to the above formula, and the value matrix before clustering is used to verify the value matrix after clustering.

The principal component analysis method introduced above is used to carry out the weight value calculation of each indicator. Usually, the hierarchical analysis method is used to calculate the weight value of each indicator of the RFM model, which requires scoring the R/F/M value given to each consumer and then constructing a judgment matrix based on the score of each indicator. However, the hierarchical analysis method has less quantitative data and more qualitative components, which is not easily convincing, while when there are too many indicators, the data statistics are large and the weights are difficult to determine. The principal component analysis method is to take the indicators with linear correlation and recombine them, so as to obtain a new set of linearly unrelated composite indicators to replace the original ones. [8]

2.4 Data Clustering

The main steps of the experiment are.

(1) install the Python tool;

(2) import the processed data into the dataset and set the title for the dataset;

(3) import the K-Means analysis tool into the Python environment;

(4) normalize the imported data;

(5) build the analysis model and set the clustering K value;

(6) load the feature vector to be analyzed and perform the algorithm operation;

(7) obtain the classification results and merge them with the dataset;

(8) output Final results. The main procedure is as follows.

```
# import the K-Means analysis tool into the python environment and import the dataset using pandas

from sklearn.cluster import KMeans

Import pandas as pd

#Read the data used for clustering and create the teaching data table

loan_data=pd.DataFrame(pd.read_csv('SJRY_data.csw',header=0))

loan_data.columns

Index(['XM','ZW','GT','WT','age','DT','zc','ZCJB','XL','ZY','JZJ'],dtype='object')

#Process the training set and normalize the analysis data features

loan_data_zs=1.0*(loan_data-loan_data.mean())/loan_data.std()
```

build the analytical model, load the feature vector "job title, time in the workforce and age" that needs to be clustered, the number of clusters is 3 loan=np.array(loan_data[['ZW','WT','age']])

clf=KMeans(n_clusters=3)

#Substitute the data into the clustering model

clf=clf.fit(loop)

#Perform simulation training to get predicted values

Cluster=clf_KMeans.fit_predict(X)

print(cluster)

#Add clustering result labels to the original data table

loan_data['label']=clf.labels_

#View the clustering results

print(loan_data)

2.5 Cluster feature analysis

The clustered dataset was analyzed and the number of users in each clustering category is shown in Table 1.

Table 1 Number of users for each category

Clustering	Number of cases in each cluster
1	5199.000
2	8474.000
3	4307.000
4	7020.000
Effective	25000.000
Missing	.000

The mean, maximum and minimum values of each indicator were calculated for each category, as shown in Table 2.

Table 2 Average, maximum and minimum values of indicators in each category

Indicators Category		The first category	The second category	The third category	The fourth category
R	Average value	38.09	143.91	37.13	142.57
	Maximum value	171	181	174	181
	Minimum value	1	8	1	4
F	Average value	1.34	1.24	1.33	1.15

	Maximum value	7	7	7	7
	Minimum value	1	1	1	7
M	Average value	187.48	181.89	63.17	61.96
	Maximum value	277	278	271	265
	Minimum value	12.8	17.3	5.81	5.71

From the above table, it can be seen that the R indicator reflects the time of the customer's last consumption from now, the customers of the first and third category have a shorter time from the last consumption, which means that the higher the trust of these two categories of customers to the enterprise, proving that these two categories of customers are quality customers, but the values of the second and fourth categories are larger, and from the maximum value, we can conclude that some customers have not made repeat purchases in a long time cycle. the F indicator is M indicator refers to the average amount of money spent by customers over a period of time, and the higher value of the average consumption of customers in the first and second categories indicates that they contribute more to the profit of the company and may be loyal customers of the store.

To further measure each indicator more accurately for different categories of users so that the categories of users can be identified, the average value of each indicator is introduced and the results are shown in Table 1.

Table 3 RFM averages

Indicators	R	F	M
Average value	95.24248	1.2352	109

For the first category of users, their purchase period is less than the average, but their indicators are basically higher than the average, which means that these users are very loyal to the store and are high-value customers of the store; for the second category of users, their indicators are basically above the average except for the higher R indicator, which is relatively important to the store, but have a greater risk of churn, so define This type of user is defined as a key recovered customer. For the third category of users, all indicators are relatively average, so this category of users is defined as general customers. For the fourth category of users, their purchase cycle is greater than the average, and other indicators are lower than the average of indicators. This category of users is of little value to the store and is likely to have churned, so the fourth category of users is named churned customers. The value of each category is roughly the same as the results of the previous section based on the clustering results, with the highest value of high-value users, followed by key recovered customers, which also confirms the feasibility of applying the RFM model and K-means clustering algorithm to shoe e-commerce user segmentation and precision marketing, and the results are of some reference value. [9

3 CONCLUSION

In recent years, with the continuous development and maturity of the e-commerce industry, more and more enterprises have started to stay in e-commerce platforms, expecting to further expand their profits through e-commerce channels, but the entry of a large number of enterprises has led to a greater competitive pressure among e-commerce companies. In the face of a large customer base, the use of traditional marketing methods will have large investment costs, uncontrollable returns and other defects, so precision marketing has become a more reasonable marketing methods. Through big data technology to classify the store's customers, respectively calculate the value of each type of customer size, for different categories of customers to develop different marketing strategies.[10] In this way, we can save a lot of manpower and material resources, and also improve the efficiency of marketing.

This paper establishes an RFM model for the user data of a shoe store on Taobao platform, and borrows the idea of principal component analysis to calculate the weight coefficient of each index in the RFM model when calculating the user value, and then calculates the user value according to the weight coefficient. In terms of user classification, this paper firstly adopts the clustering algorithm to classify users according to each index, and then compares the average value of different indexes of each category of users with the average value of the overall indexes to classify the customers of the store into four major categories: high-value users, key recovery users, general users and lost users. For each category of users, the total value of their users was calculated to verify the clustering results of the clustering algorithm, and the results were found to be correct and reliable.

REFERENCES

[1] Ernawati E,Baharin S S K,Kasmin F. A review of data mining methods in RFM-based customer segmentation[J]. Journal of Physics: Conference Series,2021,1869(1).

[2] John.Trade intermediation and the organization of exporters[J].Review of International Economics, 2017,19(4): 634-648.

[3] Manikandan K.,Niveditha V. R.,Sudha K.,Magesh S.,Radha Rammohan S.. Design and Development of Customer Relationship Management Recommendations by Clustering and Profiling of Customers Using RFM [J]. International Journal of e-Collaboration (IJeC),2021,17(4).

[4] Mohiuddin Ahmed, Raihan Seraj,Syed Mohammed Shamsul Islam. The Algorit hm: A Comprehensive Survey and Performance Evaluation[J].Electronics,2020, 9(8):1295.

[5] Rahim Mussadiq Abdul, Mushafiq Muhammad, Khan Salabat, Arain Zulfiqar Ali. RFM-based repurchase behavior for customer classification and segmentation[J]. Journal of Retailing and Consumer Services, 2021,61.

[6] Tan Chun Kit, Nurulhuda Firdaus Mohd Azmi. Customer Profiling for Malaysia Online Retail Industry using K-Means Clustering and RM Model[J]. International Journal of Advanced Computer Science and Applications (IJACSA),2021,12(1).

[7] Wendy K. Bendoni, Social Media for Fashion Marketing: Storytelling in a Digital World (Required Reading Range) [M]. London: Kogan Page Ltd, 2017.

[8] Wang Xiaozhi and Huang Jieqiong and Jain Deepak Kumar. Enterprise Decision-Making and Analysis Based on E-Commerce Data Mining[J]. Wireless Communications and Mobile Computing, 2022.

[9] Yang J. Research on the marketing strategy of imported cross-border e-commerce in the context of big data [D]. Jiangxi University of Finance and Economics, 2020.

[10] Zong Yi, Xing Hao. Customer stratification theory and value evaluation—analysis based on improved RFM model[J]. Journal of Intelligent & Fuzzy Systems,2021,40(3).