Quantifying User Experience of Mobile Applications Using a Sentimental Analysis Approach

Sameer Kharel, Mikael Fernstrom
Interaction Design Centre, Department of Computer Science and Information Systems, University of Limerick
Sameer.Kharel@ul.ie Mikael.Fernstrom@ul.ie
Keywords: Computer Science and Information Technology

Abstract
Every year the number of mobile applications grow exponentially. Among the many mobile applications, only a few remain competitive. The success of any mobile application depends upon its contextual use and User Experience (UX). Many User Interface (UI) and User Experience (UX) problems have been found using classical methods like Think Aloud, Cognitive Walkthrough and Heuristic Evaluation for mobile applications but literature has shown gaps between the problems found by using these methods. We have tried to find the UI and UX problems by analysing user sentiments towards apps through their text review. Our approach can help app developers/designers to focus on particular UI and UX problems of applications from user text reviews.

1. Introduction
The increasing number of mobile applications on the market give alternatives for users to accomplish their task using apps that provide better UX. An increasing numbers of applications have created competition between application developers. Users can easily search and buy different apps from app stores. App stores and different web site provide users opportunity to share their opinions about apps in text reviews. Aika et al. have used user reviews for finding the “helpfulness” of hotels [1]. Sentiment analysis has been used for text review using a corpus to map words to the topic of interest [2] but have not considered UI or UX. User reviews have also been used to find information about each feature and improvement need for that feature and new requirements of apps [3].

2. Methodology
Our methodology started by selecting six apps; Temple Run, Google Map, Viber, Messenger, Skype and YouTube. These were considered as good apps by reviewing the different information provided on-line. We collected data using the online review tool available from https://www.appannie.com/ for each of aforementioned apps. Data was collected in January 2016 and more than 600 reviews were collected for each app. We used Python to convert reviews into a corpus. Each corpus was analyzed into positive and negative sentiment words. Due to difference in sample size i.e. number of reviews, we used percentages i.e. total words used from reviews for each app divide each word frequency of that app and multiplied by hundred. From the corpus of each app, we listed positive and negative sentiment words. For each app, we summed the percentage of positive sentimental words and negative sentimental words. By doing this we can know which apps have more positive sentiments from users and which have negative sentiments. We found repetition of common positive and negative sentiment words used among reviews of the six apps. Our objective of finding common words was to discover whether apps have common UI or UX problems related to this word use. Additionally, some sentences provided ambiguous results, containing either a positive sentiment with a negative meaning or vice versa. E.g. “I don’t like it doesn’t offer the record a video option anymore”. Here “like” have a positive sentiment but is applied negatively. With “Can’t think of anything bad to say about”, “bad” is a normally negative sentiment words but here has a positive meaning.

For the second stage, we selected two apps that were considered bad apps from Internet sources. User text reviews for these apps were very few compared with good apps. The purpose of this was to find whether users used the same positive or negative sentiment words for text reviews as exemplified by the good apps. For the third stage, we selected two smartphone apps and their web versions and try to compare UI and UX problems between them. At every stage we used this corpus and sentiment analysis approach. At the fourth stage, we analyzed sentences associated with positive and negative sentiment words of two apps and their web version. Our target was to find UI problems associated with positive and negative sentiment words and compare them. To understand the physiology behind each UI and UX problems and to know more about the causes of the problems, we used the Thinking- Aloud method.

3. Discussion
Our research shows that good review apps have a more positive sentiment percentage than a negative sentiment word percentage. Text review sentiment and star rating are not directly related, with a higher star rating than would be expected by the text review, as in Isa et al [4].
From the above figure we can see that not all positively rated apps have a higher percentage of positive sentiment words. YouTube have more negative sentiments words percentage than positive words. YouTube’s star rating was 2 which is satisfactory but sentiment analysis shows users had completely negative sentiments towards it. From figure 1, we can say Temple Run is number one among other apps because it has a higher sentiment words percentage than negative. After that Viber, Google Maps, Skype, Messenger and YouTube. From this, it is clear that users have more positive sentiments for games apps than other. Among our six good apps, only YouTube has more negative sentiment words than positive sentiment words. Results show that highly rated apps may not have more high positive sentiment words than negative. We only got eleven common positive sentiment words from our six apps which are love, great, best, good, awesome, amazing, simple, cool, thanks, enjoy, like, well and happy. Only four negative sentiment words were found common among six apps which are hard, crash, bad and fix. Positive sentimental words like “useful” and “helpful” were found in five apps except Temple Run. This words have been missing from Temple Run due to being in a different domain. The result show that to compare different apps, we also need to consider the domain. We did sentiment analysis of two low rated apps which were American Airlines and Fingerprint Login. The American Airlines Apps positive sentiment words were 0.075% and negative sentiment words 1.4%. Fingerprint Login positive sentiment is 1.5% and negative sentiment words 1.19%. Badly rated apps clearly show that it is not necessary that badly rated apps have a higher negative sentiment words percentage than positive sentiment words. Common positive sentiment words for these two apps are like, thanks, great and glad and negative sentiment words are worst and awful. From common words we can say positive sentiment words for good and bad are consistent but for negative sentiment words it seems to vary.

Sentence analysis showed more diagnosis of problems were needed to find UI and UX problems. To do that we used the Thinking-Aloud method. Before performing Thinking-Aloud we selected two apps and their corresponding web sites to analyse differences in UI and UX related problem. We wanted to see the different and common associated UI and UX problems related. For that purpose we selected Google Maps and YouTube. Our results show Google Maps web version has 1.96% positive sentiment words and 1.06% negative sentiment words percentage. YouTube’s webpage version has 3.04% of positive sentiment words and 0.8% negative sentiment words. From the data we can say YouTube’s website are better than the app, in term of sentiment. Our sentence analysis has shown that web have few UI and UX problems compared with the app version. Thinking-Aloud performed on Google Maps and YouTube has shown that most problems related with web were trivial but problems related with apps are more serious in term of UI and UX problems.

4. Conclusion and Future Work
This study contributes to finding UI and UX problems using sentiment analysis, especially for mobile applications. Results show that user text review using sentiment analysis cannot be directly used for finding deep UI and UX problems related with mobile applications but can be a useful indicator for UI and UX problems. The Thinking-Aloud method was used to further investigation the problems. Our future work will be to related UI and UX problems of apps, collected with the Thinking-Aloud method and further analysed based on Hartson’s User Action Framework (UAF) [5].

5. References