An Exploration of User-Driven Assessments of Travel Enhancing Apps

Lidija Lalicic [a],
Christian Weismayer [b]

[a] Department of Tourism and Service Management
MODUL University Vienna, Austria
lidija.lalicic@modul.ac.at

[b] Department of Applied Statistics and Economics
MODUL University Vienna, Austria
christian.weismayer@modul.ac.at

Abstract

Distribution platforms of apps help tourists to select the most supportive apps. These platforms represent information sources offering fruitful input for app designers and marketers designing apps. This study analyzed 240 travels enhancing app reviews based on reviews topics, furthermore making a distinction between price and app category. Interestingly, price shows to be a discriminative issue for receiving informative reviews to enhance consumers' experiences. Furthermore, differences between app categories concerning the review content are detected. Finally, a set of recommendations for app vendors are given to effectively steer tourists' experiences.

Keywords: travel apps, app reviews, content analysis, user-driven assessments

1 Introduction

Smartphones are a dominant force shaping visitors' behaviour (Wang et al., 2014). Recent studies on tourists' smartphone behaviour while travelling illustrate the supportive role of apps (Tussyadiah & Wang, 2014; Wang et al., 2014). Hence, tourists use different apps to meet their needs while travelling and enhance their experiences on the go. As a result, Destination Management Organizations (DMOs) started to integrate and/or promote apps into their marketing strategies and steer visitors' experiences. Practitioners are eager to learn about tourists experience with the apps while travelling in order to optimally develop accurate and enjoyable apps. The various application distribution platforms allow users to search, buy and deploy software apps for mobile devices (Harman et al., 2012; Maalej & Nabil, 2015; Fu et al., 2013). Furthermore, these platforms allow users to share opinions about the app in form of text reviews (Ali et al., 2012; Maalej & Nabil, 2015).These reviews, on the one hand, help users to navigate and decide which app to download. On the other hand, app reviews serve as a communication channel between users and developers (Panichella et al., 2015). Users often provide detailed information about ideas for new features, documentation of released features, or requirements, that help vendors to develop, maintain or advance their software (Guzman, & Maalej, 2013; Maalej & Nabil, 2015; Panichella et al., 2015). Consequently, various studies highlight the importance of reviews for an app success. Given the need to optimally steer the travel experience by the effective design of travel enhancing apps, an understanding of app flaws is required. Despite the importance of apps, there is limited research on app flaws and opportunities to modify apps to subsequently enhance for example tourists' experiences. Therefore, the first aim of this study is to explore if users praise or criticize apps that are used for travel enhancing purposes. Second, the study indicates the various flaws among different app categories. Third, the study offers insights into the difference between charged and free apps as well as between various app categories. Fourth, this study provides recommendations for app designers seeking to steer tourists' experiences. Marketers need to integrate these user-driven assessments for three reasons.

2 Method

The app reviews used in this study were collected from the two most popular app distribution platforms; Google Play and Apple Store. As different apps may furthermore lead to different user comments, a broad travel related scope of app features is necessary to be able to capture the typical travel app review content varying upon different categories. Therefore, a pre-selection was made categorizing each app into five different types (communication e.g. Skype or Facebook Messenger, entertainment e.g. Flicker or Instagram, facilitation e.g. Google Earth or Yahoo Weather, information e.g. Booking.com or Expedia, and social marketing e.g. WhatsApp Messenger or QQ Messenger) (based upon Tussyadiah & Zach, 2012). From each category the five most popular apps (highest rank) in the year 2015 were selected. A balanced level of star ratings is intended to capture the broad mass of typical review content. Thus, frequencies per star rating were outweighed in advance and are given in brackets next to their evaluation: 1 star (48), 2 stars (48), 3 stars (49), 4 stars (46), and 5 stars (49). Furthermore, a cross tabulating charged versus free apps and star ratings reveals a balanced frequency of observations. Gender is quite balanced too, with a slight overhang of women: men 44.3%, and women 55.7% respectively. Age descriptives reveal themselves as follows: minimum 12 years, maximum 66 years, arithmetic mean 25.22 years, and std 12.65 years. More than ¾ of all pre-selected reviews were posted within a time frame of one year prior to the data collection approach to keep the probably time-varying review content up-to-date. In total 240 reviews were collected; for each review the following aspects were collected: (i) the application name reviews were collected, (ii) the app category, (iii) the operating system, (iv) star rating, (v) price, (vi) profile of the review and (vii) review text. Then, the resulting review text was analysed by 20 coders. The content analysis was guided based upon the taxonomy of Pagano and Maalej (2013) and extended by items of Khalid et al.'s study (2015) resulting in 20 items: 'praise', 'helpfulness', 'shortcoming', 'crash report', 'feature removal request', 'missing feature', 'reference to other apps', 'recommendation to use app', 'noise-meaningless information', 'dissuasion to use/buy app', 'missing content report', 'improvement report', 'dispraise', 'how to use the app', 'other feedback', 'compatibility', 'hidden costs', 'network problems', 'privacy issues' and 'unresponsive apps' (based upon Pagano et al.; 2013; Khalid et al., 2015). The items were based upon binary coding of the reviews analysed, defining whether this sort of content showed up in the review or not. Furthermore, the quantitative content analysis was used to classify the reviews according to the categories and counter for their occurrences and provide statistical inferences with text populations.

3 Discussion

There are three different price levels for iPhone apps (frequencies are mentioned in brackets): € 0 (209), € .99 (21), and € 2.99 (10), but four different price levels for the same app running on an Android operating system: € 0 (209), € .75 (10), € .99 (11), and € 2.99 (10). One app was cheaper if downloaded for Android compared to iPhone, € .75 versus € .99. To get an overview of the general difference between charged vs. free apps, the price levels are collapsed into just two categories in the first attempt: free apps (209 reviews) vs. charged apps (31 reviews). Typically, if groups are known a priori and there are only two groups, discriminant analysis is used. This leads to just one single possible discriminant function representing some kind of higher information separating the two groups. If there are more than two groups, multiple discriminant function analysis, as CVA (canonical variates analyses), is mostly preferred, especially, if it comes to the point when multivariate relationships are analysed in an exploratory manner making use of visualizations to gain first insights into new phenomena. CVA focuses on the optimal graphical representation of observations and items and maximizes the between-group information to the within-group information in the joint space that a clear group separation becomes apparent. The idea is similar to the one behind cluster analyses. The resulting space is consequently highly dependent on the group membership classification contained in the grouping variable. The predicted group membership compared with the a priori known group membership gives insight into the appropriateness and heterogeneity of the groups in terms of the observed predictors. However, CVA will be used for the binary price coded group membership as well as for the three price level coding to be explained later. The former one does not take the different observed price levels into consideration. The following rounded mean relative absolute percentage errors presented in ascending order can be interpreted as real distances on the original dichotomous scale: hidden costs (9%), privacy issues (10%), how to use the app (11%), other feedback (12%), reference to other apps (12%), noise (22%), crash (25%), dissuasion (25%), recommendation (27%), unresponsive (29%), feature removal request (30%), network problem (31%), praise (34%), content request (37%), compatibility (38%), helpfulness (41%), missing feature (42%), shortcoming (43%), dispraise (44%), and improvement request (46%). Reviews on charged apps are located closer to items like network problems, improvement requests, missing features, feature removal or request, and compatibility. Furthermore, shortcomings are mentioned most often for both price categories, unresponsiveness comments more or less solely in free app reviews. Statements on the helpfulness or improvement requests are predominant in charged app reviews. Interestingly, issues such as ‘how to interact with the app’, ‘privacy issues’, ‘hidden costs’, and some others are hardly mentioned, neither for charged nor free apps. A closer look at the review content differences between the three price level solution is reached by collapsing all .75 and .99 € price apps into one category resulting in three different price levels: no price, 2.99 €, and the in between price levels with an interval between .75 and .99 €

representing one single category. However, due to the bad matching between observed and predicted values, only very careful first interpretations are given. Highest values for the 2.99 € group were found for missing features, other apps, and network problems; highest values for the .75 to .99 € interval were found for 'improvement request' and 'noise'. Albeit other content might be higher presented in the charged reviews, these are the variables differentiating the two price categories from the rest the most. Concerning the app categories, the category 'facilitation' is well represented by the first dimension, 'entertainment' moderately by the first and bad by the second dimension. 'Information' and 'communication' are well represented by both dimensions and 'social marketing' is moderately represented by the second dimension. It follows that, the proposed 20 items are not able to properly capture the review content of 'social marketing', but the review content of 'facilitation', 'communication' and 'information' in a convincing way. Lastly, hierarchical cluster analysis (method: Ward, distance: squared Euclidean) reveals five different used assessment topics: requests (improvement request, feature removal request, missing feature, content request), complaints (shortcoming, dispraise), compliments (praise, helpfulness, recommendation, compatibility), divers (noise, dissuasion, other app, how to, privacy issues, other feedback, hidden costs), and problems (crash, network problem, unresponsive). Aggregation and dichotomization of each dimensions' sub items result in five binary coded dimensions. Those variables were used to gain insight into the frequency of patterns and their differences between charged and free reviews. Group specific patters out of the 32 possible combinations did not give a clear picture in terms of group differences.

4 Conclusion

In the field of tourism, the integration of app reviews as an important knowledge asset to learn how to enhance tourists' experiences is limited (i.e., Wang et al., 2014). Therefore, this study aimed to explore the various items addressed in app reviews, which can be usable for marketers to design and/or modify their travel enhancing app. In particular, this study analysed the distinction of charged or free apps and its effects on the usability of user-driven app assessments. First, the study illustrated how reviewers dominantly criticize apps when writing reviews compared to the reviews addressing praise. Second, the study highlights differences between app review topics among different app price categories. The study indicated that reviewers who paid for the app also give serious feedback how to improve the app (i.e., feature request). Third, this study demonstrates how reviewers heterogeneously address different topics between various app categories. Fourth, the study demonstrates the categories per app influencing items mentions in reviews. Hence, this study helps marketers who aim to integrate various apps in their product/service/experience design to pay attention to and focus on concrete issues. First, marketers who offer a charged app benefit more from user-driven assessment, as such evaluations help to improve their apps and enrich indirect consumer experiences. Second, based on the review content, free apps receive reviews with a lack of solid reasons. Given that these users do not have to pay for the app, they can easier switch to other apps and/or do not feel the need to help to improve the app. As this study shows, depending on the app price, it is likely to attract different kinds of users and thus reviews. Furthermore, the study shows how app developers should differentiate between review statements that are relevant for the topic category their app is located in. Given the exploratory state of this study, the study was able to indicate the possibilities to explore topics in user-driven assessments to develop accurate functioning apps. Hence, further studies should repeat the method and test it on a bigger sample size. Development of measurement scales capturing the review content of multiple app categories shows to be a promising area for future research. Moreover, Pagano and Maalej (2013) state that the value of app store analyses open up a broad scope of possible future research directions. There is a need to understand how consumers evaluate ICT-enhancing tools in order to continue implementing and improving them for different situations, such as travelling.

Acknowledgements
Many thanks go to Niël Johannes le Roux for his advice regarding the *BiplotGUI* package (la Grange et al., 2009).

References

Ali, R., Solis, C., Omoronyia, I., Salehie, M., & Nuseibeh, B. (2012). Social adaptation: when software gives users a voice. In: *ENASE 2012: 7th International Conference Evaluation of Novel Approaches to Software Engineering*. Wroclaw, Poland.

Fu, B., Lin, J., Li, L., Faloutsos, C., Hong, J., & Sadeh, N. (2013). Why people hate your app: Making sense of user feedback in a mobile app store. In *Proceedings of the 19th ACM International Conference On Knowledge Discovery And Data Mining*, 1276-1284.

Guzman, E. & Maalej, W. (2013). How do users like this feature? a fine grained sentiment analysis of app reviews. In *Proceedings of the 22nd IEEE International Requirements Engineering Conference,* 153-162.

Harman, M., Jia, Y., & Zhang, Y. (2012). App store mining and analysis: MSR for app stores. In *Proceedings of the 9th IEEE Working Conference- Conference on Mining Software Repositories,* 108-111.

la Grange, A., le Roux, N. and Gardner-Lubbe, S. (2009). BiplotGUI: Interactive Biplots in R. *Journal of Statistical Software 30*(12), 1-37.

Khalid, H., Shihab, E., Nagappan, M., & Hassan, A. (2014). What do mobile app users complain about? A study on free iOS apps. *Software, IEEE*, 32(2), 70-77.

Maalej, W., & Nabil, H. (2015). Bug Report, Feature Request, or Simply Praise? On Automatically Classifying App Reviews. In *Proceedings of 23rd IEEE International Requirements Engineering Conference,* (in press).

Pagano, D., & Maalej, W. (2013). User feedback in the appstore: An empirical study. In *Proceedings of 21st IEEE International Requirements Engineering Conference,* 125-134.

Panichella, S., Di Sorbo, A., Guzman, E., Visaggio, C., Canfora, G., & Gall, H. (2015). How Can I Improve My App? Classifying User Reviews for Software Maintenance and Evolution. In *Proceedings of the 31st International Conference on Software Maintenance and Evolution* (*ICSME 2015).* Bremen, Germany..

Tussyadiah, I. P., & Zach, F. J. (2012). The role of geo-based technology in place experiences. *Annals of Tourism Research, 39*(2), 780-800.

Tussyadiah, I. P., & Wang, D. (2014). Tourists' Attitudes toward Proactive Smartphone Systems. *Journal of Travel Research*, 0047287514563168.

Wang, D., Xiang, Z., & Fesenmaier, D. R. (2014). Adapting to the mobile world: A model of smartphone use. *Annals of Tourism Research*, 48, 11-26.