12-13-2015

An Empirical Investigation of Factors Impacting Application Downloads in Mobile App Stores

Ying Wang
Texas Tech University, ying.wang@ttu.edu

Jaeki Song
Texas Tech University, jaeki.song@ttu.edu

Miguel Aguirre-Urreta
Texas Tech University, miguel.aguirre-urreta@ttu.edu

An Empirical Investigation of Factors Impacting Application Downloads in Mobile App Stores

Ying Wang
Texas Tech University
ying.wang@ttu.edu

Jaeki Song
Texas Tech University
jaeki.song@ttu.edu

Miguel Aguirre-Urreta
Texas Tech University
miguel.aguirre-urreta@ttu.edu

ABSTRACT

The customer value theory has been widely applied to investigate salient factors that influence customers' purchasing intention in the context of e-commerce. This study extends this literature by combining predictors of product performance with customer value framework to explore the effect of in-store information on mobile application downloads. It calls for understanding the unique characteristics of mobile environment. In this study, we apply text-mining techniques to analyze customers' reviews and product descriptions in the mobile application store and find the embedded meaningful information valued by customers. We also find that for mobile applications, price, number of raters, and helpful information in customer reviews and product descriptions have significant impacts on the number of downloads, while average rating has no influence on application downloads. Theoretical and practical implications are discussed.

Keywords

Mobile application, customer value, product performance, in-store information, text mining

INTRODUCTION

Mobile application has become an important role that permeates people's daily life. As of July 2014, there are 1.2 million applications available in Apple Application Store and 75 million applications have been downloaded from it (Global mobile statistics, 2014). With intensive competition among mobile applications, some applications are very popular, having hundreds of millions of downloads; while others just have thousands of downloads. Therefore, what factors cause this huge difference on application performance becomes a critical problem for application developers.

A large number of studies have been conducted to investigate salient factors affecting sales of products and customers' purchasing behavior in the context of e-commerce, such as average rating and online customer review (Duan et al., 2008; Mudambi and Schuff, 2010), but few researchers have attempted to prove that such product information is still valid in the mobile environment. Mobile customers' behaviors are different with those in the context of web environment due to the characteristics of mobile devices as well as the mobile environment. The mobile devices' screen display limitations restrict mobile users' access to rich multimedia contents (Pham et al., 2000). Due to the inconvenience of searching for multimedia contents and the disability of displaying information from multiple resources synchronously, the mobile devices makes users rely on in-store information only to distinguish applications when they make downloading decisions. Therefore, the specific content of in-store information becomes critical in the mobile environment to make applications outstanding and attractive to potential users. Despite of the hyper-competitive in the mobile app stores, we do not have enough knowledge about how mobile applications can appeal to customers. The principal objective of this study is therefore to fill in this gap in both theoretical and practical fields by investigating what information about applications valued by customers has impacts on application performance.

In this study, we apply customer value theory to analyze the impact of in-store application information considered predictors of application performance, on the number of downloads. We use text-mining techniques to find the impact of text-related information, such as product description and online customer review. Our study makes theoretical contributions by revealing a customer's purchasing mechanism in the mobile environment and practical contributions by providing mobile application developers with useful guidance on taking effective actions to increase the number of downloads.

The rest of this paper is organized as follows: First, we briefly discuss the relevant literature and propose our hypothesis. Next, we describe the data collection and develop our research method. Then we present the empirical results and related discussion. Finally, we discuss contributions and limitations in this study and point out possible directions for future research.

CONCEPTUAL BACKGROUND

In marketing and e-commerce literature, customer value coming from different sources is considered an important predictor of customer buying behavior or purchase decision (Babin et al., 1994; Dodds

et al., 1991). Among the sources, customer value coming from products is created by value-chain activities associated with new product development (Smith and Colgate, 2007). A large number of studies have been conducted to find predictors of new product development, which can be commonly categorized as product, strategy, process, and marketplace characteristics (Henard and Szymanski, 2001). Therefore, these predictors of new product development could be the sources of customer value that influences customer purchasing behavior.

Customer Value Theory

Woodruff (1997) describes customer value as "a customer's perceived preference for and evaluation of those product attributes, attribute performance, and consequences arising from use that facilitate (or block) achieving the customer's goal and purposes in use situations." Based on this definition, Smith and Colgate (2007) develop a comprehensive framework of customer value by building on the strengths of previous frameworks and mitigating their key weaknesses. They propose that there are four types of value that can be created by organizations: functional value, experiential value, symbolic value, and cost value. These four types of value may come from different sources, which cover five aspects: information, products, interactions, environment, and possession transfer. In the mobile environment, there is limited information available from multimedia and very few associated activities. Therefore, for mobile application users, the customer value mainly comes from the characteristics of products. The in-store information that reflects application characteristics is the source of customer value.

Predictors of Product Performance

Since product innovation is increasingly valued as a critical component of sustainable business success, there are a large number of studies to explore the drivers of new product success (Cooper and Kleinschmidt, 1987; Montoya-Weiss and Calantone, 1994). The drivers can be commonly classified into four categories: product, strategy, process, and marketplace (Henard and Szymanski, 2001). In this study, the research object is mobile applications, not new mobile applications. The new product development process and market response to new product introduction are not within our research scope. Therefore, we only consider the performance drivers within the first two categories: product characteristics and strategy characteristics. We propose that an application performance is a function of product characteristics and strategy characteristics, which are the sources of customer value.

HYPOTHESIS

The hypotheses are proposed on the basis of the customer value framework built by Smith and Colgate (2007). We argue that in-store information describing an application's product characteristics and strategy characteristics, which are the sources of customer value, can serve as factors having impact on the application performance.

Functional Value

Functional value refers to the extent to which a product or a service has desired characteristics or performs a desired function (Smith and Colgate, 2007). A product's functional value comes from its characteristics or attributes. It can be measured in either monetary or functional terms. We use functional quality to indicate functional value in this research. Functional quality is the perceived overall excellent and expected performance of an application.

In the digital world, many users can post their opinions about a product by giving a rating or writing an online customer review, showing their perceived functional quality of the product. Previous studies show that reviewers' average rating can influence product sales (Ye et al., 2009) and online customer reviews influence product sales and consumer decision-making (Dellarocas, 2003; Duan et al., 2008). Google Play Store also indicates that prospective users consider ratings and reviews as key benchmarks indicating the quality of applications. Therefore, rating and online customer reviews are key factors that influence potential users' downloading behavior. Additionally, the number of raters is also one indicator that reflects the popularity of an application. The distribution of word-of-mouth increases consumer awareness (Liu, 2006) and the popularity of product is positively associated with its sales (Oberholzer-Gee and Strumpf, 2007. Therefore, we hypothesize that

Hypothesis 1a (H1a): Average rating has a positive impact on the number of downloads.

Hypothesis 1b (H1b): Helpful information in online customer review has a positive impact on the number of downloads.

Hypothesis 1c (H1c): Number of raters has a positive impact on the number of downloads.

Experiential Value

Experiential value refers to the extent to which a product or a service creates appropriate experience, feelings, and emotions for the customer (Smith and Colgate 2007). Experiential value is derived from certain feelings that a customer associated with a product. This feeling depends principally on how the product looks and the extent to which customers relate to it.

In the mobile application store, to attract potential users and impress them, the developers have the option of uploading a demonstration video and screenshots to provide a vivid description of their applications. According to advertising strategies, vivid information can be used to influence consumers' attitudes

towards brands and products (Appiah, 2006). Many advertising scholars and marketing professionals held a general assumption that increasing the vividness of a message enhances its persuasiveness (Appiah, 2006). Therefore, uploading a demonstration video and screenshots is a good strategy to attract potential users' attention and increase the number of downloads.

Besides demonstration video and screenshots, the developers can also use words to describe their applications. Product description is complementary to a demonstration video and screenshots, considering the availably of network and data traffic in the mobile environment. When customers have limited access to the Internet, they can look at product description to get information about applications. From the product description, users can get helpful information about various aspects of applications, such as functions, operations, privacy policy, etc. Therefore, we hypothesize that:

Hypothesis 2a (H2a): Demonstration video has a positive impact on the number of downloads.

Hypothesis 2b (H2b): Helpful information in product description has a positive impact on the number of downloads.

Hypothesis 2c (H2c): Number of screenshots has a positive impact on the number of downloads.

Symbolic Value

Symbolic value refers to the extent to which customers attach or associate psychological meaning to a product (Smith and Colgate 2007), which is the perceived utility of a product based on its ability to enhance the user's social well-being. The consumption of a product is a social act with symbolic meanings, such as self-image expression (Kim et al. 2011). According to the elaboration on badges in Google Play Store, "Top Developer is a badge recognizing established, respected developers for their commitment to launching high-quality and innovative applications on Android". An application with this badge will attract users who care about quality of products and consider it as an important way for self-image expression. Therefore, we hypothesize that:

Hypothesis 3: The Top Developer badge has a positive impact on the number of downloads.

Cost Value

Cost value is the perceived utility of a product based on its cost associated with purchase and ownership (Smith and Colgate 2007). Literature on price sensitivity claims that consumers will tend to focus on the price when there is little other information available to differentiate products and generally lower price leads to higher favorability (Smith et al. 2001). In mobile application store, mobile users have limited information from other sources to distinguish applications. Therefore, price can be considered as an important factor that influencing a potential customer's downloading decision. Therefore, we propose that:

Hypothesis 4: Price has a negative effect on the number of downloads.

METHODOLOGY

Data Collection

In this study, we collected the in-store information about 500 communication applications from the top-grossing list in the Google Play Android Application Store. For each application, the collected data includes price, average rating, number of raters, number of screenshots, demonstration video, Top Developer badge, product description, and top 10 online customer reviews. The final sample size is 484 after removing non-English applications.

Text Mining

By applying Latent Semantic Analysis (LSA), which is a statistical approach to analyze relationships between a set of documents and their terms (Deerwester et al. 1990), the online review and product description were transformed into independent variables with interval data type. We use SAS Enterprise Miner 12.1 (SAS Institute Inc., Cary, NC, USA) to do text-mining analysis.

A customer review or a product description of an application was considered a document, and documents were then transformed into a list of terms. After the term parsing and the stemming steps, we removed the unique terms that only appear in one document and terms that do not add value to our analysis (auxiliary, conjunction, determiner, interjection, number, preposition, pronoun, and particle). Then the frequency of a term in each document was counted to generate a raw term-document matrix. We use a weighting and normalization scheme known as inverse document frequency weighting (TF-IDF) to transform the raw term-document matrix into a weighted matrix (Sidorova et al. 2008). The transformed term frequency matrix was then subjected to a singular value decomposition (SVD) (Sidorova et al. 2008). In this way, we got factor loading of terms and factor loading of documents. After factors were generated through text mining, latent semantics in each factor were interpreted. On a given factor, a set of terms with high-factor loadings was identified, creating a factor interpretation. On the given factor, each document had a factor loading that would be used in regression model in later steps.

In this study, we generated the factors by setting the number of factors q as 2, 3, 4, 5, and 10 separately. By comparing the results generated in different settings, we chose the best one, which was generated when q was 3. For both customer review and product description, when the number of factors is equal to 3, we can find the meaning of one factor according to the meaning of terms

in it. When q is equal to other numbers, the meaning of terms in any factor cannot be summarized to represent the meaning of a factor. In this way, three factors in online customer review were identified, representing three main aspects of communication applications that customers care about: application feature, connection, and cost. Three factors in product description were also identified, representing three main aspects of communication applications that developers introduce: shareness, voice recognition, and quality of connection.

Empirical Model Specification

Based on the research hypotheses presented above, we set up the following multiple regression model:

$$\text{Log (No. of download)} = \beta_0 + \beta_1\text{Price} + \beta_2\text{Rating} + \beta_3(\text{No. of rater}) + \beta_4\text{DemoVideo} + \beta_5(\text{No. of screenshot}) + \beta_6\text{TopDeveloper} + \beta_7\text{PD}_{\text{shareness}} + \beta_8\text{PD}_{\text{VoiceRecognition}} + \beta_9\text{PD}_{\text{QualityOfConnection}} + \beta_{10}\text{CR}_{\text{AppFeature}} + \beta_{11}\text{CR}_{\text{Connection}} + \beta_{12}\text{CR}_{\text{Cost}} \quad (1)$$

RESULT AND DISCUSSION

We evaluate the influence of in-store information on application downloads by considering the text and non-text information at the same time. The model fitting statistics indicate that our model becomes significant and gives a fairly good fit of the dependent variable in the regression. R Square indicates that although we considered lots of main in-store information in our model, which explained 37.5 percent of variation of the dependent variable, some information that mobile users would consider when they download applications was still not included, such as public information or a friend's recommendation.

Independent Variables	Standardized Coefficient
Price	$(0.025)^{***}$
Rating	0.153
No. of Raters	0.0008^{***}
No. of Screenshot	(0.003)
Demonstration Video	0.105
Top Developer	0.498
$\text{PD}_{\text{Shareness}}$	2.392^{***}
$\text{PD}_{\text{VoiceRecognition}}$	0.369
$\text{PD}_{\text{QualityOfConnection}}$	1.290
$\text{CR}_{\text{AppFeature}}$	1.633^{**}
$\text{CR}_{\text{Connection}}$	2.109^{**}
CR_{cost}	4.710^{***}

**: p-value<0.01; ***: p-value<0.001

Table 1. Parameter Estimates of Regression

The parameter estimates of regression are shown in Table 1. Among 12 independent variables, six become significant: price, number of raters, $\text{PD}_{\text{Shareness}}$, $\text{CR}_{\text{AppFeature}}$, $\text{CR}_{\text{Connection}}$, and CR_{cost}. First, price is negatively associated with the number of downloads. It suggests that in the mobile environment, where no information from other sources is provided, the high price will make users hesitate to download the application, especially when free alternatives are available. This result is consistent with previous research, showing that higher price leads to lower favorability (Smith et al., 2001).

Second, the number of raters is positively correlated with the number of downloads. According to previous study, as the popularity of product increases, sales of products will increase (Oberholzer-Gee and Strumpf, 2007). Duan (2005) also argued that in movie industry, the influence of word-of-mouth volume on product sales is positive, which means the product with more reviews and a higher number of ratings is more likely to have high sales volume. Our results show that the findings in previous study also work in the mobile environment.

Third, for communication applications, when the developers emphasize the shareness characteristics in the product description, the number of download will increase. The main function of communication applications is to enable people to communicate with each other and share information. Therefore, introducing this main feature will definitely help potential users have a clear understanding of the application and enhance their intention to download it. This finding is consistent with the one in a previous study that investigates the helpful information embedded in customer reviews for "News & Magazine" mobile applications (Wang and Song, 2015), which emphasizes the importance of making improvements on application features to increase the application downloads.

Fourth, our findings are consistent with previous studies (Duan et al., 2008), showing that online customer review influences the sales of applications. Three aspects of communication applications are addressed in customer review and all of them have influences on the number of downloads. It suggests that mobile users care about application features, application connection, and cost of usage. Therefore, mobile application developers should pay attention to these factors and make improvements to attract potential users.

There are six independent variables that are not significant: rating, number of screenshot, demonstration video, Top Developer badge, $\text{PD}_{\text{VoiceRecognition}}$, and $\text{PD}_{\text{QualityOfConnection}}$. First, the average rating has no influence on the application downloads. This finding is different with that in previous research, which indicates that the average rating is a salient factor influencing the sales of product (Ye et al. 2009). This difference is caused by the small variance of rating. Among 484 applications, 299 have a rating between 4.0 and 4.6. Therefore, it is hard for mobile users to distinguish applications by checking their average rating. Although we only analyze one category, no big difference existing among

applications in terms of average rating may be a common phenomenon across all mobile application categories. It is necessary to examine other categories to see whether the average rating, does not work in the mobile environment.

Second, neither number of screenshot nor demonstration video becomes significant, which indicates that vivid information does not work as desired in the mobile environment. Nisbett and Ross (1980) describe vivid information as being "likely to attract and hold our attention and to excite the imagination." Mobile users pay little attention to this kind of information when they download communication applications. It is because the users are likely to focus on the utility and functionality of communication applications, which are not quite related with exciting the imagination. However, this result may be different for other categories. For example, when people download game applications, their focuses are enjoyment and playfulness. The demonstration video and screenshots are likely to excite the imagination about how playful the game application is.

Third, the Top Developer badge is not significant, which indicates that obtaining Top developer badge is not helpful in increasing the number of downloads. The evaluation from the mobile application store is not an important indicator influencing customers' downloading decisions. However, this result cannot prove that the third-party's evaluation does not work in the entire mobile environment. In our dataset, among 484 communication applications, only 5 applications' developers have the Top developer badge, which is not a common case across all application categories. Therefore, it is necessary to compare applications within different categories to get a comprehensive understanding about how the platform's evaluation plays a role that influences application downloads.

Finally, regarding product description, two factors have no impact on the number of downloads. The voice recognition and the quality of connection are supplementary features, not main functions, provided by communication applications. Although application developers provide information about these two factors in product description, mobile users may not care about them. Therefore, these two factors have no influence on application downloads.

CONCLUSION AND LIMITATIONS

In this paper, we examine an important research question concerning product information in the mobile environment: what specific information influences download of applications? We applied customer value theory to support our arguments from a theoretical perspective and employed text-mining techniques to validate them. This study makes contributions in three aspects. First, this paper highlights the importance of detailed content in product description and online customer review. Our findings indicate that besides investigating the overall impact of text information on product sales, it is also necessary to find out specific content in product description and customer reviews. By revealing the meaningful information embedded in the texts, the relationship between in-store text information and customer's purchasing behavior becomes clearer and more understandable. Second, this paper reveals that the factors influencing a consumer's purchasing decisions in the mobile environment are different than those in the context of e-commerce. In the e-commerce environment, average rating usually works as an important factor that influences product sales. However, in the mobile environment, an average rating does not work in the expected way. Third, this study provides empirical implications that guide mobile application developers in improving the features of mobile applications, which may increase sales volume. It helps prevent mobile application companies from making vain efforts since our findings indicate that some information on applications has significant impact on the number of downloads, while some does not.

This study can be improved from three perspectives and its limitations also call for future research. First, our sample only included applications in one category. Although it is the second largest category in the Android platform and almost by the definition representative of the population of application downloads, there are still many differences between communication applications and other types of applications in terms of their functionality. People may focus on different aspects of information when downloading different types of applications. Therefore, the predictors of application downloads may vary among categories. The results of this study may not be generalizable to other types of applications. Collecting data from different categories and using a larger data set is necessary for future research to get a more robust result. Second, we did not consider the impact of tenure of applications on the volume of download. The volume of downloads for a new application is relatively lower than that for an old application. Therefore, tenure of an application should be controlled in future study. Third, there is some other information in the mobile application store have not been considered in this study, such as quality of screenshots. The information embedded in screenshots is also available for customers and may also influence the number of download. In future research, it is necessary to consider all available information at the same time to have a comprehensive understanding of the impact of in-store information on downloads.

REFERENCES

References will be provided upon request.