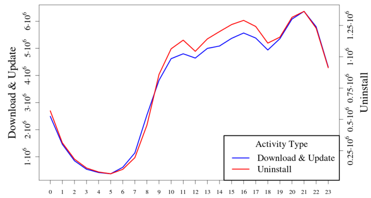

Figure 1: Activities peak in late afternoon and television time.

We hypothesize that users may have a routine schedule of housekeeping the apps on their mobile devices. To further investigate this, we plotted the distribution of the **time intervals** between any two consecutive activities of the same user (Figure 2). Most consecutive activities are conducted within less than a hour, which are likely in the same session. However, when the intervals are larger, there is a peak at every 24 hours. This suggests that a user does have a routine time period of a day for housekeeping - she may not do it every day, but when she does, it is likely to be closer to the same time of a day.

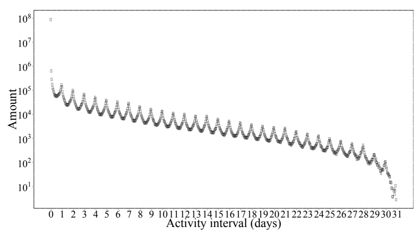

Figure 2: Intervals between activities peak at every 24 hours.

3. INFERRING APP QUALITY

An app is considered to be of high quality if it receives a higher average rating by its users. In practice, however, the online ratings suffer from data sparseness and biases.

We crawled user ratings of all apps from both Wandoujia and Google Play, which is the native Android market place. For the apps that are rated by the users on both marketplaces, we investigated the correlation of the number of ratings an app receives on the two sites, plotted in Figure 3. The number of ratings given to the same app at the two marketplaces are generally positively correlated. However, the noticeable vertical lines in the left and horizontal lines in bottom indicate considerable biases. These data points refer to the apps that have many ratings on one market but very few on the other. For example, the `Facebook` app has 25,169,686 ratings on Google Play, but has only 1,644 rating on Wandoujia.

We then investigated whether the same app receives similar ratings on the two marketplaces. Figure 4 presents a positive correlation between the average *scores* on Google Play and the *likerates* (e.g., number of positive ratings divided by number of all ratings) on Wandoujia for apps which receive at least five ratings on Wandoujia. It seems that the user ratings are overall coherent on the two marketplaces. However, we can also identify many different or even contradictory ratings for the same apps. The correlation is poor if we include apps with few ratings on Wandoujia.

We hypothesize that signals from app management activities may address the biases and sparseness of online ratings. Our goal is to identify weak signals that are good indicators of user preference. We found that "Uninstallation-Downloading (UD)" activities, i.e., a user uninstalls an app and later on re-installs it, may be a good indicator that the user likes the app (that's why he installed it back). To verify this, in Figure 5 we plot the frequency of "UD" patterns among the activities of all users of an app and correlate it with the likerate of the app. Since not all apps have an "UD" activity sequence, we only plot those who have. The average number of "UD" sequences per user is in general positively correlated with the likerate of the app. This is promising, indicates that some user activity sequences may be good indicators of user preferences and app quality.

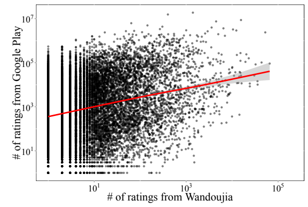

Figure 3: Numbers of user ratings are correlated in different marketplaces, but significant biases exist.

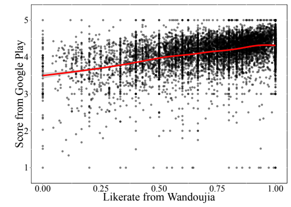

Figure 4: Average ratings at different marketplaces are correlated if there are abundant of ratings.

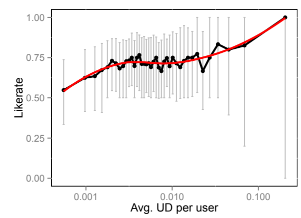

Figure 5: "Uninstallation-Downloading" sequences are positively correlated with Likerate.

4. DISCUSSION

It is encouraging to observe that certain sequential patterns of app management activities are good indicators of user preferences and app quality, which effectively supplement the biases and sparsity of online ratings. Next we will explore how to identify all such indicative patterns from large-scale behavioral data and how to construct a machine learning model that predicts app quality. Our study serves as a preliminary step of understanding individual and collective app using behaviors of smartphone users.

5. REFERENCES

[1] S. Lim, P. Bentley, N. Kanakam, F. Ishikawa, and S. Honiden. Investigating country differences in mobile app user behavior and challenges for software engineering. *IEEE Transactions on Software Engineering*, 2014.

[2] B. Liu. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167, May 2012.

[3] M. Ott, C. Cardie, and J. Hancock. Estimating the prevalence of deception in online review communities. page 201. ACM Press, 2012.