The Study on High-Value User Identification of Localized Information Platform

Jinyi Li, Sheng Cao*
Management School, Wuhan Donghu University, Wuhan, Hubei, 430212, China
*Corresponding author’s e-mail: caosheng@wdu.edu.cn

Abstract. In the current business environment, enterprise competition is fierce, high value customers are increasingly important to enterprises. However, with the development of social networks, the traditional RFM model of user value stratification is not accurate enough. Therefore, this paper introduces the characteristics of customer social attributes and establishes RFMS model to classify customer groups more accurately and effectively and identify high-value customers. Thus, it can help enterprises allocate resources reasonably and carry out customer operation in a targeted way.

1. Introduction
In the current Internet environment, enterprise competition is increasingly fierce, and the premise for enterprises to achieve profits is customer access or purchase, customer is the source of enterprise profits. For enterprises, different customer values are different. From the popular point of view of marketing, it costs 4 or 5 times as much to acquire a new customer as it does to retain an existing one. Therefore, how to identify high-value customers, so as to operate them, and reasonably allocate the limited resources in the maintenance and development of the relationship with high-value customers is the key factor for the survival and development of enterprises. Only in this way, enterprises can reduce costs as much as possible, obtain competitive benefits and win the highest benefits.

In the measurement of customer value, RFM model is a classic hierarchical model. However, with the development of social networks, the accuracy of RFM model analysis is insufficient, and the dimension setting is not perfect enough to effectively classify customer value. This paper introduces the social attributes of customers and establishes the RFMS model.

2. Establishment of RFMS model
2.1. RFMS model
RFM model is a classic customer stratification model, which uses the three core dimensions of the general transaction link--Recency, Frequency and Monetary to segment customer groups.

With the emergence of social networks, a consumer's consumption behavior will be more and more able to influence other consumers and be influenced by other consumers. However, in the current new business environment, the relationship between brands and customers is not only about consumer shopping and pushing products, but more about interaction and mutual cognition. The communication attribute of a single customer has received more and more attention and application. Therefore, this paper introduces the characteristics of customer social attributes and establishes the RFMS model.
RFMS model adds the feature of S (Sociality) after customers visit the platform in a limited statistical period. The improved RFMS model has four features: Recency, Frequency, Monetary and Sociality.

R (Recency) refers to the time interval between the customer's last consumption before the statistical cut-off date. The shorter the interval, the more interested the customer is in the current product or service.

F (frequency) refers to the number of times customers consume during a limited statistical period. Generally, the more frequent customers consume, the more satisfied they are with goods or services and enterprises. At the same time, increasing customers' purchase frequency means taking market share from competitors and increasing your own turnover.

M (monetary) refers to the consumption amount of customers in the limited statistical period. Generally, the more amount of consumption, the stronger the consumption power of customers, and such customers should be the focus of the enterprise customer groups.

S (Sociality) refers to the number of times a customer shares with others in a limited statistical period. A high number of times of sharing indicates that the customer thinks the product or service is of high quality and is willing to share it with people around him. The consumption behavior of such customers can influence other consumers or potential consumers, which has a very high social value.

2.2. User value stratification based on RFMS model

The core of dimension confirmation is score determination. Score each customer's R/F/M/S value. The division of the score interval should be combined with the actual product, and the size of the score depends on the characteristics of the product. The more accord with the behavior of the product sales, the higher the score.

After the score of each customer's characteristics is obtained, each of the four RFMS scores of each customer is compared with the average score of this item. If the score is higher than the average score, it is recorded as "1", while lower than the average score, it is recorded as "0". Finally, 16 customer value classification results will be obtained.

3. Data analysis

3.1. Data Sources

The data set adopted by the model in this paper is the user data of "Donghu Circle" WeChat applet, a campus E-commerce localization information platform. The statistical time width is four months, and the data set contains 565 pieces of user profile information, access data, social behavior data, etc.

"Donghu Circle" WeChat app is a localized information platform focusing on Wuhan Donghu University and its surrounding areas, which mainly integrates and publishes all kinds of information and services. It was launched in November 2020 and is now in the operation stage, mainly serving college students.

3.2. "Donghu Circle" high value user identification based on RFMS model

Before using RFMS model, the meaning of each feature in the model should be determined according to its product characteristics. Since the "Donghu Circle" does not involve the transaction behavior between users and platforms, "M (Monetary)" is not used in the use of the model.

Taking "Donghu Circle" as an example, the meaning of features in the model is as follows:

(1) Feature R

R (Recency) refers to the time interval between the user's last visit before the statistical deadline, in days, where END_DATE represents the statistical deadline, and ACCESS_DATE represents the user's last access date.

\[
R = \text{END\_DATE} - \text{ACCESS\_DATE}
\]

(2) Feature F

F (Frequency) refers to the number of times the user visits during the limited statistical period, and the unit is times.

\[
F = \text{ACCESS\_COUNT}
\]
(3) Feature S

S (Sociality) refers to the number of times a user sends out sharing links within a limited statistical period, the unit is times.

$$S = \text{SHAREING\_COUNT}$$

3.2.1. Data analysis charts of “Donghu Circle”

Figure 1/2/3/4 below shows the original data analysis diagram of the "Donghu Circle":

![Figure 1. Data box counting chart](image1)

![Figure 2. Data heat map](image2)
3.2.2. “Donghu Circle” user value stratification

3.2.2.1. User value scoring mechanism

RFMS model was scored by 5-point system. According to the industry experience, the R value is set as a span of 20 days, and the interval is left closed and right open:

| R score | R value | meaning                           |
|---------|---------|-----------------------------------|
| 1       | (80, ∞) | Not visited for more than 80 days |
| 2       | [60, 80) | not visited for 60-79 days        |
| 3       | [40, 60) | not visited for 40-59 days        |
| 4       | [20, 40) | not visited for 20-39 days        |
| 5       | (0, 20)  | not visited for 0-19 days         |

Figure 5. R value setting
3.2.2.2. User value stratification

Users are divided into 8 categories according to whether each R/F/S feature score is higher than the average value, as shown in the following table:

- **Figure 8. User value stratification**

An auxiliary column of the crowd numerical is introduced to connect the three values of whether R/F/S is greater than the average value:

$$rfs[CN] = (rfs[IF R > Average] * 100) + (rfs[IF F > Average] * 10) + (rfs[IF S > Average] * 1)$$

- **Figure 9. The crowd numerical**

- **Figure 10. Connect R/F/S in series**
The "crowd number" is a numeric type, and the preceding "0" is automatically skipped. For example, "1" represents "001", which corresponds to the group of "general maintenance customers"; "10" stands for "010", which corresponds to the group of "new customers".

In order to get the final crowd tag, a judgment function is defined to return the corresponding classification tag by judging the value of crowd value:

```python
# If R/F/S > Average
def transform_label(x):
    if x==111:
        label = 'IV_User'
    elif x==011:
        label = 'IR_User'
    elif x==101:
        label = 'IC_User'
    elif x==001:
        label = 'IRe_User'
    elif x==110:
        label = 'P_User'
    elif x==100:
        label = 'N_User'
    elif x==010:
        label = 'GM_User'
    elif x==000:
        label = 'L_User'
    return label
```

Figure 11. Judgment function

"IV" _ "User" = important value user; "IR" _ "User" = important return user; "IC" _ "User" = important cultivation user; "IRe" _ "User" = important retention user; "P" _ "User" = potential user; "N" _ "User" = new user; "GM" _ "User" = general maintenance user; "L" _ "User" = lost user.

Finally, the tag classification function is applied to the crowd value column:

```python
rfs['TC'] = rfs['CN'].apply(transform_label)
rfs['head']
```

Figure 12. Applied to the crowd value column (1)

| visitor | R  | F  | S  | R-score | F-score | S-score | R>average value | F>average value | S>average value | crowd numerical | customer type               |
|---------|----|----|----|---------|---------|---------|----------------|----------------|----------------|----------------|---------------------------|
| user1   | 10 | 25 | 0  | 5.0     | 5.0     | 1.0     | 0              | 1              | 0              | 10             | General maintenance user |
| user2   | 98 | 4  | 1  | 1.0     | 4.0     | 1.0     | 1              | 1              | 1              | 100            | potential users           |
| user3   | 106| 2  | 0  | 1.0     | 2.0     | 1.0     | 1              | 0              | 0              | 100            | new user                  |
| user4   | 3  | 5  | 3  | 5.0     | 5.0     | 3.0     | 0              | 1              | 1              | 11             | Important return users    |
| user5   | 51 | 3  | 0  | 3.0     | 3.0     | 1.0     | 0              | 0              | 0              | 0              | Lost users                |

Figure 13. Applied to the crowd value column (2)

3.3. Analysis of user value stratification conclusion of "Donghu Circle"  
After layering, each user has his own RFS tag. The visual report is as follows:
We define the importance of each customer group according to the different characteristics of each customer group.

(1) The number of important value users accounts for 0.88%. These people are users who visit recently, visit frequently and are willing to share the products of E-commerce information platform on campus. In the later operation, more resources should be inclined to provide personalized services and additional services.

(2) The number of important return users accounts for 0.53%. These groups have many visits and share more times, but they have no access recently, which can provide them with useful resources and win back them through renewal or update products.

(3) The number of potential users accounts for 6.55%, which has many visits and frequent recent visits, but the sharing times are less, which needs to be excavated.

(4) The proportion of new users is the most--48.67%, which has recently been visited, but the frequency is not high, the sharing frequency is not much, and it is easy to lose. But it has promotion value, can improve customer interest and create brand awareness through activity marketing.

(5) The number of general maintenance users accounted for 10.62%. This kind of people visited more times, but there was no visit recently, and the sharing times were less, generally maintained. You can re-connect with them with valuable resources sharing, discount hot products, etc.

(6) The number of lost users accounts for 32.74%. This kind of group belongs to hibernating customers, and users need to be restored, otherwise, they need to give up the worthless users temporarily.

(7) The proportion of important cultivation users and important retention users is 0. Because the sampling product is the initial campus E-commerce information platform product, the two types of data are 0, which is likely to be related to the content quantity and quality of the products of this sampling platform. It is necessary to observe the product in the later stage of operation.

(8) Important cultivation users have access in the near future, and share more times, but the recent visit frequency is not high, so it needs to be identified; the most important retention users visit times, but they haven’t visited for a long time, so they need to contact with each other to improve the retention rate.

Figure 14. The visual report
4. Conclusion
This paper aims at the current social E-commerce industry user value mining, analyzes its shortcomings, and proposes a user value evaluation method based on RFS model. This method is different from the existing RFM model of customer value classification, which only considers the characteristics of customer consumption interval, consumption frequency and consumption amount, but introduces the characteristics of customer social attributes to analyze the value of users to enterprises more accurately and effectively. Since the data in this paper are based on a campus E-commerce information platform product that has just started, in order to ensure the stability of the model, it is necessary to accumulate enough data in the later period and update the model results. After each update, follow-up observation and monitoring shall be conducted. If the difference is too large, adjustment shall be made in time.

Acknowledgments
This work was supported by the education research project of Wuhan Donghu University, which established in 2019 and name “The study on E-commerce Innovation and Entrepreneurship Cultivation Driven by Subject Competition ‘Donghu Model’ ” in finance. And Professor Sheng Cao is the principal of this project.

References
[1] HUANG H. (2016) Deyang Mobile Company's high-value customer loss analysis and countermeasures [D]. Sichuan: University of Electronic Science and Technology
[2] AN X.Q. (2012) The study on customer value recognition in C2C environment based on RFM model [D]. Sichuan: Southwestern University of Finance and Economics.
[3] YANG X, XU X.R. (2021) Customer value data mining program based on LRFMC model [J]. Computer knowledge and technology, 17(06):22-25.
[4] YAN J. (2015) The study on Customer Value and Customer Segmentation in Social Network Environment-- Based on the Influence of Internet Word of Mouth[D]. Tianjin: Tianjin University of Finance and Economics.
[5] DAI L.D. (2019) 10 minutes, quickly understand the RFM user analysis model [EB/OL].https://lidong.blog.csdn.net/article/details/103588782?ops_request_misc=&request_id=&biz_id=102&utm_term=rfm%E6%A8%A1%E5%9E%8B&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-3-103588782.pc_search_result_hbase_insert.