Chapter 14
The Study of User Download Behavior in Application Stores and Its Influencing Factors

Changzheng He, Wei Shu and Dongyue Du

Abstract The study of user download behavior and its influencing factors will contribute to a deep understanding of the application stores and provide some practical guide to the operation of application stores. Based the classic RFM model and the actual situation of the application stores, this paper develops the TDRFM model to describe the application stores' user download behavior. We use the K-means clustering and Group Decision Making method based the behavioral indicators and obtains four user types: the high-value uses, general-value users, loss user. Then, we study the impact of the system upgrade and the product attributes to users' download behavior by using the statistical analysis. The results show that the application store upgrade has no significant impact on the high-value users while have a significant effect on the general-value and potential-value users download behavior. The impact of application type, development type price, review and application size on users has been verified. This paper provides a method of studying user behavior in application stores.

Keywords Application store · User download behaviour · TDRFM model · Influencing factors

14.1 Introduction

The mobile internet is innovative business mode, which is growing fast. It combines the mobile communications and the internet as a whole. Its pace of development has caught up with the traditional internet. It's an important driving force for the network

---

C. He (✉) · W. Shu
Business School, Sichuan University, Chengdu 610064, P. R. China
e-mail: hechangzheng@scu.edu.cn

D. Du
Planning and Construction Department, Sichuan University, Chengdu 610064, P. R. China

J. Xu et al. (eds.), *Proceedings of the Seventh International Conference on Management Science and Engineering Management (Volume 1)*,
Lecture Notes in Electrical Engineering 241, DOI: 10.1007/978-3-642-40078-0_14,

economic. The application stores are the key nodes of the mobile internet industry chain. After the App Store pioneered by Apple gained great success, various types of application stores are springing up all over the world. The study of application stores' user download behavior and its influencing factors will contribute to a deep understanding of the application store and provide some practical guides for the operation of application stores.

As a new business mode, little attention has been paid to the study of the application stores. Current research consists of three parts. First, the product development in application stores. For example, Bergvall-Kareborn and Howcroft [1] considered the problem of application product development; Liu et al [2] analyzed the development trend of the medical applications in the application stores; Ning [3] analyzed the developer's incentive strategy in application stores. Second, the operating strategy of application stores. For example, Ghezzi et al [4] analyzed core competitiveness of the application store industry. Tuunainen [5] discussed the critical factors for application stores' success. Third, marketing strategy in application stores. For example, Bellman [6] verified the validity of brand strategy. Gans [7] explored product pricing issues in application stores. In addition, the study also included empirical analysis of the consumers' use intention and usage behavior by Li [8].

The application store is a typical two-sided market platform. The platform operators provide application product services to both buyers and sellers to facilitate transactions. On the one hand, the application stores provide one-stop service from purchase (downloads) to use for mobile internet users. Therefore, how to match the interests of the both sides and enhance the integrated value of each node of the industrial chain has become the urgent problem of the application stores. Therefore, the research on user download behavior as well as customer value evaluation plays a vital role.

14.2 User Segments Based on Behavioral Variables in Application Stores

14.2.1 Behavior Variable Selection

User value identification is the basis of value-oriented differentiation in marketing. The classic approach use the RFM model to measure the user value by using past buying behavior [9]: purchase time (Recency, R), the number of purchases (Frequency, F) and the purchase amount (Monetary, M). For the application store operators, they are not only concerned about the direct profit of application download but also indirect profit caused by the free application download. So this paper develops a TDRFM model based on the original RFM model which describes the user's download behavior in the application stores. The meaning of each variable in the

TDRFM model as follows:

T : last download time, which measures the time interval of last users download.
D : users' total number of download in the study period.
R : last purchase time, which measures the time interval form the most recent purchase.
F : total number of purchases.
M : total purchase amount in the study period.

14.2.2 Data Processing

The data for model building was from the database of China's W application store, including users download recorded data from May 10, 2010 to September 22, 2011. We sorted out TDRFM variable values by excel.

First, delete the users which download less than 10 times and the users have more than 180 days away from last download. This is because these users enter into the application stores by chance. Obviously they do not belong to regular users. We gained 1280 users' download records (N) as the sample data, which includes 245 purchase records (n). The statistics describe the characteristics of the users are shown in Table 14.1.

All the data are standardized; because the indicator variables T, D and R, F, M in the model have different units of measurement. The Max-Min standardization method is used as follow:

$$X' = (X - X^s)/(X^l - X^s), \tag{14.1}$$

$$X' = (X^l - X)/(X^l - X^s), \tag{14.2}$$

where X' is the standardized value, X is the original value, X^s is the minimum value of the variable , X^l is the maximum value of the variable.

There is a negative relationship between $T(R)$ and customer value, so their standardization should use Equation (14.3). However, the index D, F and M have a positive relationship with user value, so their standardization should use Equation (14.1).

Table 14.1 Sample descriptive statistics ($N = 1280$, $n = 245$)

Variable indicators	Mean	Std.Dev.	Min	Max
Last download time (T)	205	174.936	1	500
users' total number of download (D)	94.86	119.067	1	1395
Last purchase time (R)	297.66	118.464	1	499
Total number of purchase (F)	5.20	5.810	1	48
Total amount of purchase (M)	14.873	20.261	0.1	195.07

14.2.3 *K*-means Clustering

K-means is a commonly used clustering method in marketing segmentation. The algorithm flow is simply described as follow:
Step 1. Randomly selected K object from the data set as initial cluster centers;
Step 2. Find which cluster it belongs to for each object based on its distance from the initial centers;
Step 3. Update the cluster centers based on the average of each cluster;
Step 4. Repeat Step 2 until each cluster centers no longer change.

Numerous studies of user value clustering based on the RFM model [10–12] have been done. Similarly, we get final number of clusters of 5. After clustering, we will analyze each cluster to see if the indicators are greater than (equal) or less than the overall customer base average. Table 14.2 shows the clustering results.

Table 14.2 K-means clustering result

Cluster number	Count	T-means	D-means	R-means	F-means	M-means	Type represents
I	568	77.268	98.364	445.586	2.172	5.56	$T\downarrow D\uparrow R\uparrow F\downarrow M\downarrow$
II	527	393.918	40.524	444.385	3.385	9.98	$T\uparrow D\downarrow R\uparrow F\downarrow M\downarrow$
III	165	60.176	194.982	242.697	4.909	13.04	$T\downarrow D\uparrow R\uparrow F\downarrow M\downarrow$
IV	9	94.556	294.333	185.667	27.556	98.41	$T\downarrow D\uparrow R\downarrow F\uparrow M\uparrow$
V	11	12.455	851.636	319.333	6.667	18.47	$T\downarrow D\uparrow R\uparrow F\uparrow M\uparrow$
Overall	1280	205	94.86	297.66	5.20	14.873	—

Definition 14.1. If a cluster average of R (or F, M, T, D) value is greater than (or equal to) the R (or F, M, T, D) mean over all the customers. We denote this cluster of as R (or F, M, T, D)$\uparrow$; If a cluster average of R (or F, M, T, D) values less than the mean R (or F, M, T, D), then this cluster of is represented by R (or F, M, T, D)$\downarrow$.

From the Table 14.2, we can see:
Cluster I: Users have the shorter download cycle times. They have higher D-value than the overall average. Their purchase frequency and purchase amount are very small.
Cluster II: Users in this cluster have the largest T-value and the second largest R-value which indicates that there is a long time that these users are not logged in the application stores. Their purchase frequency and purchase amounts also are very low.
Cluster III: Users download a large amount, but the number of purchases and the purchase amount is smaller or slightly lower than the overall average.
Cluster IV: has the highest number of download (F), the largest total amount of purchase (M-value).
Cluster V: Users have a very short period for download (T). They often login on the

application store application to download and belong to the stores' loyal users, but they rarely pay for the applications.

14.2.4 User Value Evaluation

Application store is an innovative business mode. Different people have different behavior which is reflected in the weight of the variables in TDRFM model. In order to determine these weights, we use the famous group decision making method Analytic Hierarchy Process (AHP) [13]. Eight members are invited which consist of two applications store personnel, two loyal users, two mobile internet observers and two scholars. The relative importance of behavioral variables in TDRFM is obtained by the formation of scale judgment matrix through pairwise comparison. The final weight is calculated as follows.

$$W = (W_T, W_D, W_R, W_F, W_M) = (0.0601, 0.2632, 0.1098, 0.2829, 0.2840).$$

After getting the weights, we use the following formula to calculate the composite score to measure the customer value.

$$\text{Score}_i = \sum_{k=T,D,R,F,M} (k_b - \text{means})_i \times W_k \ (i = 1, 2, 3, 4, 5), \tag{14.3}$$

where Score_i represents the value scoring for the customer in cluster i. It is the weighted average of the variable after the standardization. After calculating, we gained the value of each cluster of application store user as shown in Table 14.3.

Table 14.3 User value segmentation

Cluster number	T_W-means	D_W-means	R_W-means	F_W-means	M_W-means	Score	Ranking
I	0.0509	0.0184	0.0118	0.0071	0.008	0.096	4
II	0.0128	0.0075	0.0120	0.0144	0.0144	0.061	5
III	0.0530	0.0366	0.0565	0.0235	0.0189	0.1887	3
IV	0.0488	0.0554	0.0691	0.1598	0.1432	0.476	1
V	0.0587	0.1606	0.0396	0.0341	0.0268	0.320	2

We can see from Table 14.3, the value of cluster IV and V users rank the highest, then come cluster III and I. The value ranking of cluster II is the lowest of all, which is also consistent with our previous qualitative description. The users of high loyalty correspond to the higher value. The users' loyalty is low because a long time not logging in the application stores (such as cluster II), so their value is low.

14.2.5 Analysis of User Behavior Characteristics

We combine the results of Tables 14.2 and 14.3 to analyze the characteristics of the user download behavior of different sub-groups. Table 14.4 gives the description.

Table 14.4 The description of user type

Cluster number	Number of users	Number of purchase	Value ranking	Type description
I	568	29	4	Regular free users
II	527	39	5	Loss user
III	165	165	3	Regular paid users
IV	9	9	1	Golden paid users
V	11	3	2	Golden free users

(1) High-value users

Users in cluster IV download a large amount and buy the highest number. The amount of consumption is also the largest for this cluster and their value rank the highest. We call them the gold paid subscribers. They are the direct source of application store revenue.

Users in cluster V have the largest number of download and the download cycle is very short. They often login on the application stores to download and belong to the stores' loyal customers. Their value rank the second. They are called the application store golden free users. Although such users will not directly provide profit to the stores, they play a crucial role in the operation of the entire application store platform.

The golden paid users and golden free users belong to the high-value users with high degree of loyalty. Application stores should pay more attentions to maintain good client relationships with them.

(2) General-value users

Users of cluster III are regular paying customers. They download a large amount. Each user has purchase behavior, but the purchase amount is smaller than the overall average. Their values rank the third. They are the general-value user.

Such users are mainly paid subscriber. They have the habit of buying application software. They have great consumption potential. So the application store operators should stimulate and guide their purchase. For example, they should take the promotional tools to enhance their buying consumer strength. However, they only occupy 12.89% of the overall users, which indicates that the few people are willing to pay customers in the application stores and most users are unwilling to pay.

(3) Potential-value of the users

Users in cluster I are regular free users which represent the typical user groups of the application stores in China. They do not want to spend money to buy mobile phone application software. Only a handful of users will accidentally buy the product.

Regular free users are the users with potential-value. Their free consumption behavior is caused by consumption habits. The application stores cannot expect to change their spending habits, but can gather the profit by business mode innovation such as advertising. They will bring indirect profit. So to develop the value of such users can also become an important source of profit.

(4) Loss users

Users in this cluster are loss users. They have a very long time (mean 13 months) not logging in the application stores to download. The operators do not need to put more efforts to this users.

14.3 Impact of Application Stores' Upgrade on the User Download Behavior

Application store upgrade is the operational development of a series of functions carried out by the application stores which include the improving, repairing, and operation. The purpose of upgrade is to make the application stores basis platform more complete and powerful and attract more users.

In this paper, we select a major upgrade in the *W* Application Store on March 6, 2011 to analyze the customer value change before and after this upgrade. The study time period is divided into two stages:

Before the upgrade: May 10, 2010 to March 2011, 5th, a total of 300 days.

After the upgrade: March 6, 2011 to September 2011, 21, a total of 200 days.

Table 14.5 The results of t-test of user value change before and after upgrade

Categories of users	Behavioral indicators	Test results	Analysis
High-value users	Number of download per hundred days Number of download per hundred days Purchase amount per hundred days	Accept the null hypothesis Accept the null hypothesis Accept the null hypothesis	Application stores upgrade had no significant effect on high-value users' download behavior
General-value users	Number of download per hundred days Number of download per hundred days Purchase amount per hundred days	Reject the null hypothesis Reject the null hypothesis Reject the null hypothesis	Application stores upgrades have a significant effect on the general-value and potential-value users download behavior
Potential-value users	The number of download per hundred days	Reject the null hypothesis	

We set the research hypothesis H_0 as: There is no significant change in behavioral indicators before and after the application store upgrade. By using independent samples t-test analysis, we try to verify this hypothesis. The results are shown in Table 14.5.

Table 14.6 Various types of classification standards used on the indicators

Type of application	Price level (x stands for price)	Score level (x stands for score)	Review number (x stands for the number of review)	Application size (x stands for application size, units: M)
Game class	Low price ($x < 3$) Medium price ($3 \leq x < 5$) High price ($x \geq 5$)	Poor evaluation ($x \leq 3$) Medium evaluation ($3 < x \leq 4$) High evaluation ($x > 4$)	Less review ($x \leq 1$) Medium review ($1 < x \leq 3$) Much review ($x > 3$)	Small applications ($x \leq 0.5$) Medium applications ($0.5 < x \leq 2$) Large applications ($x > 2$)
Tools class	Low price ($x < 5$) Medium price ($5 \leq x < 10$) High price ($x \geq 10$)	Poor evaluation ($x \leq 3$) Medium evaluation ($3 < x \leq 4$) High evaluation ($x > 4$)	Less review ($x \leq 1$) Medium review ($1 < x \leq 3$) Much review ($x > 3$)	Small applications ($x \leq 0.5$) Medium applications ($0.5 < x \leq 1$) Large applications ($x > 1$)
Reading class	Low price ($x < 2$) Medium price ($2 \leq x < 4$) High price ($x \geq 4$)	Poor evaluation ($x \leq 3$) Medium evaluation ($3 < x \leq 4$) High evaluation ($x > 4$)	Less review ($x \leq 1$) Medium review ($1 < x \leq 3$) Much review ($x > 3$)	Small applications ($x \leq 0.5$) Medium applications ($0.5 < x \leq 1$) Large applications ($x > 1$)

14.4 Impact of Product Attribute on User Download Behaviour

The product attributes of application of include application type, developer type, quality, price, review, and application size. During the time of the download, users more or less will consider these attributes factors. By using independent samples t-test and ANOVA analysis, we investigate the impact of above factors on the applications downloads. We focus on the high-value, general-value and potential-value users of and of the user. A total of 1749 applications are selected, which include 517 game applications, 483 application of the tool and 749 reading class application. In order to facilitate the examination and analysis of price level, score level, the number of reviews and application size, these four indicators is discretized into three types. The criteria for the classification are shown in Table 14.6. The test results are shown in Table 14.7.

Table 14.7 Various types of classification standards used on the indicators

Factors		Test methods	Test results
Application type (games, tools and reading)		Single factor analysis of variance	Application type has impact on download
The type of developer Individual (developers and corporation developers)		Independent samples t-test	Developer type has effect on download
Pricing	Whether free	Independent samples t-test	Whether free has a significant effect on application downloads
	Price level	Single factor analysis of variance	Price level has significant effect on the tool application but no effect on the game and reading application
Review	With or without reviews	Independent samples t-test	Users download more applications with review than application without reviews
	Score level	Single factor analysis of variance	Only high score level only tools have significant impact on download.
	Number of reviews		The more number of comments, the more the downloads
Application size		Single factor analysis of variance	Application size has effect on the game and reading application download but no significant effect on tool application downloads

14.5 Conclusions

Based on the user data in application stores, this paper analyzed the users' download behavior variables, segments users based on their value. Three types of user are identified: high-value customers, general-value of users and potential-value user. We also analyzed the impact of application store upgrade and product attributes on the user download behavior. The following conclusions can be drawn:

- High-value users are golden users in application stores and they are direct source of profit. The general-value users are accustomed to pay to download applications. Application stores operators should through enhance their purchase motivation by using operator promotions;
- Upgrade of the application stores significantly affect user download behavior of high-value user but do not have a greater impact on the general-value users and potential-value users.

- Application type, pricing, reviews and application size have a significant effect on users download behavior.

This paper provides a deep understanding of customer behavior in application stores.

Acknowledgements This work is supported by National Natural Science Foundation of China (Grant No. 71071101).

References

1. Bergvall-Kareborn B, Howcroft D (2011) Mobile applications development on apple and google platform. Communications of the Association for Information Systems 29:565–580
2. Liu C, Zhu Q (2011) Status and trends of mobile-health applications for iOS devices: A developer's perspective. Journal of Systems and Software 84:2022–2033
3. Ning X (2011) Developer of mobile application store incentives for research. Beijing University of Posts and Telecommunications, Beijing
4. Ghezzi A, Balocco R, Rangone A (2010) How a new distribution paradigm changes the core resources, competences and capabilities endowment: The case of mobile application stores. In: Proceedings of Ninth International Conference on Mobile Business 33–42
5. Tuunainen VK, Tuunanen T, Piispanen J (2010) Mobile service platforms: Comparing Nokia OVI and Apple App store with the IISI Model. In: Proceedings of 10th International Conference on Mobile Business 74–83
6. Bellman S, Potter R (2011) The effectiveness of branded mobile phone apps. Journal of Interactive Marketing 25:191–200
7. Gans J (2012) Mobile application pricing. Information Economics and Policy 24:52–59
8. Li L (2011) Consumer mobile application store with intention and use behavior. Nanjing University of Posts and Telecommunications, Nanjing (In Chinese)
9. Hughes A (1994) Strategic database marketing. Probus Publishing Company
10. Zhao X, Huang X, Sun F (2005) An optimization model for promotion mix strategy based on RFM analysis. China Management Science 13:60–64 (In Chinese)
11. Liu DR, Shih Y (2005) Integrating AHP and data mining for product recommendation based on customer lifetime value. Information & Management 42:387–400
12. Jiang G, Liu P, Huang T (2007) A customer value segmentation based on AHP method. Computer Engineering and Applications 8:238–241
13. Qin S (2003) Comprehensive evaluation of theory and application. Publishing House of Electronics Industry