Fall 9-11-2020

Examining the Impact of User Reviews On Mobile Applications Development

Miriam Erne
Singapore University of Social Sciences, miriam.erne@googlemail.com

Zhiying Jiang
Singapore University of Social Sciences, jiangzhiying@suss.edu.sg

Vanessa Liu
Singapore University of Social Sciences, vanessaliusw@suss.edu.sg

EXAMINING THE IMPACT OF USER REVIEWS ON MOBILE APPLICATIONS DEVELOPMENT

Miriam Erne
Singapore University of Social Sciences
Miriam.erne@googlemail.com

Zhiying Jiang
Singapore University of Social Sciences
jiangzhiying@suss.edu.sg

Vanessa Liu
Singapore University of Social Sciences
vanessaliusw@suss.edu.sg

ABSTRACT

User reviews were often collected to enlighten mobile applications (apps) developers on areas for improvement and novel features. However, users might not always possess the required technical expertise to make commercially feasible suggestions. The value of user reviews also varied due to their unmanageable volume and content irrelevance. In our study, over 40,000 user reviews with 50 apps would be analyzed using Python coding and regression analysis to examine the impacts of innovation and improvement led by users on apps performance in terms of revenues and user ratings. The developers' lead time in responding to user reviews would be included as a moderator to investigate whether apps performance would be enhanced if developers respond faster. Our study should represent one of the first few attempts in offering empirical confirmation of the value of co-creation of apps with users.

KEYWORDS

user reviews, user led innovation, user led improvement, data analytics

INTRODUCTION

Nowadays application distribution platforms such as Apple App Store and Google Play provide millions of different mobile applications (apps) to users. As of the fourth quarter of 2019, there were around 2.57 million apps for android users and 1.84 million apps for App Store users available (Statista, 2020). Survival in such a "hyper-competitive" mobile market was challenging to apps developers (Comino et al., 2016). Unwanted or unpopular apps could be phased out very shortly after launch, resulting in a waste of development cost and effort. To sustain competitiveness, it is therefore becoming increasingly important for apps developers to pursue continuous improvement and launch novel features that meet user needs (e.g., see Maalej et al., 2016; Maalej and Hadeer, 2015; Chen et al., 2014).

One key channel for users to voice out their feedback is through submitting reviews. Indeed, most apps actively elicit user comments as they are enlightening to the app's developers in terms of areas for improvement and novel features. As user needs vary significantly and the usage of the apps could differ across contexts, users may be a good source of creative ideas for development of innovative functionalities. Through gaining actual experience with the apps, users are also likely to be able to spot a non-working feature than the apps developers. For example, one apps feature may work in one version of smartphones but not another. User reviews could therefore help spotting bugs and enable continuous improvement of the apps.

User reviews, however, may not be fully utilized as conducting full screening of all reviews could be challenging. For instance, online gurus like Facebook could generate as high as at least 2,000 user reviews per day (Chen et al., 2014). The issues covered in the reviews could be highly diverse, ranging from complaints about the price of the apps to the frequency of advertisements. Manual processing of these reviews is simply impossible. Management of user reviews could be overwhelming and expensive to apps developers. Following up on user reviews is yet even more time-consuming. Users may provide feedback that is unconstructive and commercially infeasible when they do not possess related technical expertise. It may not always be economically justified for developers to translate user feedback into actual apps features (Ives and Olsen, 1984). Most prior researchers focused on the development of analytical tools for categorization of user reviews (e.g., Maalej et al., 2016; Maalej and Hadeer, 2015). It is not clear whether addressing user comments could result in better apps performance.

To address this gap, we would categorize and analyze over 40,000 user reviews associated with 50 apps to examine the impact of user reviews on apps performance in terms of user ratings and apps revenues. Specifically, we conceptualize reviews with innovative suggestions as "user led innovation" and those with bug-fixing suggestions as "user led improvement". We would investigate whether and how adopting the user suggestions would lead to higher user ratings and increased revenues from apps purchases. This correlation, however, may be moderated by developers' responsiveness. If the apps developers take a long time to respond and address the user comments, the innovativeness of the suggestions may depreciate. Competitors might have already built similar features in their latest apps during the response lead time. Similarly, users might get disappointed if the developers did not promptly address the errors they pointed out. We hence include developers' responsiveness as a moderator on the relationship between user led innovation/improvement and apps performance. In other words, the effect of user reviews on apps performance would be greater if the developers were more responsive and attend to the user feedback faster.

The remainder of this paper is structured as follows: first, we will explain the conceptual framework and the related past studies. The research methodology and the data analysis procedure will then be presented. Finally, the findings will be discussed and the theoretical and managerial implications will be drawn.

CONCEPTUAL FRAMEWORK

User Involvement

The notion of user involvement was well documented in the literature, referring to the level of personal relevance and importance attached by users to the system (Barki and Hartwick, 1989). In broad terms, it is defined as "direct contact with users" (Kujala, 2003). Recently, it was observed that customers had become more and more involved in the product development (Prahalad and Ramaswamy, 2013). User involvement was essential and indispensable for system/ software developers as it helped to collect more accurate user requirements and enable quality improvement, resulting in better fulfillment of user needs and higher user satisfaction (Kujala, 2008; Kaulio, 1998). User involvement was therefore recognized by previous researchers as beneficial to the improvement of quality and performance (Berger et al., 2005). Terms such as co-creation or co-design had emerged to describe the collaboration between developers and users. Other terms included quality function deployment (QFD), user-oriented product development, concept testing, Beta testing, consumer idealized design, lead user method and participatory ergonomics (Kaulio, 1998). In the collaborative process, users may assume the roles of providers of information, commentators or objects for observations.

In the context of mobile apps, users could submit complaints or suggestions (Khalid et al., 2015; Panichella et al., 2015). Users' wish list of additional features could inspire potential new apps development (Barlow and Moeller, 1996). Users may point out bugs such as incompatibility or poor functionality (Khalid et al., 2015; Panichella et al., 2015). However, the number of user reviews received could be immeasurable and unmanageable. More importantly, not all feedback is useful. Almost 65% of apps reviews were found to be noisy and irrelevant (Chen et al., 2014). Some suggestions might be solely emotional and commercially infeasible for implementation.

Many tools were therefore developed to aid the search, screening, and extraction of useful information from user reviews. A review of the current literature showed that different tools were built with different mining objectives. Examples included MARK (Mining and Analyzing Reviews by Keywords) (Vu et al., 2015), MARA (Mobile App Review Analyzer) (Iacob and Harrison, 2013), ALERTme (Guzman et al., 2017), and AR-Miner (App Review Miner) (Chen et al., 2014). These tools made use of techniques like natural language processing, topic modeling, clustering and machine learning algorithms to search, classify, extract, group and rank user reviews based on pre-defined keywords or categories. Abstract and Keywords

Every submission should begin with an abstract of no more than 150 words, followed by a set of keywords. The abstract should be a concise statement of the problem, approach, and conclusions of the work described. It should clearly state the paper's contribution to the field.

User Led Innovation and User Led Improvement

User reviews, if carefully and properly screened and processed, could be vital to innovativeness and ongoing improvement of apps development. For example, users might point out interesting and novel features that could be added for iPhone users. They may also highlight the incompatibility of an apps with iPhones. As user profiles and needs are so diverse, it is difficult for apps developers to consider all possible new features. Smartphones and software updates (e.g., iOS) are so frequent that it is also complicated to test functionalities and compatibility thoroughly. Users are in a better position to detect bugs specific to a particular phone model. User reviews could be a good source to

identify creative solutions and usability issues. Though some users may be tech-non-savvy, the problems experienced by them might never be foreseen in the development process. Addressing their comments could make the apps more user-friendly to novice users.

We therefore conceptualize user reviews with new feature requests as user led innovation. It denotes requests from users on new features to be added to the apps or new apps development. User reviews with suggestions on improvement are conceptualized as user led improvement. It denotes reports from users about unwanted errors, bugs, annoying advertisements and other usability problems. If addressed properly and adequately, user led innovation and user led improvement should contribute to higher user satisfaction and hence user ratings. More users will be attracted to purchase the apps, leading to higher apps revenues (Kujala, 2008).

Accordingly, we hypothesize that:

H1a: User led innovation has a significant and positive impact on apps user ratings.

H1b: User led innovation has a significant and positive impact on apps revenues.

H1c: User led improvement has a significant and positive impact on apps user ratings.

H1d: User led improvement has a significant and positive impact on apps revenues.

Developer Responsiveness to User Reviews

The time taken by developers to respond to user reviews may matter (Vaniea and Rashidi, 2016). After a user submitted his/her feedback, he/she may tend to expect the developer to address the suggestion or bug quickly. For example, if the developer response is slow, the current users may continue to experience the bugs in the regular apps usage and may eventually rescind usage or even uninstall the apps. Conversely, users may tend to be more positive about the apps if their concerns and problems were addressed promptly. Timely responses are even more crucial for suggestions of new features. The degree of novelty would be diminished and the risk of being copied by competitors would increase over time. Developers should, however, be cautioned about the frequency of apps dates. Frequent issuance of updates upon minor bugs fixing may actually cause disruptions to users. In general, reasonable responsiveness to user reviews should lead to better quality and performance of apps (Schenchk, 2013; Armerding, 2012). Accordingly, we hypothesize that:

H2a: Developer responsiveness significantly and negatively moderates the relationship between user led innovation and apps user ratings.

H2b: Developer responsiveness significantly and negatively moderates the relationship between user led innovation and apps revenues.

H2c: Developer responsiveness significantly and negatively moderates the relationship between user led improvement and apps user ratings.

H2d: Developer responsiveness significantly and negatively moderates the relationship between user led improvement and apps revenues.

Our research model is presented in Figure 1.

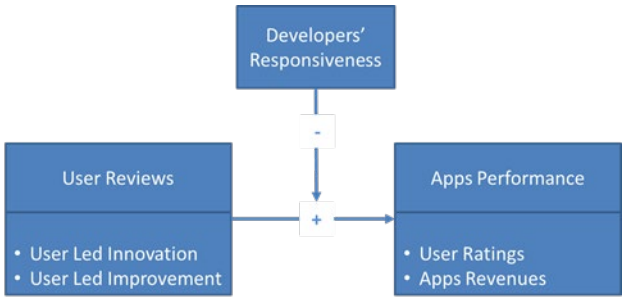

Figure 1. The Research Model

RESEARCH METHODOLOGY

Research Context

Data would be collected through App Annie, a business intelligence company. It stores data on a collection of health and fitness apps, including the apps user reviews and revenues generated from each app. In our study, only apps that have been active for at least one year would be included in the sampling. Active apps should provide more valid results as it was common in the mobile apps industry that numerous apps could have been removed before their official launch. A total of 50 apps would be selected for our analysis as their revenue constitute almost 75% of the total revenue in the health and fitness apps market. There are 189,527 user reviews available for these selected apps.

In order to measure the effect of user reviews on apps performance, a specific research time frame would be defined. Only reviews posted after the second last updated version and before the latest version of the apps would be included in our samples. This enables us to examine whether the user reviews led to development of new features in the resultant update of the apps. The final sample is estimated to consist of a total of 40,619 user reviews, representing 21.4% of the total reviews associated with the selected apps.

Operationalization

User led innovation

User led innovation would be operationalized as a feature request by users. It refers to the wish or idea proposed by users on a new functionality that should be added but does not exist yet (Cheung, 2013; Wiggins, 2015). Examples of such review are "Needs to have a value for calories burned for strength training too" and "Missing a compatible version for Apple Watch". Keywords to be used to screen for bug reviews would be "add", "please", "need", "prefer", "request", "suggest", and "wish" (Maalej and Hadeer, 2015).

User led improvement

User led improvement would be operationalized as a bug review and a complaint. A bug review reports on an unwanted error in an app. It could be any kind of problems with the app, a crash, an error or a performance issue arising from programming failure by the developer (Maalej and Hadeer, 2015). Examples of such review are "*it's not letting me sign up and I deleted the app and re-downloaded it but it's not working*" and "*if you open the app in the watch it tries to connect for a minute (literally a minute) then crashes*". Keywords to be used to screen for bug reviews would be "bug", "fix", "problem", "issue", "defect", "crash", and "solve" (Maalej and Hadeer, 2015).

A complaint could be about user annoyance with certain unwanted features, such as too many popped up advertisements. Reducing the number of ads might be room for improvement for the developer. Examples of these reviews are "*paid for the ap. Still get ads pushed to me. Don't advertise to me if I paid the money for the non-ad version*" and "*The avalanche of ads makes it unusable unless you pay $3 each and every month*".

Developer responsiveness

Developer responsiveness would be measured by the time interval (number of days) from the first posted date of the user review to the update date when the comment would be addressed (e.g., the new feature is added; the bug is fixed or the pop-up advertisements are removed).

Apps performance

Apps performance could be operationalized in a number of ways such as the number of downloads and apps ratings etc. In this study, apps performance would be measured using the user ratings and the revenues generated from the apps during the research time frame. In particular, capturing the revenue changes allows us to examine the financial impact on the apps developers more directly. Revenues could include purchases of apps, micro-transactions within an app or in-app advertisement (IADV) (Ghose and Han, 2014). The revenues for each app would be computed by a summation of the daily revenues for the research time frame.

Data Analysis

A subtraction and categorization process would be conducted to identify the user reviews that specifically pertained to user led innovation and user led improvement.

First, *generic* reviews would be subtracted to isolate the *specific* reviews (Chen et al., 2014). *Generic* reviews are noisy and irrelevant reviews that do not provide any information. Examples of such review are "*by far the best app on meditation!*" and "*I love this app and have done since the moment I started using it. Potentially helped me get*

through a period of anxiety…". *Specific* reviews, on the other hand, are those that represent a feature request, a bug report and a complaint.

Next, the specific reviews would be categorized to shortlist the innovation- and improvement-related reviews. As the data for specific reviews took the form of written texts, steps would be taken to convert the text data into numerical data. Each review would be enumerated with a Python code respectively according to its match with the categories of "feature request, "bug report" or "complaint".

Regression Analysis would then be conducted to analyze the correlations in the measurement model.

CONCLUSION

User reviews have been regarded as useful for applications development in the extant literature of user involvement. This presumption might not hold in the context of mobile apps, where hundreds or thousands of users may easily participate in the apps design through submitting reviews online. The volumes of user reviews might be hardly manageable, and the usefulness of the reviews might also be impaired by the users' lack of technical expertise. It was questionable whether the users' creative solutions could lead to actual improvement of the app's performance. Our study therefore attempts to fill this gap in the literature by empirically examining the effect of user led innovation / improvement over apps performance in terms of revenues and user ratings. Our findings are expected to demonstrate the value of co-creation of apps with users (Gustafsson et al., 2012). By testing the moderating effect of developers' effectiveness, we seek to provide guidelines to apps developers in setting response time to user reviews.

WORK TO BE COMPLETED

This research is currently in the stage of data analysis. All data has been collected and the initial sample selection is nearly completed. The data subtraction and categorization processes described in the "data analysis" section would then be performed, followed by the text-to-numeric data conversion procedure to reach a set of final sample user reviews. A regression analysis would then be performed. Finally, the findings would be reported, along with discussion of theoretical and managerial implications.

REFERENCES

1. Armerding, T. (2012). Why Users Don't Often Upgrade Software When They Should. Available at https://www.csoonline.com/article/2132061/security-awareness/References 48 why-users-don-t-often-upgrade-software-when-they-should.html.
2. Barki, H., & Hartwick, J. (1989). Rethinking The Concept of User Involvement. *MIS Quarterly*, pp.3–63.
3. Barlow, J., & Møller, C. (1996). A Complaint Is A Gift: Using Customer Feedback As A Strategic Tool. Berrett-Koehler Publishers.
4. Berger, C., Möslein, K., Piller, F., & Reichwald, R. (2005). Co-Designing Modes of Cooperation at the Customer Interface: Learning From Exploratory Research. *European Management Review* (21), pp. 70-87.
5. Chen, N., Lin, J., Hoi, S., Xiao, X., & Zhang, B. (2014). AR-miner: Mining Informative Reviews for Developers from Mobile App Marketplace. in *Proceedings of the 36th International Conference on Software Engineering*, pp. 767–778.
6. Cheung, K. W. (2013). A Feature Request Is A Bug...Is a Dumb Detail. Available at https://www.getdonedone.com/a-feature-request-is-a-bug-is-a-task-is-a-to -do-is-a-dumb-detail/.
7. Comino, S., Manenti, F. M., & Mariuzzo, F. (2016). *Updates Management in Mobile Applications*. iTunes vs google play.
8. Ghose, A., & Han, S. P. (2014). Estimating Demand for Mobile Applications in the New Economy. *Management Science* (606), pp.1470–1488.
9. Gustafsson, A., Kristensson, P., & Witell, L. (2012). Customer Co-Creation in Service Innovation: A Matter of Communication? *Journal of Service Management* (23), pp. 311–327.
10. Guzman, E., Ibrahim, M., & Glinz, M. (2017). A little bird told me: Mining tweets for requirements and software evolution. In *2017 IEEE 25th International Requirements Engineering Conference*, pp. 11–20.
11. Iacob, C., & Harrison, R. (2013). Retrieving and analyzing mobile apps feature requests from online reviews. In *2013 10th IEEE Working Conference on Mining Software Repositories MSR*, pp. 41–44.

12. Ives, B., & Olson, M. H. (1984). User Involvement and MIS success: A Review of Research. *Management Science* (305), pp. 586–603.
13. Kaulio, M.A. (1998). Customer, Consumer and User Involvement in Product Development: A Framework and A Review of Selected Methods. *Total Quality Management* (91), pp.141–149.
14. Khalid, M., Asif, M., & Shehzaib, U. (2015). Towards Improving the Quality of Mobile App Reviews. *International Journal of Information Technology and Computer Science* (710), pp. 35.
15. Kujala, S. (2003). User Involvement: A Review of the Benefits and Challenges. *Behavior & Information Technology* (221), pp. 1–16.
16. Kujala, S. (2008). Effective User Involvement in Product Development by Improving the Analysis of User Needs. *Behavior & Information Technology* (276), pp. 457–473.
17. Maalej, W., & Hadeer, N. (2015). Bug Report, Feature Request, or Simply Praise? On Automatically Classifying App Reviews. In *2015 IEEE 23rd International Requirements Engineering Conference*.
18. Maalej, W., Kurtanović, Z., Nabil, H., & Stanik, C. (2016). On The Automatic Classification Of App Reviews. *Requirements Engineering* (213), pp. 311–331.
19. Magnusson, P. R. (2003). Benefits of Involving Users In Service Innovation. *European Journal of Innovation Management* (6:4), pp. 228–238.
20. Mooi, E., & Sarstedt, M. (2011). *A Concise Guide to Market Research*. Springer Berlin Heidelberg.
21. Panichella, S., Di Sorbo, A., Guzman, E., Visaggio, C. A., Canfora, G., & Gall, H. C. (2015). How Can I Improve My App? Classifying User Reviews for Software Maintenance and Evolution. In 2015 IEEE International Conference on Software Maintenance and Evolution ICSME, pp. 281–290.
22. Prahalad, C. K., & Ramaswamy, V. (2013). *The Future of Competition: Co-Creating Unique Value with Customers*. Harvard Business Press.
23. Schenck, Stephen. (2013). Why I don't update apps. Available at: http://pocketnow.com/2013/ 09/25/dont-update-apps.
24. Statista (2019). Number of apps available in leading app stores as of 4th quarter 2019. Available at: https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/
25. Vaniea, K., & Rashidi, Y. (2016). Tales of Software Updates: Software. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 3215–3226.
26. Vu, P. M., Nguyen, T. T., Pham, H. V., & Nguyen, T. T. (2015). Mining User Opinions In Mobile App Reviews: A Keyword-Based Approach. In *2015 30th IEEE/ACM International Conference on Automated software engineering ASE*, pp. 749–759.
27. Wiggins, N. (2015). The difference between a Bug, Error and Feature. Available at: https:// www.webigence.com/blog/the-difference-between-a-bug-error-and-feature.