The Need to Account for Geographical Diversities in Mobile Data Research

Akhil Mathur
Bell Labs, Nokia
akhil@bell-labs.com

Fahim Kawsar
Bell Labs, Nokia
fahim.kawsar@bell-labs.com

ABSTRACT

Large-scale mobile data studies have the potential to provide valuable insights regarding usage behavior of smartphone users. A major challenge in generalising the findings of these studies is the inherent population diversity in large-scale smartphone usage data. Many published studies however do not account for population heterogeneities in their analysis, and instead present their findings based on the aggregate usage data. In this paper, we investigate the effects of geographical diversity on smartphone data collected from 130 users from India, Europe and the US over a period of four months. Our results show significant differences in daily usage duration, session-level usage, and temporal usage patterns across various geographies, and stress the need to account for population heterogeneities in mobile data research and its application in real-world systems.

Keywords
Mobile Data, Population Diversity, User Study, Android

1. INTRODUCTION

Mobile phones have evolved from simple communication tools into powerful information, communication and entertainment devices. By the year 2020, it is projected that 5.4 billion people in the world will have a mobile phone, higher than those projected to have electricity (5.3 billion), running water (3.5 billion) or cars (2.8 billion) [2]. Across different app stores, there are more than 4 million applications that the users can download on their phones [1], many of them capable of capturing rich data about users and their environment. By accurately inferring user and environment context from this data, these apps can provide personalised, contextual services to the end-users and substantially improve the user experience.

Consequently, we have seen a number of studies that have collected smartphone usage data at scale, and analysed it to infer usage patterns. Falaki et al. [8] conducted a comprehensive study of smartphone usage to characterise user interactions with the device and various applications, and the impact of these interactions on network and energy consumption. Bohmer et al. [6] extended their analysis to understand contextual factors that influence application usage – for example, they found that “news applications are most popular in the morning and games are at night, but communication applications dominate through most of the day”. Other works have studied the effect of mobility on usage [21], notification delivery and reception on smartphones [15, 17, 23], energy consumption [12], and predicting application usage [25].

A major challenge in generalising the findings of these user studies is the inherent population diversity in large-scale smartphone usage data. In the human-computer interaction community, researchers aim to control for population diversity in their experiment design, e.g., by selecting participants with similar skills or demographics. In mobile data research however, this issue of population diversity is even more critical – as many of the large-scale data collection studies are done by publishing an application on app stores (e.g. Apple Store or Google Play) [6, 15, 23], it indeed becomes difficult to control for user diversity in the data.

Our aim, in this work, is to highlight one particular source of user diversity (viz. geographical diversity) – and how it impacts smartphone usage. While it is not surprising to expect diversity in smartphone usage due to geographical variations, we observe that many published studies on mobile data (e.g. [6, 23, 25]) do not account for such variations even when analysing the data. Other studies have acknowledged the user diversity (e.g. [8]), however they did not provide in-depth analysis of how it affects smartphone usage.

In this paper – by analysing the data gathered in a 130-user longitudinal study, we demonstrate several key differences in smartphone usage between users in India, Europe and United States. We also contrast our findings with past research on smartphone usage in another country (i.e., Korea). We found that there are significant differences across various geographies in terms of daily usage duration, session-level usage, and temporal usage patterns – and if these differences are left unaccounted for, they may result in very misleading findings about real-world smartphone usage. The paper concludes by discussing implications of these geographical variations for mobile data studies, application development, and mobile-based data inference models.

2. RELATED WORK

In this section, we overview two categories of past research that are relevant to our work on investigating population heterogeneities in mobile data.

Mobile Data Studies. Over the last few years, there have been a number of studies on mobile data usage in-the-wild. Falaki et al. [8] conducted a comprehensive study of smartphone usage to characterise user interactions with the phone, application usage, network...
traffic and energy usage. Bohmer et al. [6] analysed contextual factors that influence application usage. Oulasvirta et al. [16] discussed the possibility of mobile users developing a checking habit that involves brief and frequent content consumption.

Various techniques have been proposed [19, 20, 22] for continuously and unobtrusively gathering mobile usage and sensor data. Based on such data, researchers have proposed smart home screens and launcher apps to improve user experience [4, 28] and device battery life [9], as well as developed inference models for predicting the user’s next application [25, 26]. Other works have studied the effect of mobility on usage [21], notification delivery and reception on smartphones [15, 17, 23], and energy consumption [12].

**Data Heterogeneities.** Researchers have identified various hardware and software heterogeneities that impact inference techniques on mobile phones. Blunck et al. [5] discussed how variations in GPS duty cycling across platforms can adversely affect data quality and performance of inference models. Sisten et al. [27] highlighted that run-time factors such as instantaneous I/O load can lead to an unstable sensor sampling rate, thereby compromising the data quality. Das et al. [7] found significant hardware variations in acoustic components such as microphones and micro-speakers of a mobile device.

As a solution to such device heterogeneities, Lu et al. [13] proposed a framework which can identify stress using microphones in diverse acoustic environments. Lu et al. [14] presented a continuous sensing platform robust to accelerometer variations and biases. To account for population diversity in sensor data, Lane et al. [10] looked at incorporating inter-person similarity measurements from crowd-sourced sensor-data into the classifier training process.

### 3. RESEARCH STUDY

In this section, we present our data collection system, and give an overview of the participant demographics and our data analysis methodology.

#### 3.1 System

In order to collect smartphone usage data, we developed an Android application and released it on Google Play. The app, implemented for Android 5.0+, runs as a background service on the user’s phone and passively records all usage sessions on the device. More specifically, the application collects data on system-level permissions and application usage data on their phone. Both these permissions are categorised as system-level permissions in Android, and and must be manually approved by the user from the Settings Menu on their phone. The apps guides the user through this approval process, and once the permissions are granted – the app starts running in the background to collect data logs. The logs are stored locally on the external storage of the phone, and are periodically uploaded to a remote server.

#### 3.2 Participants and Data

After publishing the app on Google Play in December 2015, we advertised it on social networks and email lists to solicit participation. In total, the app was installed by 130 users from 12 different countries. The raw data set obtained from the users until April 2, 2016 contains 40,414 call events, 630,241 notification events, 890,945 application events, and 901,624 screen events.

Out of 130 users, we selected 86 users (25 female) for our analysis who were i) based in our target geographies, i.e. India, Europe or the US, and ii) had kept the app installed on their phone for at least one month, which we believe is a sufficient duration to get representative usage data.

The shortlisted users were mostly young adults aged between 18 to 30 years, and 65% were students. Of the 86 users, 54 were from India, and the remaining 32 were from 6 different European countries and the US. For our geographical analysis, instead of analysing data at a per-country level, we decided to group the users from Europe and US into one category called ‘Western Users’ and compared their usage to ‘Indian Users’. There were primarily two reasons for this choice: i) to ensure that we have sufficient number of users in each comparison group, and ii) Europe and US, although geographically separate, are both developed regions with high smartphone penetration for a number of years. On the contrary, smartphone adoption in India is a relatively recent phenomenon, and the smartphone penetration is much lower [3]. Hence, even with the aforementioned categorisation, it will be possible to contrast the smartphone usage among two diverse groups.

#### 3.3 Data Analysis

Similar to the prior work on mobile usage analysis, we also centred our analysis around a usage session as shown in Figure 1. A usage session is defined as the period between screen-on and screen-off – as such, we processed the Screen Events in the logs to generate usage sessions. On Android, events such as notification arrival and battery charging can switch on the screen automatically – therefore, for an accurate representation of usage sessions, we filtered out the sessions that did not elicit any user interactions (i.e., screen unlocking). In each session, a user can access 0, 1 or multiple applications – as such, we processed the Application Events in the logs to find the apps used in each session.

Smartphone usage sessions can be triggered by external cues (e.g., through a notification or a call), or internal cues (i.e., user motivated). To differentiate the two kinds of sessions, the notification and screen events were processed together; we categorised a session as ‘externally cued’ if i) the session was initiated within 30 seconds of any notification arrival, or ii) if any notifications from the apps used in the session arrived between the current session and the preceding session [11]. All other sessions were categorised as ‘internally cued’.

Finally, before comparing the usage between Indian and Western users, we measured the skewness of each comparison metric (e.g. session duration, session count, application count) within the group, and found that the data distributions for all metrics were within the bounds of a normal distribution \((-1 < \text{skewness} < 1\).)

![Figure 1: Example of a smartphone usage session.](image)
4. STUDY RESULTS

We now analyse the aggregated usage, session-level usage, and temporal usage across the two user groups (Indian users and Western users), and also contrast them with prior research on mobile data usage.

4.1 Aggregate Usage

We begin by comparing smartphone usage of the two groups in our dataset (Indian and Western) through unpaired two-tailed t-test, and report the p value, t statistic and Cohen’s d. In addition, we also contrast the usage of these two groups with the findings of Lee et al. [11] regarding smartphone usage of young ‘Korean’ users. In their study, Lee et al. [11] investigated smartphone addiction among Korean students; as such, they collected smartphone usage data from ‘addicted users’ and ‘regular users’ for nearly one month. For our geographical analysis, we use the data from ‘regular users’ in [11] and compare it to the smartphone usage of participants in our dataset. The participants in our dataset also belong to a similar age group as those in [11] and majority of them are students; therefore, we believe it is reasonable to compare the two datasets.

Table 1: Comparison of aggregate usage across user groups. Results of unpaired two-tailed t-test are provided for groups in our dataset (* indicates significant difference between groups, p < 0.05)

<table>
<thead>
<tr>
<th></th>
<th>Indian</th>
<th>Western</th>
<th>p</th>
<th>t</th>
<th>d</th>
<th>Korean [11]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usage duration (mins) per day</td>
<td>149</td>
<td>113</td>
<td>0.04*</td>
<td>0.51*</td>
<td>1.96</td>
<td>207</td>
</tr>
<tr>
<td>Sessions per day</td>
<td>91</td>
<td>66</td>
<td>0.01*</td>
<td>2.6</td>
<td>0.56</td>
<td>100.1</td>
</tr>
<tr>
<td>Unique apps</td>
<td>80.5</td>
<td>82.4</td>
<td>0.8</td>
<td>-0.15</td>
<td>0.1</td>
<td>NA</td>
</tr>
<tr>
<td>Entropy top-5</td>
<td>3.42</td>
<td>3.55</td>
<td>0.38</td>
<td>-0.87</td>
<td>0.18</td>
<td>1.96</td>
</tr>
<tr>
<td>Entropy top-10</td>
<td>3.89</td>
<td>4.03</td>
<td>0.34</td>
<td>-0.95</td>
<td>0.2</td>
<td>2.53</td>
</tr>
</tbody>
</table>

4.2 Session-level Usage

Table 2 shows the comparison of session-level usage across different groups. Our analysis did not reveal any significant difference in the mean session duration between Indian and Western users ($p = 0.8$), however we observe that Korean users have 30% higher mean session duration than the other two groups. We did observe a significant difference in the inter-session gap between the two groups ($p < 0.02$); results showing that Indian users start a usage session every 9.6 minutes on average as compared to 16.6 minutes for Western users and 13.61 minutes for Korean users.

<table>
<thead>
<tr>
<th></th>
<th>Indian</th>
<th>Western</th>
<th>p</th>
<th>t</th>
<th>d</th>
<th>Korean [11]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean session duration (s)</td>
<td>97.8</td>
<td>95.3</td>
<td>0.8</td>
<td>0.22</td>
<td>0.05</td>
<td>129.9</td>
</tr>
<tr>
<td>Inter-session gap (min)</td>
<td>9.6</td>
<td>16.6</td>
<td>0.02*</td>
<td>2.78</td>
<td>1.07</td>
<td>13.61</td>
</tr>
<tr>
<td>App launches per session</td>
<td>3</td>
<td>2.4</td>
<td>0.01*</td>
<td>2.95</td>
<td>0.58</td>
<td>3.16</td>
</tr>
<tr>
<td>Unique apps per session</td>
<td>1.91</td>
<td>1.71</td>
<td>0.01*</td>
<td>2.71</td>
<td>0.55</td>
<td>NA</td>
</tr>
</tbody>
</table>

Table 2: Comparison of session-level usage across groups. Results of unpaired two-tailed t-test are provided for groups in our dataset (* indicates significant difference between groups, p < 0.05)

Next, we analysed the app launch count in each session (i.e. how many apps were launched in a session, including repetitions). A significant difference was found ($p < 0.005$), and results show that Indian users launch more apps (mean = 3.0) per session than Western users (mean = 2.4). Similarly, we found a significant difference ($p < 0.01$) in number of unique apps used in a session, with Indian users exhibiting a higher unique app count.

4.3 Session Analysis

Now we analyse the nature of individual sessions. As we described in the previous section, we categorised the sessions into two classes: internal and external depending on how they were initiated. Both user groups had a similar distribution of sessions, with external sessions accounting for 17.7% of all sessions for Indian users and 18.1% of sessions for Western users. Lee et al. [11] however reported a very different session distribution for Korean users, where externally cued sessions accounted for 79% of all sessions!

In order to account for varying session durations, we further categorised the sessions into three classes (following the methodology of Banovic et al. [4]): glance, short, long. Glance sessions are those sessions in which no apps are used – here a user may unlock the phone to glance at a notification (or check the time), but does not launch any application. A Short session is less than 60 seconds in duration, and involves the use of 1 or 2 apps, whereas a Long session lasts for more than 60 seconds and may involve 1 or multiple apps. Therefore in total, we had six categories for sessions (3 based on duration and usage x 2 based on initiation type).

<table>
<thead>
<tr>
<th></th>
<th>Internal</th>
<th>External</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Indian</td>
<td>Western</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glance</td>
<td>53.4%</td>
<td>45.9%</td>
<td>8%*</td>
<td>3.2%*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short</td>
<td>11.1%</td>
<td>16.4%</td>
<td>4%</td>
<td>3.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long</td>
<td>15.9%</td>
<td>19%</td>
<td>5.5%</td>
<td>6.2%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Comparison of session types across groups. Cells marked with * indicates a significant difference between groups ($p < 0.05$).

Table 3 presents the session distribution across both groups. We
found a significant difference in the percentage of External-Glance ($t = 2.27, d = 0.62, p < 0.05$) sessions between the two groups. This suggests that on receiving notifications (i.e., external cues), Indian users show a higher preference to glance at its content, rather than opening the application associated with it. We also found a significant difference in the percentage of Internal-Short session ($t = -2.6, d = 0.76, p = 0.01$) between the two groups, suggesting that Western users self-initiate more short duration sessions than Indians.

4.4 Temporal Analysis

Finally, we analyse the temporal usage patterns in a day for all the groups. To facilitate comparison with Lee et al. [11], we divided the day in four equal buckets of six hours, and grouped the smartphone usage by buckets. Figure 2 presents the temporal distribution of session counts and session durations across the user groups.

As evident from the figure 2, we found significant differences both in session duration and session count between 12noon - 6pm and 6pm - 12am (all $p < 0.05$), with Indian users having higher session counts and durations at these times than Western users. Korean users exhibited higher session duration and session count than the other groups during 12noon - 6pm and 6pm - 12am.

5. DISCUSSION

We now discuss the implications of our findings for mobile data studies, and also broadly highlight the challenges for application design and user modelling due to geographical variations in mobile data.

Implications for Mobile Data Research. We present two empirical examples which clearly highlight the challenges posed by geographical diversity for mobile data research. Firstly, we found that the percentage of externally cued sessions for Korean users (79%) was much higher than for Indian (17.7%) and Western users (18.1%). This clearly suggests that app developers should focus on designing better external cues (i.e., notifications) to increase smartphone engagement among Korean users. Now hypothetically, if this analysis on session initiation was done over aggregate data from the three user groups (i.e., without accounting for geographical diversities) – it would have presented a misleading picture on session initiation behavior of the users, and may not have resulted in a concrete design guideline for app developers.

In another example, we look at the large-scale smartphone data collection exercise by Bohmer et al. [6] which attracted participation of users from 20+ countries. In their analysis on the aggregate data, they found that almost half (49.8%) of all smartphone sessions were shorter than 5 seconds. On the other hand, our dataset reveals that the percentage of micro-sessions (i.e., sessions shorter than 5 seconds) was only around 22% for both Indian and West-

6. CONCLUSIONS

The immense growth in the smartphone industry, proliferation of mobile app stores, and availability of software frameworks for continuously gathering mobile usage and sensor data has enabled us to study mobile device usage at an unprecedented scale. However, for mobile app and platform developers to confidently apply these research findings to develop real-world systems, it is extremely important to improve the external validity and generalisability of mobile data studies. In this paper, we showed that geographical diversity in the population can lead to significant differences in smartphone usage patterns, which in turn, can affect the external validity of mobile data studies and even lead to misleading results about user behavior. As such, we urge fellow mobile data researchers to account for population heterogeneities while analysing smartphone usage data or developing mobile-based data inference models. We also hope that our findings on geographical variations in mobile data can trigger interesting discussions at the workshop, which lead to design of better mobile data studies and systems that are robust against population variations.
7. REFERENCES


