Asia Pacific Journal of Information Systems
Vol. 26 No. 4 (DECEMBER 2016), 577-592

ISSN 2288-5404 (Print) / ISSN 2288-6818 (Online)
http://dx.doi.org/10.14329/apjis.2016.26.4.577

Investigating the Value of Information in Mobile Commerce: A Text Mining Approach

Ying Wang[a], Miguel Aguirre-Urreta[b], Jaeki Song[c,*]

[a] *Ph.D. candidate, Information Systems and Quantitative Sciences, Rawls College of Business Administration, Texas Tech University, USA*
[b] *Assistant Professor, Information Systems and Quantitative Sciences, Rawls College of Business Administration, Texas Tech University, USA*
[c] *Professor, Information Systems and Quantitative Sciences, Rawls College of Business Administration, Texas Tech University, USA*

A B S T R A C T

The proliferation of mobile applications and the unique characteristics of the mobile environment have attracted significant research interest in understanding customers' purchasing behaviors in mobile commerce. In this study, we extend customer value theory by combining the predictors of product performance with customer value framework to investigate how in-store information creates value for customers and influences mobile application downloads. Using a data set collected from the Google Application Store, we find that customers value both text and non-text information when they make downloading decisions. We apply latent semantic analysis techniques to analyze customer reviews and product descriptions in the mobile application store and determine the embedded valuable information. Results show that, for mobile applications, price, number of raters, and helpful information in customer reviews and product descriptions significantly affect the number of downloads. Conversely, average rating does not work in the mobile environment. This study contributes to the literature by revealing the role of in-store information in mobile application downloads and by providing application developers with useful guidance about increasing application downloads by improving in-store information management.

Keywords: Mobile Application, Customer Value, Product Performance, In-store Information, Latent Semantic Analysis

I. Introduction

As mobile has overtaken fixed Internet access, mobile applications have increasingly permeated people's daily lives. This can be seen in the increase in mobile users, the frequent interaction between users and applications, and the increasing number of available applications in mobile application stores (Lella and Lipsman, 2014). In 2013, 91% of American adults owned cell phones, and 50% of phone owners used the devices to download applications (Duggan, 2013). As of July 2015, there are 1.5 million applica-

*Corresponding Author. E-mail: jaeki.song@ttu.edu Tel: 18068342024

tions available in the Apple Application Store, and over 100 million applications have been downloaded from it (Global mobile statistics, 2015). Among mobile applications, the number of downloads ranges from hundreds to hundreds of millions, which directly represents customers' purchasing decisions in application stores and has an influence on revenue generated from applications. Therefore, what factors cause this huge difference in customers' download behavior and application performance in the mobile environment becomes a critical problem worthy of investigation and discussion.

A large number of studies has been conducted to investigate salient factors affecting sales of products and customers' purchasing behaviors in the context of e-commerce (e.g., price, average rating) (Duan et al., 2008; Mudambi and Schuff, 2010). Few researchers have attempted to prove that such information is still valid in the mobile environment. Despite a previous study indicating that the major sources for organizations' competitive advantage likely come from customers (Woodruff, 1997), most firms do not have enough knowledge about how to compete on superior customer value delivery, especially in the mobile environment. Mobile customers' behaviors are different from those in the context of the web environment due to the characteristics of mobile devices as well as the mobile environment. Mobile devices' screen display limitations restrict mobile users' access to rich multimedia content (Pham et al., 2000). Due to the inconvenience of searching for multimedia content and the disability of displaying information from multiple resources synchronously, mobile device makes users rely on in-store information only to distinguish applications when they make downloading decisions. In addition, in the mobile environment, the application store provides a unified layout across all applications so that all applications demonstrate information in a uniformed structure. Therefore, the specific content of in-store information becomes critical in the mobile environment to make applications outstanding and attractive to potential users. To fill in the gap from both theoretical and practical perspectives, this study investigates which in-store information influences mobile application download and provides mobile application developers with empirical guidance on in-store information management.

In this study, we apply customer value theory to analyze the impact of in-store application information considered as predictors of application performance on the number of downloads for 500 applications in the communication category in the Google Play Android Application Store. We use latent semantic analysis approach to analyze text-related information, such as product description and online customer review, and apply the regression method to analyze the combined impact of text and non-text information on application downloads. In early studies about online review, the relationship between number of reviews and product sales was the primary issue (Chevalier and Mayzlin, 2006; Duan et al., 2008). In later studies, the discussion shifted to the detailed information embedded in the review, such as the level of helpfulness and abstractness (Liu et al., 2013). Therefore, we applied a text-mining technique to analyze the content of text information. Our study makes theoretical contributions by revealing a customer's purchasing mechanism in the mobile environment. We find that mobile users value certain detailed content about applications in text information, which indicates that, besides investigating the overall impact of text information on product sales, it is also necessary to determine if embedded specific messages have a deeper understanding of how customers make their purchasing decisions. We

also find that some information that has been widely accepted as salient factors affecting product sales in the context of e-commerce does not work in the mobile environment. Our study also makes 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.

Ⅱ. Conceptual Background

Customer value has been considered a critical predictor of customers' purchasing behaviors in e-commerce literature (Babin et al., 1994; Dodds et al., 1991; Kim et al., 2007). Customer value comes from different sources associated with products or the process of exchanging products, such as information about products, characteristics of products, interactions with people and environment during the transaction, and possession transfer (Smith and Colgate, 2007). Specifically, customer value coming from products is created by value-chain activities associated with product development (Smith and Colgate, 2007). Therefore, the predictors of product development success, such as product, strategy, process, and marketplace characteristics (Henard and Szymanski, 2001), could be the source of customer value, influencing customers' purchasing behaviors. In this study, we identify predictors of product performance and merge them into a customer value framework. We propose that in-store information describing a mobile application's characteristics is the source of customer value and influences mobile users' downloading behaviors.

2.1. Customer Value Theory

Customer value has been defined differently in existing literature. In this study, we consider 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" (Woodruff, 1997). Based on this definition, numerous studies have been conducted to identify key precursors of customer value and propose a related framework to demonstrate that perceived customer value has an influence on an individual's purchasing behavior (Chen and Dubinsky, 2003; Schechter, 1984).

Among different conceptual frameworks, 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 (Smith and Colgate, 2007). These four types of value may come from different sources, which cover five aspects: information, products, interactions, environment, and possession transfer. First, the customer value coming from information about products can help customers realize performance and outcomes, make associations, and interpret meaning. It is important for customers to have a correct understanding of products and realistic expectations when evaluating them. Second, products directly provide features, functions, and characteristics that allow per-

formances and outcomes, which generate value by enabling customers to have sensory and emotional feelings towards products. Third, the activities associated with the processes of purchasing and consuming products also generate values for customers; in this study, we mainly consider the value coming from products due to the unique characteristics of the mobile environment. In the mobile environment, there is limited information available from multimedia and very few associated activities. Therefore, for mobile application users, customer value mainly comes from the characteristics of applications. The in-store information that reflects application characteristics is the source of customer value. We use customer value framework to understand how the valuable in-store information about applications serves as a predictor of application performance.

2.2. Predictors of Product Performance

Due to the increasingly important role of product innovation in sustainable business success, many studies have been conducted to explore the drivers of product development success (Cooper and Kleinschmidt, 1987; Montoya-Weiss and Calantone, 1994). Although the drivers are different among related research, researchers agree that these drivers can be commonly classified into four categories: product, strategy, process, and marketplace (Henard and Szymanski, 2001). Product characteristics refer to elements pertaining to a product or service, such as price and features. Strategy characteristics are a firm's planned actions that have the potential to provide competitive advantage in the marketplace, such as obtaining icons and badges valued by the market. Process characteristics refer to the elements associated with the new product development process and its execution. Finally, marketplace characteristics are the elements that describe the target market and include market potential, competitive activity, and the intensity of that activity in response to new product introductions. 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 consider only the performance drivers within the first two categories: product characteristics and strategy characteristics. In this study, an application performance is a function of product characteristics and strategy characteristics, which are the sources of customer value.

Ⅲ. Hypothesis Development

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 influencing application downloads.

3.1. Functional Value

Functional value, considered a key influence on consumer choice (Kim et al., 2007; Sweeney and Soutar, 2001), refers to the extent to which a product or service has desired characteristics or performs a desired function (Smith and Colgate, 2007). The functional value of a product comes from its characteristics or attributes, which can be measured in functional terms like quality (Mazid, 2012). In this study, we use functional quality to indicate functional value, which is the perceived overall excellence and expected performance of an application.

In the digital world, consumers have various ways to express their opinions on products, such as giving ratings and writing online reviews, which address their perceived functional quality of products. Previous studies show that use-generated content such as online rating and review are salient factors in e-commerce, having a significant influence on product sales and consumer decision-making (Ye et al., 2009; Dellarocas, 2003; Duan et al., 2008). By analyzing the information generated by online users, organizations can have a deep understanding of consumers' perceptions of products and identify consumers' purchasing behaviors (Chin et al., 2015). Additionally, the distribution of word-of-mouth increases consumer awareness and reflects the popularity of a product (Liu, 2006). Hanson and Putler (1996) demonstrated that consumers use the relative popularity of products as an indication of both the quality and the appropriateness of the products when making their choices. Number of raters demonstrates consumer awareness and reflects the popularity of an application, which is another indicator of consumers' perceptions of products and influences product sales. For example, the popularity of an album is positively associated with its sales (Oberholzer-Gee and Strumpf, 2007). Therefore, we hypothesize that

H1a: An average rating has a positive impact on the number of downloads.

H1b: Helpful information in online customer reviews has a positive impact on the number of downloads.

H1c: The number of raters has a positive impact on the number of downloads.

3.2. Experiential Value

In the framework proposed by Smith and Colgate (2007), experiential value refers to the extent to which a product or service creates appropriate experiences, feelings, and emotions for the customer. This value depends principally on how the product looks and how it relates to customers. In the mobile application store, to attract potential users and impress them, developers have the option of uploading a demonstration video and screenshots to provide vivid descriptions 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 hold a general assumption that increasing the vividness of a message enhances its persuasiveness (Appiah, 2006). Video-based information enhances message effectiveness and makes the message more attractive, vivid, and salient. Xu et al. (2015) argue that dynamic and moving videos can capture human attention by vividly presenting stimuli for different processing channels like hearing and seeing to enhance cognition, which is also associated with feelings and experience. Therefore, uploading a demonstration video and screenshots is an effective strategy to attract attention, provide appropriate feelings, and increase number of downloads.

Besides vivid information, 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 descriptions to obtain helpful information about various aspects of applications, such as functions, operations, privacy policy, etc. Therefore, we hypothesize that

H2a: Demonstration video has a positive impact on the number of downloads.

H2b: Helpful information in product descriptions has a positive impact on the number of downloads.

H2c: Number of screenshots has a positive impact on number of downloads.

3.3. Symbolic Value

Symbolic value refers to the extent to which customers attach psychological meaning to a product, which helps customers enhance self-concepts (Smith and Colgate, 2007). For example, purchasing goods with a "recycle" icon on the package is a way to express a person's self-concepts and self-worth related to environment protection, which makes them feel good about themselves. The top developer badge indicates the application developer has the qualification of providing qualified applications. The privacy policy shows customers how their information will be collected and used. The "good quality" and "privacy protection" icons make people who care about these aspects feel good about themselves by downloading applications with these two indicators. Therefore, we argue that people who care about quality and privacy are more likely to download applications with top developer badges and privacy policies, which increases download volume. Therefore, we hypothesize that

H3: The Top Developer badge has a positive impact on the number of downloads.

3.4. Cost Value

Cost value refers to the perceived utility of a product based on the transaction costs involved in its purchase, ownership, and use (Smith and Colgate, 2007). In e-commerce literature, it has been verified that price and price elasticity are important factors influencing customers' purchasing behaviors in the context of the web environment (Gu et al., 2013). The literature on price sensitivity also claims that consumers tend to focus on price when there is little other information available to distinguish products and favor products with lower prices (Dodds et al., 1991). In the mobile application store, mobile users have limited information from other sources to distinguish applications due to the mobile devices' displaying limitations and interacting difficulty (Froehlich et al., 2007). Therefore, price can be considered an important factor influencing potential customers' downloading decisions. Therefore, we propose that

H4: Price has a negative effect on number of downloads.

Ⅳ. Methodology

4.1. Data Collection

In this study, we collected data from the Google Play Android Application Store, which dominates the application market in terms of number of available applications and number of users. The millions of mobile applications in the store are classified into twenty-seven categories, such as Games, Books & Reference, Businesses, etc. Within each category, applications are ranked based on a combination of criteria such as ratings, number of downloads, download growth, uninstalls, and usage. We chose to analyze the "Communication" category in this study because it is the second most popular category ranked by download volume, which attracts attention from both users and developers on the Android platform. Additionally, compared with the "Game" category, which has the largest proportion of downloads, the

<Table 1> Description and Summary of Statistics on Independent Variables

Variables	Description	Mean	Std.dev.	Min	Max
Price	The charged price when installing the application; It is 0 when the application is free;	3.31	12.17	0.00	190.00
Rating	The listed rating value of application in application store;	4.03	0.52	1.00	5.00
No. of raters	The number of people who evaluated the application;	65,882.67	515,365.87	0.00	6,998,788.00
DemoVideo	Whether application developer provided demonstration video; Dummy variable: 1 means the developer provided demo video and 0 otherwise;	0.28	0.448	0.00	1.00
No. of screenshot	The number of screenshots provided by application developer;	6.63	3.57	0.00	24.00
TopDeveloper	Whether application developer obtained Top Developer badge from application store;	0.01	0.10	0.00	1.00
$PD_{Shareness}$	The loading of shareness factor generated from text analysis on product description;	0.12	0.10	0.01	0.59
$PD_{VoiceRecognition}$	The loading of voice recognition factor generated from text analysis on product description;	0.06	0.13	0.00	0.99
$PD_{QualityOfConnection}$	The loading of quality of connection factor generated from text analysis on product description;	0.08	0.08	0.09	0.38
$CR_{AppFeature}$	The loading of application feature factor generated from text analysis on customer review;	0.13	0.09	0.10	0.46
$CR_{Connection}$	The loading of connection factor generated from text analysis on customer review;	0.09	0.09	0.06	0.57
CR_{cost}	The loading of cost factor generated from text analysis on customer review;	0.12	0.09	0.19	0.56

content of applications in the "Communication" category is similar, which enables us to focus on analyzing the impact coming from the features of applications and minimizing the impact of differences coming from the preference of customers.

We collected in-store information about the top 500 communication applications from the top-grossing list, which includes both free and paid applications and ranks them according to their generated revenue. For each application, the collected data includes price, average rating, number of raters, number of screenshots, demonstration video, Top Developer badge, product description, and online customer reviews. The description and summary statistics of independent variables are listed in <Table 1>. We removed 5 non-English applications and 11 outliers with extremely high or low numbers of downloads. Therefore, the final sample size is 484.

4.2. Text Mining

The independent variables, online review, and product description were in text data type and could not be directly used in our empirical model to investigate their effects of application downloads. By applying Latent Semantic Analysis (LSA), 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. SAS Enterprise Miner 12.1 (SAS Institute Inc., Cary, NC, USA) was used to conduct text mining.

A customer review or a product description of an application was considered a document, and documents were then transformed into a list of terms. We first removed all non-English words, punctuation, and numbers. The terms with parts of speech, such as conjunction, interjection, preposition, pronoun, determiner, auxiliary, and particle, had limited information and were eliminated from the dataset. Next, we removed some words that appear frequently but do not add any value to our further analysis (e.g., application, app). After all pre-processing steps, we conducted stemming, which refers to finding the root of each term and marking terms having the same root as one. Then, we counted term frequency and removed terms that appeared only once in the dataset. Based on remaining terms, we created a raw term-document matrix with dimensions of n x m (n was the number of terms, and m was the number of documents).

For a term, the value in the raw term-document matrix demonstrates how important the term is in both a certain document and the whole dataset. For example, a reviewer's language habit might have caused a term to uncommonly occur in a document many times and therefore bias the effective information. Thus, the raw term-document matrix should be weighted to dampen the biases brought by the terms having high frequency.

Let f_{ij} be the ij^{th} frequency in the raw term-by-document matrix, or the frequency of term i in document j. And then the weighted frequency of element f_{ij} in the term-by-document frequency matrix is determined by

$$W_{ij} = \left[log_2\left(f_{ij}+1\right)\right]\left[1+\Sigma_j \frac{\left(\frac{f_{ij}}{g_i}\right)log_2\left(\frac{f_{ij}}{g_i}\right)}{log_2(m)}\right] \quad [1]$$

where g_i is the frequency of term i in all documents, and m is the number of documents (Dumais, 1991).

After term weighting, the raw term-document matrix was converted into a weighted term-document matrix with the same dimension of n x m, which is a huge matrix and requires high computation capability. To reduce the complexity of the weighted term-document matrix, but at the same time keep as much information as possible, we applied the Singular Value Decomposition method (SVD) to reduce noise in the matrix. The SVD method decomposes W into three new matrices－T, S, and D, such that $W=TSD^T$, where T is orthonormal matrix with dimensions n x r, D^T is orthonormal matrix with dimensions r x m, and S is a diagonal matrix of singular values with dimensions r x r and all numbers on the diagonal are positive. Row vectors in D represent documents, and similarities between documents are obtained by calculating differences between rows in the matrix DS. Similarly, row vectors in T represent terms, and similarities between terms are obtained by calculating differences between rows in the matrix TS.

Matrices D, T, and S are truncated in such a way that only the first q column remains. The truncated matrices D and T specify the positions of documents and terms in a q-dimensional space. Coordinates of the q-dimensional space can be considered as a factor, and the coordinate system is rotated to make the positions of documents and terms close to the new coordinates. The coordinates of the rotated q-dimensional space are defined as new factors, or topics. The factor loadings of documents for the rth factor ($r=1,2\cdots q$) are values in the rth column of matrix DS. Similarly, the factor loadings of terms for the

r^{th} factor (r=1,2⋯q) are values in the rth column of matrix TS.

After factors were generated through text mining, we interpreted latent semantics in each factor. For a given factor, a set of terms with high factor-loadings was identified, generating a factor interpretation. For each factor, each document had a factor loading that would be used to investigate the effect of the factor on application download.

We used text mining to analyze online reviews and product descriptions separately. We generated 500 documents since we collected information from about 500 applications. However, some applications do not have product descriptions or customer reviews. After the text parsing and filtering, online reviews had 484 documents and 1,186 terms, and product descriptions had 494 documents and 1,169 terms. In this study, we generated the factors by setting the number of factors q as 2, 3, 4, 5, and 10 separately. A higher number of factors would lead to a lower level of abstractness in the results of text analysis, which may help to find meaningful factors that represent the content embedded in text information. However, a higher number of factors may also increase the similarity between factors, making several factors have similar interpretations, especially when the sample size is small. Based on the settings in previous studies, we argued that it was not necessary to set q as a number higher than 10 since our sample size was relatively small.

By comparing the results generated in different settings, we chose the one that represents semantic meaning in a better way and used it in regression analysis. When the number of factors is equal to 3, we can clearly identify the factors according to high-loading terms in each factor. When we chose other numbers, either terms were cross-loaded or it was hard to identify factors. Therefore, three factors in online customer review were identified, representing three main aspects of communication applications that customers care about. The first factor is application feature, which is the way an application tells the user what capabilities it exhibits to function properly (Gary, 2012). For communication applications, the main capability is enabling people to contact each other, such as dialing and texting others, group calling, and receiving notifications. The second factor is connection, which refers to the application's performance in terms of server connection, compatibility with platform, and Internet connection. The terms in this factor describe the connection performance, such as browser, connect, vpn, android, and wifi. The third factor is cost, which refers to how much users must pay when installing or using the application. Users care about whether the application is free, how much they will be charged, how to earn credit when using it, whether they will waste their money, etc.

In the same way, 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. First, for a communication application, the main function that it provides is enabling people to contact each other, sharing information between individuals or among groups. Therefore, chatting with friends, sending messages, and making national or international phone calls should be the focus addressed in the product description. Second, one type of communication application is designed to provide smooth call flow, which means calling with no wait and no confusion (Spoken Communication, 2014). Another type is to transform text messages into voice messages or vice versa. To provide a clear introduction about applications, the voice recognition function should be ad-

<Table 2> Terms in Each Factor

	Factors	Terms in Factor
Online Customer Review	CR_Applicaiton Feature	contact, message, app., group, text, sms, dialer, notification, widget, great
	CR_Connection	work, browser, connect, version, vpn, android, great, update, support, wifi
	CR_Cost	credit, text, number, voice, free, money, pay, waste, earn, charge
Product Description	PD_Shareness	number, phone, free, message, friend, text, international, user, group, chat
	PD_Voice Recognition	voice, pronunciation, correction, aloud, language, read, navigation, speak, read, run
	PD_Quality of Connection	device, contact, version, server, mode, connection, network, access, pc, widget

dressed thoroughly by application providers, such as pronunciation correction and multiple language support. Third, in the mobile environment, the quality of connection is important, especially for communication applications. Without having access to the network or servers, most of them will not provide main functions, such as contacting others via mobile devices. Therefore, it is necessary to emphasize the quality of connection in the product description.

These factors containing information about online reviews and product description were used as independent variables in the further analysis. The top 10 terms with high factor-loading in each factor are listed in <Table 2>.

4.3. Empirical Model Specification

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

$$\begin{aligned} Log(No.of\ download) = {} & \beta_0 + \beta_1 Price + \beta_2 Rating + \\ & \beta_3(No.of\ rater) + \beta_4 DemoVideo + \\ & \beta_5(No.of\ screenshot) + \beta_6 TopDeveloper + \\ & \beta_7 PD_{Shareness} + \beta_8 PD_{VoiceRecognition} + \\ & \beta_9 PD_{QualityOfConnection} + \beta_{10} CR_{AppFeature} + \\ & \beta_{11} CR_{Connection} + \beta_{12} CR_{Cost} \end{aligned} \quad [2]$$

Since the number of downloads ranges from 1,000 to 500,000,000 with a high skewness, the dependent variable is the log-transformed number of downloads. After taking log-transformation, the dependent variable is normally distributed.

Ⅴ. Results and Discussion

We evaluated 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 the variation in the dependent variable, some information that mobile users would consider when they download applications was still not included. For example, although we counted on vivid information such as demonstration video and screenshots, we did not analyze the quality of such information. In addition, for some mobile users, their motives for downloading certain applications may not be because of in-store information, but information from other sources, such as public information or friends' recommendations.

The parameter estimates of regression are shown

<Table 3> Parameter Estimates of Regression

Variables	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
PDShareness	2.392***
PDVoiceRecognition	0.369
PDQualityOfConnection	1.290
CRAppFeature	1.633**
CRConnection	2.109**
CRcost	4.710***

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

in <Table 3>. Among 12 independent variables, 6 become significant: price, number of raters, $PD_{Shareness}$, $CR_{AppFeature}$, $CR_{Connection}$, 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, high price will make users hesitate to download applications, especially when free alternatives are available. This result confirms that higher price leads to lower favorability (Smith et al., 2001).

Second, number of raters positively correlates with number of downloads. According to a previous study, as the popularity of a product increases, its sales will also increase (Oberholzer-Gee and Strumpf, 2007). Duan (2005) argued that, in the 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. High volume of reviews generates more conversation and increases the awareness of applications.

Third, for communication applications, when developers emphasize shareness characteristics in product descriptions, the number of downloads increases. The importance of making improvements on applications features to increase application downloads is also emphasized in a related study (Wang and Song, 2015), indicating that it is critical to highlight key features in product descriptions. Because the main function of communication applications is to enable people to communicate with each other and share information, introducing this main feature definitely helps potential users have a clear understanding of applications and enhances their intentions to download them.

Fourth, the results show customers value other people's opinions when making downloading decisions. Three main aspects of communication applications are addressed in customer reviews, and all of them have influences on number of downloads. This 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 screenshots, demonstration video, Top Developer, $PD_{VoiceRecognition}$, and $PD_{QualityOfConnection}$.

First, average rating has no influence on application downloads. This finding is different from that in previous research, which indicates that average rating is a salient factor influencing product sales (Ye et al., 2009). This difference might be caused by the small variance of rating. Among 484 applications, 299 have a rating between 4.0 and 4.6. It is hard for mobile users to distinguish applications purely by looking at numeric ratings. Narrow range of average rating is a common phenomenon across all categories in mobile application stores, especially for popular applications. Therefore, although average rating is a salient factor influencing product sales in e-commerce, it might not have influence on application downloads in the mobile environment.

Second, neither number of screenshots nor demonstration video becomes significant, which indicates that vivid information does not work as desired. Vivid information is "likely to attract and hold our attention and to excite the imagination" (Nisbett and Ross, 1980). However, mobile users pay little attention to this kind of information when they download communication applications, perhaps because users are likely to focus on the utility and functionality of communication applications, which are not likely related to exciting the imagination. However, this result may be different for other categories. For example, when people download game applications, their motivations are enjoyment and playfulness. Vivid information like screenshots and demonstration video exhibits playful aspects about applications, providing customers with experience of how playful the game is. Therefore, for game applications, screenshots and demonstration video are more likely to influence users' purchase decisions.

Third, the coefficient of the Top Developer badge is not significant. This indicates that obtaining a Top Developer badge is not helpful in increasing the number of downloads for communication applications. In our dataset, among 484 applications, only 5 have the Top Developer badge, demonstrating that this badge has not been widely recognized by customers when they download communication applications. However, it is not a common case across all application categories. Therefore, it is necessary to compare applications within different categories to develop a comprehensive understanding of how the evaluation from application platforms (e.g., Google Application Store) influences application downloads.

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

Ⅵ. Conclusion and Limitations

In this paper, we examine an important research question concerning product information in the mobile environment: what in-store information influences mobile application downloads? Our data included both text and non-text information. We applied customer value theory to support our arguments from a theoretical perspective and employed text-mining techniques to validate them. A number of theoretical and practical implications can be derived from this

study.

This study makes contributions in three ways. 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. For example, mobile users care about the main features of applications and the costs associated with downloads when making purchasing decisions. Cost does not only refer to monetary cost, but also may refer to time cost and psychological cost. Therefore, to increase number of downloads, communication application developers need to address network connection, shareness, and cost in text information clearly. By revealing this meaningful information embedded in the texts, the relationship between in-store text information and customers' purchasing behaviors becomes clearer and more understandable.

Second, this paper reveals that the factors influencing consumers' purchasing decisions in the mobile environment are different from 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, average rating does not work as expected. The small range of ratings among applications could be explained by one or multiple reasons. For example, if most of applications are free, customers might not care about the differences between applications and ignore average rating since the downloading behavior does not cost them anything. Therefore, comparing several possible reasons and determining the most fundamental of them calls for further investigation.

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 number of downloads, while some does not.

This study can be improved from several perspectives, and its limitations call for future research. First, our sample included applications in only one category. Although it is the second largest category in the Android platform and almost, by definition, representative of the population of application downloads, there are still many differences between communication applications and other types of applications regarding functionality. People may focus on different aspects of information when downloading different types of applications. Therefore, predictors of application downloads may vary among categories. Additionally, 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 to get more robust results are necessary for future research. 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, the tenure of an application should be controlled in future study. Third, other information in the mobile application store, such as quality of screenshots, was not considered in this study. The information embedded in screenshots is also available for customers and may influence number of downloads. 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>

[1] Anderson, J., and Rainie, L. (2012). *The Future of Apps and Web*. Pew Research Center, http://www.pewinternet.org/2012/03/23/the-future-of-apps-and-web/

[2] Appiah, O. (2006). Rich Media, Poor Media: The Impact of Audio/Video vs. Text/Picture Testimonial Ads on Browsers' Evaluation of Commercial Web Sites and Online Products. *Journal of Current Issues & Research in Advertising (CTC Press), 28*(1), 73-86.

[3] Babin, B. J., Darden, W. R., and Griffin, M. (1994). Work and/or Fun: Measuring Hedonic and Utilitarian Shopping Value. *Journal of Consumer Research, 20*, 644-656.

[4] Bohmer, M., Hecht, B., Schoning, J., Kruger, A. and Bauer, G. (2011). Falling Asleep with Angry Birds, Facebook and Kindle: A Large Scale Study on Mobile Application Usage. in *Processing of the 13th International Conference on Human-Computer Interaction with Mobile Devices and Services*, New York, NY, 47-56.

[5] Chen, Z., and Dubinsky, A. J. (2003). A Conceptual Model of Perceived Customer Value in E-Commerce: A Preliminary Investigation. *Psychology & Marketing, 20*(4), 323-347.

[6] Cooper, R. G., and Kleinschmidt, E. J. (1987). New Products: What Separates Winners form Losers? *Journal of Products Innovation Management, 4*(4), 169-184.

[7] Dodds, W. B., Monroe, K. B., and Grewal, D. (1991). The Effects of Price, Brand and Store Information on Buyers' Product Evaluations. *Journal of Marketing Research, 28*(3), 307-319.

[8] Dellarocas, C. (2003). The digitization of word of mouth: promise and challenges of online feedback mechanisms. *Management Science, 49*(10), 1407-1424.

[9] Dhar, V. and Chang, E. (2009). Does Chatter Matter? The Impact of User-Generated Content on Music Sales. *Journal of Interactive Marketing, 23*(4), 300-307.

[10] Duan, W., Gu, B., and Whinston, A. B. (2008). The Dynamics of Online Word-of-Mouth and Product Sales – An Empirical Investigation of the Movie Industry. *Journal of Retailing, 84*(2), 233-242.

[11] Duggan, M. (2013). *Cell Phone Activities 2013*. http://www.pewinternet.org/2013/09/19/cell-phone-activities-2013/, Pew Research Center.

[12] Dumais, S. T. (1991). Improving the retrieval of information from external sources. *Behavior Research Methods, Instruments, & Computers, 23*(2), 229-236.

[13] Forman, C., Ghose, A. and Wiesenfeld, B. (2008). Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identify Disclosure in Electronic Markets. *Information System Research, 19*(3), 291-313.

[14] Froehlich, J., Chen, M. Y., Consolvo, S., Harrison, B., and Landay, J. A. (2007). MyExperience: A System for in situ Tracing and Capturing of User Feedback on Mobile Phones. *Proceedings of the 5th International Conference on Mobile Systems, Applications and Services*, 57-70.

[15] Gary. (2012). *Understanding Application Features*. https://community.jivesoftware.com/docs/DOC-68850, Jive community.

[16] Global Mobile Statistics (2014). *Home: All the Latest Stats on Mobile Web, Apps, Marketing, Advertising, Subscribers, and Trends*… http://mobiforge.com/research-analysis/global-mobile-statistics-2014-home-all-latest-stats-mobile-web-apps-marketing-advertising-subscriber

[17] Henard, D. H. and Szymanski, D. M. (2001). Why Some New Products Are More Successful Than Others. *Journal of Marketing Research, 38*(3), 362-375.

[18] Liu, J., Sarkar, M. K. and Chakraborty, G. (2013). Feature-based Sentiment Analysis on Android App Reviews Using SAS Text Miner and SAS Sentiment Analysis Studio. *SAS Global Forum*, paper number. 250.

[19] Kim, H. W., Chan, H. C., and Gupta, S. (2007). Value-based Adoption of Mobile Internet: An Empirical Investigation. *Decision Support Systems,*

43(1), 111-126.

[20] Kim, H, W., Gupta, S. and Koh, J. (2011). Investigating the Intention to Purchase Digital Items in Social Networking Communities: A Customer Value Perspective. *Information & Management, 48*, 228-234.

[21] Montoya-Weiss, M. M. and Calantone, R. (1994). Determinants of New Product Performance: A Review and Meta-Analysis. *Journal of Product Innovation Management, 11*(11), 397-417.

[22] Moe, W. W. and Schweidel, D. A. (2012). Online Product Opinion: Incidence, Evaluation, and Evolution. *Marketing Science, 31*(3), 372-386.

[23] Mudambi, S. M. and Schuff, D. (2010). What Makes A Helpful Review? A Study of Customer Reviews on Amazon.com. *MIS Quarterly, 34*(1), 185-200.

[24] Oberholzer-Gee, F. and Strumpf, K. (2007). The Effect of File Sharing on Record Sales: An Empirical Analysis. *Journal of Political Economy, 115*(1), 1-42.

[25] Pham, T., Schneider, G. and Goose, S. (2000). A situated computing framework for mobile and ubiquitous multimedia access using small screen and composite devices. *Multimedia 00' Proceedings of the eighth ACM international conference on Multimedia*, 323-331.

[26] Shaffer, G. and Zettelmeyer, F. (2002). When Good News About Your Rival Is Good for You: The Effect of Third-Party Information on the Division of Channel Profits. *Marketing Science, 21*(3), 273-293.

[27] Schechter, L. (1984). *A normative conception of value*. Progressive Grocer, 12-14.

[28] Schlosser, A. E. (2005). Posting versus Lurking: Communicating in A Multiple Audience Context. *Journal of Consumer Research, 32*, 260-265.

[29] Schlosser, A. E. (2011). Can Including Pros and Cons Increase the Helpfulness and Persuasiveness of Online Reviews? The Interactive Effects of Rating and Arguments. *Journal of Consumer Psychology, 21*(3), 226-239.

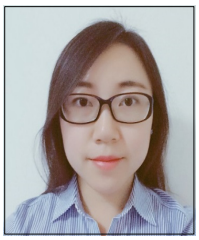
[30] Smith, J. B. and Colgate, M. (2007). Customer Value Creation: A Practical Framework. *Journal of Marketing Theory and Practice, 15*(1), pp.7-23.

[31] Wang, Y. and Song, J. (2015). The Impact of Mobile Application Information on Application Download: A Text Mining Approach. *American Conference on Information System*, 2015 proceedings.

[32] Woodruff, R. (1997). Customer Value: The Next Source for Competitive Advantage. *Journal of the Academy of Marketing Science, 25*(2), 139-153.

[33] Ye, Q., Law, R. and Gu, B. (2009). The Impact of Online User Reviews on Hotel Room Sales. *International Journal of Hospitality Management, 28*(1), 180-182.

◆ About the Authors ◆



Ying Wang

Ying Wang is currently a Ph.D. candidate of Management Information Systems in Information Systems and Quantitative Sciences at the Rawls College of Business at Texas Tech University. Before she joined the Ph.D. program in 2012, Ying obtained an M.B.A in general business from Texas Tech University, a B.S. in Chemical Engineering from Wuhan University of Technology, and a B.A. in English (dual degree) from Huazhong University of Science and Technology. She also had three years of work experience in credit risk assessment in HSBC Bank (China). Her research interests primarily focus on business analytics, adoption and diffusion of smart technology, and text analytics techniques. Her dissertation focuses on investigating the effects of online reviews on product sales by considering the structure of reviews and the effect of time. Her work has been presented at national and international conferences, such as Americas Conference on Information Systems, Association for Information Systems - Special Interest Group on Human Computer Interaction, IEEE International Conference on Service Systems and Service Management, and Big XII+ MIS Symposium.

Miguel Aguirre-Urreta

Miguel Aguirre-Urreta is an Assistant Professor in the Area of Information Systems and Quantitative Sciences, Jerry S. Rawls College of Business Administration, at Texas Tech University. Before joining TTU, he was on the faculty at DePaul University in Chicago. Miguel received his Ph.D. in Information Systems from the University of Kansas, and his MBA in Information Systems and Finance from the Kelley School of Business at Indiana University. His undergraduate degree is in public accounting from the Universidad de Buenos Aires, in Argentina. Miguel is interested in quantitative research methods, computer self-efficacy, human-computer interaction, technology acceptance and diffusion, and business analytics. His research has appeared or is forthcoming at *MIS Quarterly*, *Information Systems Research*, Communications of the AIS, *Research Synthesis Methods*, *IEEE Software*, and the *DATA BASE for Advances in Information Systems*, among others. Miguel is the Past Chair of AIS SIGHCI.

Jaeki Song

Jaeki Song is the Jerry S. Rawls Endowed Professor of MIS at the Rawls College of Business at Texas Tech University. His primary area of research is the innovations in Information Systems. Specific research issues include IT service innovations, diffusion of information technologies, social and economic consequences of information technologies, and business analytics. His research findings have appeared more than 50 refereed journal papers in *Management Science, Journal of Management Information Systems, Journal of the Association for Information Systems, IEEE Transactions on Professional Communication, Journal of Strategic Information Systems, Decision Support Systems, Information & Management, and other journals*.

Submitted: November 2, 2016; 1st Revision: December 31, 2016; Accepted: December 31, 2016