The Impact of Mobile Application Information on Application Download: A Text Mining Approach

Full papers

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

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

Abstract

The effects of customers' reviews on products have been understood in the initial digital purchase context. This study extends this literature by exploring the reviews in the mobile environment. It calls for understanding the context of reviews which types of information in the review have significant impact on customers' behavior. This study applied text mining in analyzing customers' reviews in purchasing mobile applications and find out important applications' features valuable for customers by discovering meaningful information in influential words. We find that for mobile applications, specific expressions in online customer review, companies' reply to the review, and quality certifications from application store have significant impacts on the number of download. Theoretical and practical implications are discussed.

Keywords

mobile application; customers' review; text mining; latent semantic analysis

Introduction

As mobile phones have evolved from single-purpose communication devices into smart tools that provide their users with various functions like entertainment and navigation, the number of available applications for mobile phones has been steadily increasing (Bohmer et al. 2011). As of July 2014, there are 1.3 million applications available in Google Play Store and 1.2 million applications available in Apple Application Store (Global mobile statistics 2014). With intensive competition among mobile applications that provide similar functions, how to attract prospective customers to make purchase decisions becomes a challenging issue for mobile application companies. A large number of studies have been conducted to understand what factors affect sales of products in the context of e-commerce. Among the many influencing factors, available product information has been considered an important resource for customers due to its critical ability to demonstrate product quality and help customers make more efficient and rational purchase decisions when they are faced with information asymmetry (Mudambi and Schuff 2010; Cao et al. 2011).

However, the impact of product information is not independent from its specific context. For example, Dhar and Chang (2009) found that, concerning music albums, the volume of blog posts (no further distinction between positive and negative posts) about an album is positively correlated with future album sales. Forman et al. (2008) revealed that online customer reviews play different roles under different contexts. Positive reviews help consumers make purchases they will value while negative reviews help consumers avoid purchases they may regret. Few studies on the impact of product information have been conducted in the context of purchasing mobile applications. Therefore, although it is easy for mobile application companies to collect information about the usage of applications, such as rating, number of rater, and customer reviews, it is still hard for them to figure out the relationship between information on an application and its sales. In this situation, research studies that investigate what specific application information has impacts on sales of mobile application are desired in both theoretical and practical fields.

The common information about applications in the application store can be classified into two types: text and non-text. Text information refers to customer reviews and product description; non-text information includes the average customer rating, the number of download, the number of raters, etc. Based on the results of a recent behavior survey, 98 percent of online shoppers consider online customer reviews an important factor that influences their purchase decisions (Liu et al. 2013). Google Play Store also indicates that prospective users consider ratings and reviews as key benchmarks indicating the quality of applications. Therefore, both text and non-text information is important for customers when they decide to download mobile applications and should be considered when analyzing the impact of application information on its number of download.

Although a large number of studies have been conducted to examine the impact of product information on product sales (Forman et al. 2008) and explore determinants of its features like "helpfulness" (Cao et al. 2011), few researches investigate what specific words in reviews have influence on the sales of products and how to determine customers' needs by analyzing customer reviews. The purpose of this study is to empirically investigate what kind of information about mobile application has impacts on sales of applications, or number of download, as well as to find out what specific words in the reviews have influences on the number of download. This study does not only make theoretical contributions by filling the theoretical gaps in research, but also make practical contributions by providing mobile application companies with useful suggestions for improving the applications' number of download.

The rest of this paper is organized as follows: First, we briefly discuss the relevant literature on text product information and non-text information. Next, we describe the data and develop our research method. Then, we present the empirical results and related discussion. Finally, we discuss limitations and areas for future research.

Literature Review

Text Information – online customer reviews

Online customer reviews are considered as digitalized word of mouth (Dellarocas. 2003), having influences on product sales and consumer decision-making (Duan et al. 2008). The primary issue addressed in early studies on online customer reviews was discussing the relationship between online user review valence/volume and product sales. By analyzing a dataset in book websites, Chevalier and Mayzlin (2006) empirically proved that improvements in reviews led to increases in relative sales. They also asserted that the impact of extremely negative reviews is greater than that of extremely positive reviews. Additionally, Duan et al. (2008) found that in the movie industry, valence of online customer reviews significantly influence volume of reviews, and volume of reviews, in turn, leads to higher box office performance.

Later studies on online customer reviews pay more attention to the detailed text information generated in reviews; thus, content analysis is commonly used to quantify the feedback text comments in the e-commerce environment (Pavlou and Dimoka 2006). Therefore, text mining becomes popular in IS research, and different text mining techniques are developed to investigate content embedded in online customer reviews (Larsen et al. 2008), such as latent semantic analysis. The primary issue becomes investigating characteristics of online consumer reviews, such as helpfulness of the review. One study shows that a customer-written product review with a low level of content abstractness yields the highest-perceived review helpfulness (Li et al. 2013).

Although the relationship between online customer reviews and product sales has often been discussed in early studies, most of them analyze reviews at the sentence level, discussing valence of reviews. In this study, we propose that text-mining techniques, such as latent semantic analysis, used in later researches could be applied to reinvestigate the issues in early studies. The relationship between online customer reviews and product sales can then be analyzed at word level, and the impact of online customer reviews could be traced back to find out which words in the reviews have influence on product sales.

Non-text Information

Besides text information like online customer reviews, there are also various types of non-text information that indicate quality of a product. Among non-text information, some types like average ratings are customer-generated, while some types like trust signs are provided by a third party. Unlike studies on online customer reviews, those on non-text information show that not all information has influence on product sales. Chen et al. (2004) proved that consumer ratings are not related to book sales at Amazon.com. Therefore, it is necessary to investigate which non-text information has influence on product sales and which does not. Otherwise, it is possible for sellers to attempt, in vain, to improve product information in order to increase sales.

In this study, we consider text and non-text information at the same time. Previous literature asserts that customer reviews and product ratings are correlated with each other (Schlosser 2005; Moe and Schweidel 2012). However, it is possible that the degree of matching between rating and online reviews varies among different customers. For example, consumers who give the same rating may post different content in online reviews. Inconsistencies between a review's rating and review will likely decrease customers' perceived helpfulness of the review (Schlosser, 2011), and then influence customers' purchasing decision. Therefore, it may be meaningful to consider both text and non-text information together.

Methodology

Data collection

Data for this research was collected from Google Play Android Application Store, which dominates the application market in terms of the number of applications available. It has large and varied categories of Android applications. Inside each category, applications are ranked based on a combination of ratings, reviews, download, countries, and other factors. We chose to use the "News & Magazines" category for this research to minimize the impact of differences coming from the content of applications on users' download choices. For example, although both fall in the "Games" category, Angry Bird and Temple Run are two totally different games in terms of their content and experiences; it is unlikely for a person to choose Angry Bird instead of Temple Run based only on the former's higher rating and more positive reviews. Compared to applications within categories like "Games," "News & Magazines" applications provide users with similar experiences. The content of "News & Magazines" applications has a higher level of similarities, especially for the applications focusing on providing news. Therefore, this category is an appropriate data sample for this research.

We chose the top 100 free "News & Magazines" applications and collected three types of data about each application on February 10th, 2015: interval, logic, and text. The interval data includes the number of download, the average rating, the number of raters, the number of screens, and the number of transversal screens. The logic data includes Top Developer badge, Editor's Choice badge, the reply from mobile application company, and the demonstration video. The text data refers to the top 10 most helpful reviews. The purpose of this study is to investigate the impact of in-store information on customers' download behavior, not to examine the dynamic changes of information on customers' download behavior. Additionally, customer reviews are listed in order of helpfulness, not posting time. Therefore, we have focused on cross-sectional data rather than panel data. Except Top Developer and Editor's Choice, most variables are easily understood. According to the elaboration on badges in Google Play Store, Top Developer and Editor's Choice are two signs indicating the high quality of one application. According to the definitions listed on Google Play's website, "Top Developer is a badge recognizing established, respected developers for their commitment to launching high-quality and innovative applications on Android. The Google Play editorial staff awards a Top Developer badge from time-to-time based on the cumulative work of the developer. Editors' Choice is a curated collection of applications that highlights some of the very best applications available on Android. Editors choose these applications for quality and great user interface, long-term popularity and innovative use of Android features. Applications chosen for Editors' Choice also receive a badge that is displayed wherever the name of application name is seen in Google play." A statistic summary on interval variables is shown in Table 1.

Variable	Data Type	Mean	STD	Min.	Max	Skewness	Kurtosis
No. of Download	Interval	$7.84*10^6$	$5.00*10^6$	$1.00*10^4$	$5.00*10^8$	9.34	90.27
Log(No. of Download) [a]	Interval	5.64	0.81	4	8.70	1.03	1.46

No. of Screens	Interval	11.02	5.94	2	24	0.47	-0.54
No. of Transversal Screens [b]	Interval	2.48	3.87	0	15	1.79	2.39
Rating	Interval	4.17	0.26	3.1	4.6	-0.95	2.23
Transversal Screen Ratio	Interval	0.19	0.27	0	1.08	1.51	1.48

[a] The distribution of No. of Download was far away from normal distribution, having high skewness value and high Kurtosis value. We convert it into log(No. of Download) to solve this problem.

[b] Transversal screen refers to a screen displayed in horizontal direction.

Table 1. Statistic Summary on Variables

Text mining methodology

To analyze reviews, we applied Latent Semantic Analysis (LSA), which is a statistical approach to analyze relationships between a set of documents and their terms; it produces a set of meaningful patterns related to documents and terms (Deerwester et al. 1990). Therefore, LSA is usually used to find patterns representing features of some documents. For example, in our context, a cluster containing "update, latest, content" may be a key feature of reviews indicating that users pay attention to some features of news applications, such as the latest version of application and promptly updated news. However, in this study, we used LSA to examine a more basic question of whether reviews have any impact on the number of download.

We used SAS Enterprise Miner 12.1 (SAS Institute Inc., Cary, NC, USA) to conduct text mining. LSA is used to transfer a text into high dimensional space with Singular Value Decomposition (SVD) to discover the underlying substantial knowledge (Landauer et al., 1998). Figure 1 presents the flowchart for performing text mining. For 100 applications, data sets containing review texts and values of 11 variables are imported into SAS Enterprise Miner for further processing and analyzing.

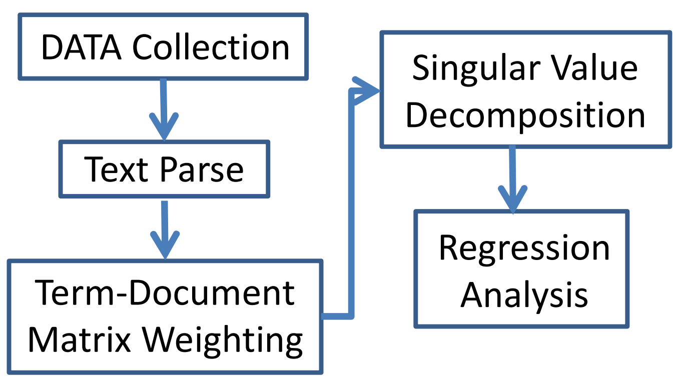

Figure 1. Flow Chart for Text Mining

Text Parse

For each application, its reviews are considered as one document. In the parse step, 100 documents were transferred into a list of terms by going through the processes of term stemming, part of speech identification, and term frequency counting. In term stemming, terms with the same stem of root form were grouped and marked as one term in order to decrease the total number of terms in the list. The part of speech of each term was identified. Terms with parts of speech, such as auxiliary, conjunction, determiner, interjection, number, preposition, pronoun, and particle, had limited information and were thus deleted from the list of terms.

After term stemming and part of speech identification, each term's frequency in all documents was counted, and the terms with a frequency less than 4 were removed from the term list due to their low frequency of occurrence. The remaining terms were used to create a 1*n (n was the number of terms) matrix for the term frequency in all documents. Further, the frequency of a term in each document was counted to generate a term-document matrix with dimensions of m*n (m was the number of documents), which was called raw term-document matrix. It would be weighted in a later step.

Term-Document Matrix Weighting

The raw term-document matrix described the correlation between terms and documents, and it was used to determine the latent semantics or effective information related to applications' number of download. The effective information in the terms that frequently occurred in all documents, like "news," was overestimated in the raw term-document matrix. Furthermore, a reviewer's language habit may 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 in order to dampen the biases brought by the high frequencies of the term in all documents and in a certain document. Let f_{ij} be the ijth 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

$$W_{ij} = [\log_2(f_{ij}+1)][1+\sum_j \frac{(f_{ij}/g_i)\log_2(f_{ij}/g_i)}{\log_2(m)}]$$

where g_i is the frequency of term i in all documents, and m is the number of documents (Dumais 1991). The first bracket on the right hand side of the equation dampens the effect of terms that occurs frequently in a document, and the second bracket dampens the terms that occur frequently in all documents.

Singular Value Decomposition

The weighted term-document matrix with dimensions 100×565 (100 documents by 565 terms) was complicated and could not be directly used to study its effect on the number of download. Thus, Singular Value Decomposition (SVD) was applied to reduce the complexity of the weighted term-document matrix and, at the same time, reduce the noises in the matrix. Let W denote the weight term-document matrix. The SVD method decomposes W into three new matrices – D, S, and T, such that $W=DST^T$, where D is orthonormal matrix with dimensions m×m (m is the document number 100), T is orthonormal matrix with dimensions n×n (n is the term number 565), and S is a diagonal matrix of singular values with dimensions m×n and non-negative real numbers on the diagonal serving as amplitudes. Matrices D, T, and S are truncated in the way that only the first q columns remained. The truncated matrices D, T, and S are multiplied to generate a new term-document matrix with lower order (q), approximating the original complex term-document matrix. The rows in the term-document matrix are mapped into a q-dimensional space as a document matrix D in which each document has q coordinates, or concepts. Similarly, each column in the term-by-document matrix can be mapped into a q-dimensional space as a term matrix T. In other words, SVD defines q concepts for terms and documents. Each term has q concept values, and each document has q concept values. SAS Enterprise Miner conducted the SVD for the weighted term-document matrix. A high number of SVD dimensions produced a high level of the approximation or prediction. Thus, the number of SVD dimensions was chosen as the maximum value 94, which was determined by SAS Enterprise Miner according to the minimum of the number of rows (100) and number of columns (565) of W.

Factor Analysis

After the SVD step, a document-concept matrix with dimensions 100×94 was generated for investigating the correlation between the latent semantics (effective information) in the documents and the number of download of the application. Each row of the document-concept matrix referred to each document; each column of the matrix referred to each concept. A term-concept matrix of 565×94 dimensions was also created in the SVD step. The term-concept matrix and the document-concept matrix shared the same concepts. We consider each concept as a factor and conduct factor analysis. The document-concept matrix was then considered as the document-factor matrix. Each column of the document-factor matrix was a factor. The document-factor matrix gave 94 factors that described the latent semantic information in the documents. The correlation between the 94 factors (or concepts) and the dependent variable log(No. of Download) was investigated.

Data Analysis

The dependent variable was set as log(No. of Download), converting the number of download into the common logarithm with base 10. The reason to use log(No. of Download) instead of the number of download was that the distribution of the number of download was far away from normal distribution, having a high skewness value (9.34) and high Kurtosis value (90.27). After the conversion, the problem of skewness was relieved. The log(No. of download) had a moderate skewness value (1.03) and Kurtosis value (1.46). Therefore, the independent variables included the text-related parameter (94 concepts in the document matrix, denoted by SVD_1 to SVD_94), and the non-text parameters (No. of Screens, No. of Transversal Screens, Transversal Screen Ratio, Rating, Demonstration Video, Editor's Choice, Company Reply, Top Developer, and Transversal Screen).

Since some variables (No. of Screens, No. of Transversal Screens, Rating, and Transversal Screen Ratio) were interval data types, their Pearson's correlation coefficient with log(No. of Download) were calculated. As a measure of the linear correlation between two variables, the Pearson's correlation coefficient values are between -1 and 1, where 1 indicates total positive correlation, 0 indicates no correlation, and −1 indicates total negative correlation. Some other variables (Variables of Demonstration Video, Editor's Choice, Company Reply, Top Developer, and Transversal Screen) were logic data types, and their effects on log(No. of Download) were evaluated by t-Tests (two-sample assuming equal variances, two-tail). The t-Tests determine whether two sets of data are significantly different from each other. For example, the value of Editor's Choice 1 means that this application has the Editor's Choice badge; the value 0 means that this application does not have Editor's Choice badge. The 100 applications were grouped into two according to the Editor's Choice values. A t-Test (two-sample assuming equal variances, two-tail) was used to compare the log(No. of Download) values of two groups. The results with P-value<0.05 indicate significant effects. The effects of other variables with logic data type on log(No. of Download) were also determined by t-Tests (two-sample assuming equal variances, two-tail).

We also conduct multivariate regression to study the correlation between the dependent variable log(No. of Download), denoted by Y, and independent variables, denoted by X_i (i=1, 2, ..., k), with the following form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

where k was the number of the independent variable, β_0 was the intercept, and β_i (i=1, 2,..., k) was the coefficient.

Three models using different independent variables were constructed. Model A only employed the text-related 94 concepts (SVD_1 to SVD_94) derived from the term-document matrix; Model B only employed the non-text variables; Model C used both text-related and non-text variables.

The independent variables have a large number (94 in Model A, 9 in Model B, and 103 in Model C). Model selection method was applied in the multiple regression models. Stepwise model selection method, which begins with no independent variable effects in the regression model, adds independent variable effects and adjusts independent variables in the model until the desired significance level or the stop criterion is met, was used.

To compare the predictions from Models A, B, and C, the R-square values were determined for these three models. In linear regression with an estimated intercept term, R-square equals the square of the correlation coefficient between the observed and predicted data values of the dependent variable. The R-square value close to 1 means a perfect prediction of the data from the regression model. Additionally, the Akaike information criterion (AIC) that measures the relative quality of a statistical model for a data set (Akaike 1974) was used to compare Models A, B, and C. AIC evaluates both the precision of the model and the complexity of the model. Let k be the number of parameters in the model and L be the maximized value of the likelihood function for the model. Then the AIC value is as follows:

$$AIC = 2k - \ln(L)$$

The best model should have the lowest AIC value.

Results and Discussion

Table 2 summarizes Pearson's correlation coefficients between log(No. of Download) and non-text variables that belong to interval data type (No. of Screens, No. of Transversal Screens, Rating, and Transversal Screen Ratio). The Pearson's coefficients of all these four variables are small with the maximum value of 0.209 and the minimum value of -0.043, indicating that these four variables have no significant relationship with log(No. of Download). However, according to previous research, rating is positively correlated with product sales (Chevalier and Mayzlin 2006). This contradictory is caused by small variance of both rating and number of download. The rating ranges from 3.1 to 4.6, with a variance of 0.068. The value of log(No. of Download) ranges from 4 to 8.6, with a variance of 0.6561. Among 100 applications, 85 have a rating between 4.0 and 4.6; 77 have a value of log(No. of Download) between 4 and 5. Since rating has many replicated values and the variation of log(No. of Download) is small, the coefficient between them becomes very small.

Independent Variable	Coefficients
No. of Screens	0.209
No. of Transversal Screens	0.019
Rating	0.008
Transversal Screen Ratio	-0.043

Table 2. Pearson's correlation coefficients between log(No. of Download) and non-text variables with interval data type

Table 3 presents the effects of non-text variables that belong to logic data type (Demonstration Video, Editor's Choice, Company Reply, Top Developer, and Transversal Screen) on log(No. of Download) calculated by using t-Tests (two-sample assuming equal variances, two-tail). The independent variables, Editor's Choice (P-value=0.00029), Company Reply (P-value=0.01057) and Top Developer (P-value<0.0001), showed significant effects on log(No. of Download); this means that whether an application has an Editor's Choice badge and a Top Developer badge, as well as whether the application company replies to customers' comments, have significant influence on log(No. of download).

Independent Variable	P-value
Demonstration Video	0.2329
Editor's Choice	0.00029
Company Reply	0.01057
Top Developer	<0.0001
Transversal Screen	0.25043

Table 3. The effects of non-text variables with logic data type on log(No. of Download) calculated by t-Tests (two-sample assuming equal variances, two-tail)

Table 4 presents fit statistics of Models A, B, and C. By comparing R-square values and AIC values, Model C with both text-related and no-text variables was shown to provide the best prediction on log(No. of Download) in multiple regression (R-square=0.6818 and AIC=-124.5873). Model A with the effects of text-related variables (SVD concepts) gave the second best prediction (R-square=0.5442 and AIC=-96.6556). Model B with only no-text variables gave the worst prediction on log(No. of Download) (R-square=0.2352 and AIC=-62.9010).

Model	R-square	AIC
Model A (SVD factor loadings)	0.5442	-96.6556
Model B (non-text variables)	0.2352	-62.9010
Model C (non-text variables + SVD factor loadings)	0.6818	-124.5873

Table 4. Model Fit Statistics

As Model C gave the best fit of dependent variable in multivariate regression, maximum likelihood estimates for parameters in Model C are summarized in Table 5. The stepwise selection model determined 15 independent variables for multiple regression of the dependent variable log(No. of Download). Fifteen independent variables included two non-text variables (Editor's Choice and Top Developer) and 13 text-related variables. Through the LSA and the regression analysis, the semantic characteristics described by SVD concepts were found to have an impact on log(No. of Download). Note that the SVD concepts determined the underline connections of term-term, term-document, and document-document, did not have explicit semantic meanings.

Parameter	Estimate	Error	t Value	P-value
Intercept	4.660	0.345	13.49	<.0001
Editor's Choice	-0.432	0.122	-3.53	0.0007
Top Developer	-0.297	0.071	-4.19	<.0001
SVD1	4.277	0.880	4.86	<.0001
SVD14	-1.544	0.471	-3.28	0.0015
SVD16	1.000	0.462	2.16	0.0333
SVD17	1.310	0.474	2.76	0.0071
SVD23	1.260	0.501	2.52	0.0138
SVD26	1.066	0.472	2.26	0.0264
SVD36	-1.091	0.510	-2.14	0.0353
SVD53	-1.765	0.588	-3	0.0035
SVD56	-2.746	0.520	-5.28	<.0001
SVD72	-1.490	0.618	-2.41	0.0181
SVD76	2.498	0.638	3.92	0.0002
SVD82	1.497	0.661	2.27	0.026
SVD90	-1.661	0.639	-2.6	0.011

Table 5. Estimates for parameters in Model C

The positive concepts included SVD1, SVD16, SVD17, SVD23, SVD26, SVD76, and SVD82; the negative concepts included SVD14, SVD36, SVD53, SVD56, SVD72, and SVD90. The positive concepts related to the increase of log(No. of Download), while the negative concepts related to the decrease of log(No. of Downloads). During the SVD step, a matrix with dimensions 565*94 (565 was the number of terms, and 94 was the number of concepts) for terms was generated from Equation 2. In the term matrix, each column represented a concept, and each term had 94 concept values. For each of these positive and negative concepts, five terms with the highest concept values were extracted, generating a table for terms related to positive concepts and negative concepts. A term having a higher concept value in a positive or negative concept generates a higher positive or negative impact on the log(No. of Download).

Negative concepts	Positive concepts
cover, deliver, search, background, space, swipe, support, navigation, format, flipboard offline, unusable, broadcast, hang, wifi, area, language, online, touch, design, tap, specific, alternative, quality, response, useless, hold, notify, rarely, garbage	application, good, news, update, fix, magazine, page, widget, display, auto, customizable, catch, newspaper, improvement, paper, medium, a lot, important, last, blogs, icon, English, solve, bar, science, sense, type, view, dev, original, suggestion, message, second, official, headline

Table 6. Terms related to positive concepts and negative concepts

Sixty-five terms from the negative concepts and positive concepts were gathered in an effective term set. Terms without useful information, like "cover," "good," and "catch," marked with gray in Table 6, were removed from the term set. The remaining 43 effective terms that reflect features of application that customers care about were further classified into three groups according to application main feature, application auxiliary feature, and application service, a similar method used by Chen et al. (2012). As shown in Table 7, mobile application users care about the application's main feature, the application's auxiliary feature, and the application's service. Therefore, mobile application companies need to make improvements on these aspects.

Application's main feature	navigation, offline, search, broadcast, wifi, news, newspaper, widget, display, blogs, icon, message, headline, magazine
Application's auxiliary feature	background, bar, deliver , swipe, flipboard, hang, area, language, touch, design, tap, notify, response, auto, customizable, type, science, English, space, view, page, dev, official
Application's service	support, update, fix, improvement, solve

Table 7. Classifications of effective terms

Conclusion and Limitations

In this paper, we examine an important research question concerning the product information in the context of mobile application purchasing: What specific information influences the sales of applications? We addressed this question by investigating the impact of two types of application information on the number of download. We categorized information from an application store into text and non-text information. Text mining techniques and linear regression models were employed to investigate 1000 reviews of 100 news and magazine applications. A number of practical and research implications can be derived from this study.

This study makes contributions in three aspects. First, it is meaningful to analyze customers' purchasing behavior in mobile environment because mobile environment is not quite the same as World Wide Web environment, such as small screen and inconvenience of searching for information. Customers usually make intuitive decisions and download mobile applications based on the information in mobile application store. This study makes theoretical contributions by revealing the mechanism of customer decision-making in mobile environment. Second, although customer rating and online review has been largely investigated in previous researches, the difference between them has not even been carefully examined. This study takes the difference into account and considers text and non-text information together, making theoretical contributions to research areas related with customer behaviors in digital world. Third, this study has theoretical implication for other behavioral studies in that our findings indicate that besides investigating the overall impact of online customer reviews on product sales, it is also necessary to find out specific influential words in their reviews. By revealing the meaningful information in the review, the relationship between online customer reviews and customers' purchasing behavior becomes clearer and more understandable.

Additionally, this study provides empirical implications that guide mobile application developers in improving the features of mobile applications, which may increase sales volumes. Our findings show that users care about applications' main features, auxiliary features, and services. Therefore, mobile application companies could make efforts to meet customers' requirements on these fields. Based on the influential key words in customers' comments, improvements could include making the content in applications more attractive, providing customization functions, making the applications easy to use, and updating the system frequently to fix technical problems. Furthermore, this study helps mobile application companies avoid making efforts in vain since our findings indicate that some information on applications has significant impact on number of download, while some does not. Therefore, the mobile application companies should take actions that are valuable for their customers, such as reply to customers' comments and get badges that indicate high quality from the application stores.

This study can be improved from many perspectives. First, we used number of download as a proxy of sales of application to indicate customers' purchasing behavior. However, when downloading free applications, customers do not need to pay anything. Although the revenue of news and magazine applications may come from advertising and the number of download may be indirectly related to sales to some extent, our results may not be able to fully reflect consumers' purchasing mechanism. Second, we did not consider the impact of tenure of applications on the volume of download. The volume of download for a new application is relatively lower than that for an old application. Therefore, tenure of an application should be a control variable in this study. Third, there are some other information in the mobile application store have not been considered in this study, such as in-application purchase, product description, and quality of screenshots. These details are also available for customers when they download application and may also influence their purchasing decisions. It is necessary to consider all available information at the same time to fully understand the impact of mobile application information on download.

The limitations in this study also call for future research. First, the data set comes from one category and only contains 100 applications. More interesting patterns between product information and product sales may be found if using a larger data set. Second, application screens are also an important indicator of the application's quality, having impacts on download behavior. Compared with the number of screens, the content in the screen is more critical and influential for prospective consumers. Combining screen content evaluation with online customer review analysis, which is a mixed methodology, could provide a more robust result. Moreover, the text information on applications does not just refer to customer reviews; product description is also important text information available for prospective consumers and may also influence the number of download. It may be more meaningful to compare the impact of customer reviews on the number of download with that of product descriptions. Customer reviews are consumer-generated and out of companies' control, whereas product descriptions are fully controlled by companies. Analyzing information from different sources would enable scholars and practitioners to have a comprehensive understanding of the impact of product information on sales of products.

REFERENCES

Akaike, H. 1974. "A new look at the statistical model identification," IEEE Transactions on Automatic Control, 19 (6), pp.716–723.

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 Processings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, New York, NY, pp. 47-56.

Cao, Q., Duan, W., and Gan, Q. 2011. "Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach," Decision Support Systems, 50, pp. 511-521.

Chen, J., Zhang, J., and Zhang, Y. 2012. "Impact Factors of Online Customer Reviews Usefulness: A Text Semantics Approach," Library and Information Services, 56(10), 119-123.

Chen, P., Wu, S. and Yoon, J. 2004. "The Impact of Online Recommendations and Consumer Feedback on Sales," ICIS 2004 Proceedings. Paper 58.

Chevalier, J. A. and Mayzlin, D. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews," Journal of Marketing Reserch, Vol. XLIII, pp. 345-354.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K. and Harshman, R. 1990. "Indexing by Latern Semantic Analysis," Journal of the American Society for Information Science, 41(6), pp.391-407.

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

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

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), pp. 233-242.

Dumais, S. T. 1991. "Improving The Retrieval of Information From External Sources," Behavior Research Methods, Instruments, & Computers, 23(2), pp. 229-236.

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), pp. 291-313.

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

Landauer, T. K., Foltz, P. W. and Laham, D. 1998. "An Introduction to Latent Semantic Analysis," Discourse Processes, 25, pp. 259-284.

Larsen, K. R. Monarchi, D. E. and Hovorka, D. S. 2008. "Analyzing unstructured text data: using latent categorization to identify intellectual communities in information systems," Decision Support Systems, 45 (4), pp. 884-896.

Li, M., Huang, L., Tan, C. and Wei, K. 2013, "Helpfulness of Online Product Reviews as Seen by Consumers: Source and Content Features," International Journal of Electronic Commerce, 17 (4), pp. 101-136.

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.

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

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

Pavlou, P. A. and Dimoka, A. 2006. "The Nature and Role of Feedback Text Comments in Online Marketplaces: Implications for Trust Building, Price Premiums, and Seller Differentiation," Information System Research, 17 (4), pp. 392.

Rao, A. R., Qu, L., and Ruekert, R. W. 1999. "Signaling Unobservable Product Quality through a Brand Ally," Journal of Marketing Research (36:2), pp. 258-268.

Schlosser, A. E. 2005. "Posting versus Lurking: Communicating in A Multiple Audience Context," Journal of Consumer Research 32, pp. 260-265

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), pp. 226-239.