Discovering Different Kinds of Smartphone Users Through Their Application Usage Behaviors

Sha Zhao[1,2], Julian Ramos[2], Jianrong Tao[1], Ziwen Jiang[1], Shijian Li[1], Zhaohui Wu[1], Gang Pan[1], Anind K. Dey[2]

[1]Department of Computer Science, Zhejiang University, China
{szhao, taojianrong0, zjuzjw, shijianli, wzh, gpan}@zju.edu.cn
[2]Human-Computer Interaction Institute, Carnegie Mellon University, USA
ingenia@andrew.cmu.edu, anind@cs.cmu.edu

ABSTRACT
Understanding smartphone users is fundamental for creating better smartphones, and improving the smartphone usage experience and generating generalizable and reproducible research. However, smartphone manufacturers and most of the mobile computing research community make a simplifying assumption that all smartphone users are similar or, at best, constitute a small number of user types, based on their behaviors. Manufacturers design phones for the broadest audience and hope they work for all users. Researchers mostly analyze data from smartphone-based user studies and report results without accounting for the many different groups of people that make up the user base of smartphones. In this work, we challenge these elementary characterizations of smartphone users and show evidence of the existence of a much more diverse set of users. We analyzed one month of application usage from 106,762 Android users and discovered 382 distinct types of users based on their application usage behaviors, using our own two-step clustering and feature ranking selection approach. Our results have profound implications on the reproducibility and reliability of mobile computing studies, design and development of applications, determination of which apps should be preinstalled on a smartphone and, in general, on the smartphone usage experience for different types of users.

Author Keywords
Clustering; User groups

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
The number and popularity of mobile applications is rising dramatically [25] at the same time as there is an accelerating rate of adoption of smartphones. Meanwhile, a great number of research studies in recent years have sought to understand users' smartphone application usage behaviors. This has ranged from how individuals download, install and use different applications [13], how many daily interactions they have with apps [6], how long the average application session lasts [1, 3], and how app usage varies with context [5, 7]. In terms of patterns of smartphone application usage, previous work has sought to investigate how often individuals revisit a particular app [11], and which apps are frequently used together [24]. Other work has focused on predicting which applications people are likely to install [31], and use [9, 12, 14, 23].

However, this past work has tended to treat smartphone users as a homogenous user population with similar usage characteristics. This has led to a lack of reproducibility and generalizability in smartphone studies, as highlighted by Church *et al.* [4], who suggest that this could be due to the existence of different user sub-populations among the larger smartphone user population. For example, Jones *et al.* [11], show evidence of the existence of at least 3 different kinds of smartphone users when looking only at their app revisitation patterns. Banovic *et al.* [1] identified 4 types of users by analyzing actions users performed on the emails displayed on their lock screen. We are very motivated by this work but, due to the analysis of a relatively small sample size (n=165 and n=27, respectively), believe that the results may be overly simplistic given the vast diversity in human smartphone behaviors. Based on these previous findings [4, 11, 1], we further investigate a large user population and show evidence of the existence of, and characterize, many (382) different kinds of smartphone users. To our knowledge, we are the first to show the great variety of user types from a larger smartphone user population.

We do not think that smartphone manufacturers, mobile carriers, app developers, and researchers, *etc.* actually believe that there is one (or a few) type(s) of smartphone users, however their actions make it clear that they are making this simplifying assumption. This has had profound consequences on:

- *the way smartphones are made:* All smartphones from a given manufacturer are very similar with slight variations in hardware and price that do not fundamentally change how they are used.
- *the way smartphones are loaded with apps:* For any given brand of smartphone and mobile carrier, they are pre-loaded with a very similar set of apps.

UbiComp '16, September 12-16, 2016, Heidelberg, Germany
 ISBN 978-1-4503-4461-6/16/09…$15.00
DOI: http://dx.doi.org/10.1145/2971648.2971696

- *the way applications are designed:* Applications do not adapt to different kinds of users. At best they have options for multiple languages, and some accessibility functionality.
- *the way researchers design studies, analyze and report results:* Many of the research studies cited above were designed for a general population; however, they may have actually been conducting studies on small groups of different kinds of users [4]. This could potentially explain why results from different studies are difficult to replicate and generalize.

That there are different kinds of users should not come as a surprise: different habits or preferences can result in different *application usage* behaviors [16, 17, 22]. For example, health [16, 17] or mental [22] status of a user can be inferred from location and application usage. Application usage behaviors also depend on demographic attributes to some extent, such as gender, age, occupation, and income. For instance, in one study population, female smartphone users used photography applications much more often than male users, who preferred to use applications related to sports, cars, and news [32]. Similarly, students used learning-related apps more frequently than business people, who used travel and navigation apps more often. It has also been shown that, application usage behavior varies with the age of users [19].

The contributions of this research are three-fold. First, we challenge the commonly-held assumption that all smartphone users are either similar or can be classified into a small number of types. Instead, we show that we were able to identify 382 distinct types of users in a population of 106,762 Chinese smartphone users, based on their application usage. Due to space constraints we focus on the top 3 biggest groups (more than 2300 users each) and 3 smaller groups that have interesting demographics, and characterize these groups based on their app usage behaviors. By doing so, we also assign them descriptive labels: *e.g.*, Night communicators, Screen checkers, Evening learners, Financial users, Young parents, and Car lovers. Each of these groups has very distinctive app usage behaviors. Through greater examination of these groups, we also demonstrate that there is a strong relationship between demographics and app usage.

Our second contribution is our clustering method for identifying types of users and finding meaningful features that enable us to describe what makes each cluster of users different and interesting. We combine k-means and MeanShift clustering and create a performance metric to evaluate results that considers penalties for complexity and non-uniform distribution of users across clusters. We also introduce our own feature ranking scheme that helps us identify meaningful differentiating features for the selected clusters.

Our third contribution is to explore the implications of our work. Based on our findings on different types of users, we provide advice for researchers to help them better focus their research and analyses on the different types of users. We also propose some design ideas for smartphone manufacturers, mobile carriers and app developers, so as to help them tailor better smartphone experiences for these different kinds of users.

In this paper, we demonstrate how to identify different types of users from a dataset from the smartphones of 106,762 users from multiple provinces in China. For each smartphone, the dataset contains hourly updates on the 10 most recently used apps, for the month of September 2015. To identify different types of users, we first extract the app usage records from the dataset and then place the used apps into different app categories. We then discover different groups of users by clustering based on the similarity of the app category usage behaviors among users. For each cluster, we identify the most characteristic behavior(s) of the users and give the cluster a representative label. Finally, based on our findings we provide recommendations for researchers to improve their analyses and generalizations and for smartphone manufacturers, mobile carriers, and app developers to improve the user experience.

RELATED WORK

As mentioned above, a growing number of analyses in recent years have sought to investigate how individuals use applications on smartphones. Some have shown great diversity in app usage [6, 29], however, they have focused on exploring and reporting averages or ranges of app use across users. Many other studies of mobile application usage [10, 21] have focused on specific user groups, since collecting large-scale user data is challenging. This has resulted in a lack of exploration about the simplifying assumption that all smartphone users can be treated similarly.

Falaki *et al.* [6] used detailed traces from 255 users to characterize smartphone usage from two intentional user activities: user interactions and application use. They did not only explore the average case behaviors, but also explored the range seen across users and time. They found that the mean number of interactions per day for a user varies from 10 to 200, the mean interaction length varies from 10 to 250 seconds, and the number of applications used varies from 10 to 90. They found that users fell along a continuum between the extremes, rather than being clustered into a small number of groups. Their study also showed that demographic information was not a reliable predictor of application usage behavior, such as application popularity and interaction time.

Xu *et al.* [29] investigated diverse usage patterns of smartphone apps via network measurements from a national level tier-1 cellular network provider in the U.S. They presented aggregate results on correlations between spatial and temporal factors and usage. They found that some apps have a high likelihood of co-occurrence across smartphones. They also found that the diurnal patterns of different genres of apps can be remarkably different. For example, news apps are much more frequently used in the early morning, sports apps in the evening, while other apps have less visible diurnal patterns and their usage is more flat during a day. Some apps are more frequently used when users are moving around, while others are prevalent when users are stationary. This work showed how, where, and when smartphone apps are used at an aggregate level but did not explore how app usage differs across users or user groups.

Jesdabodi *et al.* [10] segmented usage data, which was collected from 24 iPhone users over one year, into 13 meaningful clusters that correspond to different usage states, in which users normally use their smarphone, *e.g.*, socializing or consuming media. They analyzed how the usage states differ within and between the users. It was found that the 24 users had, on average, 13.7 usage states. They identified communication apps that were present in all usage states, which means that users are likely to use apps like SMS and Email in all states. This study focused on characterizing the usage states that exist between usage sessions rather than analyzing application usage behavior of different kinds of users.

Yan *et al.* [30] found that the majority of mobile device usage is brief: 50% of mobile phone engagement (the time period between the user unlocking and relocking the device) lasts less than 30 seconds, and 90% of engagements last less than 4 minutes. Based on a study of usage behavior from 4,100 android users, Böhmer *et al.*'s large-scale study on mobile application usage revealed that mobile phone owners use their device for an average of 59.23 minutes daily, with the average application session lasting 71.56 seconds [3]. Neither of these studies explored how application usage varies across different kinds of users. Other studies [21, 22] have similarly looked at app usage (*e.g.* the length of usage sessions, the number of sessions per hour or per day, the number of apps per day) at an aggregate level. However, they have not identified the differences in app usage behaviors among different user groups.

All of the above work analyzed smartphone users and their application usage in the aggregate. There have been relatively few pieces of work that have considered differences in smartphone users. For example, Church *et al.* show that there are great discrepancies and an inability to generalize results across different research studies [4]. More specifically they say about results from past research studies: "work of this nature is not generalizable beyond the given population and again should be interpreted as such." In general, they suggest that mobile application behavior is very specific to different populations.

Jones *et al.* [11] identified three distinct clusters of users based on their app revisitation patterns, by analyzing three months of application launch logs from 165 users. App revisitation refers to how often users return to a particular app. The three clusters are: Checkers who exhibit brief but quick revisit patterns (in less than an hour), Waiters who are split between short-medium revisitations (between 1min and 4hrs) and long revisitations (2hrs–3days), and Responsives who exhibit sometimes brief and sometimes long revisit patterns. Banovic *et al.* [1] identified four types of users by analyzing 27 users' actions on emails displayed on their lock screen. The four types of users are: Non-users who reacted to very few emails, Normal users who reacted to their emails occasionally, Power users who reacted to more than a third of their emails and Cleaners who tried to keep their inbox free of any unread messages. Both studies [11, 1] are promising and very motivating for us in suggesting that there are at least 4 kinds of users, however their scope is limited by only examining a relatively simple behavior (revisitation for [11] and actions performed on emails for [1]) for relatively small populations. We hypothesize in our work that there are many more complex and diverse behaviors that make up the smartphone user population.

To summarize, although there have been many studies on mobile application usage behaviors, they have either only scratched the surface at finding different kinds of users, analyzed application usage in the aggregate or explored the range across users. They have not explored any differences in application usage behavior between groups of users. They mostly treat all smartphone users as similar, which is a nice simplifying assumption for app developers, phone manufacturers and mobile carriers, except that it does not reflect reality. In this work, we analyze app usage behaviors of 106,762 Android users and demonstrate that there are many different types of users (382) with vastly different application usage behaviors. We will now describe our dataset and how we analyzed it.

DATA OVERVIEW

The dataset we use to identify user groups contains lists of recent apps used on Android smartphones, provided by a mobile Internet company in China. It contains 106,762 unique smartphones and 77,685 unique apps from Sep. 1st, 2015 to Sep. 30th, 2015. Among the 30 days, there were 9 holidays including weekends (Sep. 3rd-5th, 12th-13th, 19th-20th, 26th-27th), and the rest were workdays. The data was collected approximately every hour using the function *ActivityManager.getRecentTasks()*. It returns a list of the tasks that the user recently launched, ordered from most recent to oldest. The dataset consists of 52,872,129 usage records in total. A sample of the dataset is shown in Table 1, with each record consisting of a:

- User ID: the unique identity of the sampled smartphone. Each user ID is anonymized for security and privacy reasons before the data is collected.
- Time stamp: the time when the list of tasks was collected.
- List of recent tasks: each list consists of up to 10 package names that can be used to identify an app.

Table 1. Sample of lists of recent app tasks in the dataset

User ID	Time	The List of Recent App Tasks
0000751aecb005a2	2015/9/1 9:09	com.android.calendar, com.tencent.mobileqq, com.moji.mjweather
0000751aecb005a2	2015/9/1 10:09	com.miui.home, com.android.incallui, com.android.calendar, com.moji.mjweather

Basic analysis

To give a sense of the richness of the dataset, about 60,000 users have 30 days of data from Sep. 1st to Sep. 30th, and about 90% of users have more than 20 days. 25% of the days have 24 records (*i.e.*, complete data collection), and about 80% have more than 15 records. Each record can contain 1 to 10 data points, and about 30% of the records consist of 10 apps.

Demographic attributes

Demographic data about each user was collected, including gender, income level, and age range. There were three income categories: low income (monthly income $\leq$ 3,000 CNY (460 USD)), high income (monthly income $\geq$ 10,000 CNY (1,535 USD)), and medium income. There were four age categories: 0-17, 18-24, 25-34, and 35+.

APPROACH TO IDENTIFYING USER GROUPS

To identify user groups, there are a number of steps we must perform. First, we perform a *preprocessing* step, in which we extract the apps used from the dataset records and weight them by how much they were used. Further, because the number of apps in the dataset is so large, we reduce the dimensionality of the dataset by grouping apps into semantic app categories. The final step of preprocessing is to filter the data to those users who were active in using apps during the data collection. We then proceed with *clustering* the data into unique user groups based on their app category usage. We describe our clustering approach and metrics for optimization of the clustering, and then provide categories of features that can be used to describe and distinguish the user clusters.

PREPROCESSING

To discover different kinds of users, we cluster users based on similarities and differences in their app usage. To this end, we need to preprocess the data, including extracting the app usage records from the recent app task lists.

Extracting the users' app usage records

As each usage record shows the recent tasks in the past hour, by comparing two consecutive records, we can identify the change in app usage, such as launching new apps, re-opening an already used app, and killing old apps. Here, old apps refer to those apps that only appear in the former record, and new apps are those that only appear in the later record. If two consecutive records are identical, we assume that the user has not used any apps in the past hour. This occurs often in our dataset in the period from 0am to 6am, when most of our users are presumably asleep.

To detect which apps were used in the past hour, we perform a comparison of two consecutive lists. Since we do not know in which order each list changes, we focus on the difference between the two lists. For instance, the lists collected at 9:09 and 10:09 on Sep.1st, shown in Table 1, are compared to detect which apps are used in this time period. We scan the list L_2 collected at 10:09 from the last app *com.moji.mjweather*. We scan the list L_1 collected at 9:09 starting from the last app until we find *com.moji.mjweather* and label its index. In the second round, we scan L_1 starting before *com.moji.mjweather* to find *com.android.calendar*, the second-last app in L_2, and then label its index in L_1. This process will be repeated until the apps in L_2 are all scanned or we cannot find any of the remaining apps in L_1. Here, we stop when we cannot find *com.android.incallui* (the third-last app in L_2) before the labeled app *com.android.calendar* in L_1. All unlabeled apps before the first labeled app in L_1 represent the apps that the user killed, and all apps before the last scanned app (including the last scanned app it cannot be found in L_1) in L_2 are the used apps. Thus, the only killed app is *com.tencent.mobileqq*, and the used apps are *com.miui.home* and *com.android.incallui*. In the case in which there are no labeled apps in L_1, we take all the apps in L_1 as the killed apps and all the apps in L_2 as the used apps.

For each used app, we also calculate its usage weight for each hour slot (what percentage of the hour it was used). If there are more than 120 minutes between consecutive records, we only calculate the usage weight of apps over the final 120 minutes, and assume the phone was turned off for the earlier time period. For example, if a list is collected at 0:15, and the next is collected at 7:10, we only calculate the usage weight of the apps from 6am to 7am, and 7am to 8am. Table 2 shows how to calculate the usage weight.

Table 2. The calculation of the usage weight (h_1, m_1 are the hour and minute of the time stamp for the 1st list, and h_2, m_2 are the hour and minute of the time stamp for the 2nd list

Time of the 1st list	Time of the 2nd list	Hours	Usage weight
$h_1.m_1$	$h_1.m_2$ $(T < 120)$	$h_1 \sim (h_1+1)$	1
$h_1.m_1$	$(h_1+1).m_2$ $(T < 120)$	$h_1 \sim (h_1+1)$	$(60-m_1)/(60-m_1+m_2)$
		$(h_1+1) \sim (h_1+2)$	$m_2/(60-m_1+m_2)$
$h_1.m_1$	$(h_1+2).m_2$ $(T < 120)$	$h_1 \sim (h_1+1)$	$(60-m_1)/(120-m_1+m_2)$
		$(h_1+1) \sim (h_1+2)$	$60/(120-m_1+m_2)$
		$(h_1+2) \sim (h_1+3)$	$m_2/(120-m_1+m_2)$
$h_1.m_1$	$h_2.m_2$ $(T \geq 120)$	$(h_2-1) \sim h_2$	$(60-m_2)/60$
		$h_2 \sim (h_2+1)$	$m_2/60$

Table 3. A sample of one user's app usage records

User ID	Date	Hours	Used apps	Weight
0000751aecb005a2	2015-09-01	09-10	com.miui.home	0.85
0000751aecb005a2	2015-09-01	09-10	com.android.incallui	0.85
0000751aecb005a2	2015-09-01	10-11	com.miui.home	0.15
0000751aecb005a2	2015-09-01	10-11	com.android.incallui	0.15

Table 3 shows a sample of one user's app usage records extracted from the sample lists in Table 1, which contains the usage weight of the used apps in each hour slot. For each app, we calculated the sum of the usage weights across all users. The top 5 most frequently used apps are WeChat, Phone, QQ (an IM client), Contacts, and SMS, which are all used for communication and social activities. Recent work by Church *et al.* [4] found similar results: Facebook, Contacts and communication apps like SMS, Phone and WhatsApp were in their list of most frequently used apps. Each of the 77,685 apps was used by 53 users on average. There are 27,779 apps for which the sum of the usage weights is greater than 10.

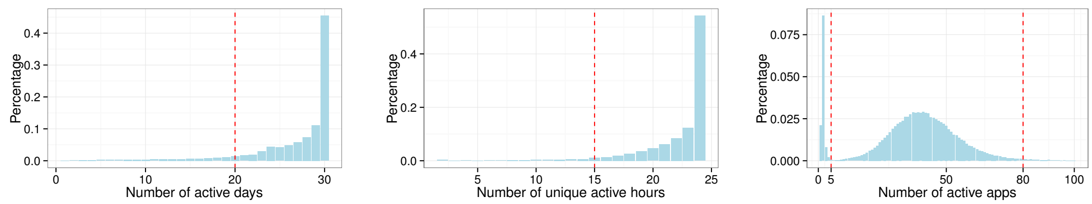
Categorizing apps

Due to the very high number of distinct applications in our dataset, it is impossible to compare directly application usage behaviors from the applications themselves. Instead, we used application categories for this comparison. Our reasoning is that users that are using apps in the same category (*e.g.*, Social) are more likely to be involved in a similar activity than a user using apps in different categories (*e.g.*, Stocks and Education). The use of app categories sacrifices precision (the name of the specific app used) in comparing across different users and obtaining a more general and less complex representation of the user behavior. App categories are useful because they have an inherent semantic meaning, *e.g.*, News, Games, and Banking, that allow us to reason more easily about phone usage than using the name of the apps alone.

We crawled the categories of the apps from appstores. The categories are largely specified by the apps' developers; as domain experts, they assign their apps to categories when uploading to appstores. For each app that was assigned to different categories in different appstores, we manually chose the category according to the app's core function. In addition, we performed some minor manual modifications. For example, we merged apps that can make people's lives easier (*e.g.* flashlight, calculator, note, and express delivery service) into a single *Lifestyle* category. Apps distributed via channels other than appstores (*e.g.* pre-installed by device manufacturers, or part of Android), or apps that have been removed from

Table 4. The number of apps in each category and some examplary apps

No.	Category	Apps	Examplary Apps	No.	Category	Apps	Examplary Apps
1	Weather	671	MoWeather, ZhiquWeather	**16**	Photography_and_beauty	1,954	Beauty Camera, Polaroid, 360Camera
2	Clock	286	ZDClock, Alarm	**17**	Education	3,280	EnglishStudy, DrivingTest
3	Calendar	418	HandCalendar, Chinese Calendar	**18**	System tool	7,128	WiFi, Settings, TencentManager
4	Theme	2,165	GODesktop, LoveWallpaper	**19**	Phone_and_SMS	1,160	SmartCall, HandcentSMS, Contact
5	News_and_reading	3,455	NeteaseNews, BaiduReader,	**20**	Car	445	CarMegazine, Mobile4s, CarNews
6	Browser_and_searching	336	UCBrowser, QQBrowser,	**21**	SON_and_IM	2,357	QQ, WeChat, SinaWeibo, Momo, Fetion
7	Business	1,241	NameCard, Mail, WPS Office,	**22**	Shopping	2,040	Taobao, BeautyShopping
8	Navigation	560	AMap, GoogleMap, Compass	**23**	Parent_and_child	1,665	Children's song, HappyParenting
9	Travel	387	Qnar, HotelBooking, XiechengTravel	**24**	Game_card_and_chess	2,776	FightTheLandlord, ChineseChess
10	Transportation	905	Train12306, 8684Bus, Uber	**25**	Game_casual_and_puzzle	16,873	EliminationGame, Fruit Ninja
11	Health_and_fitness	1,321	Peroids, Yoga, DingxiangDoctor	**26**	Lifestyle	4,143	PinkNote, Flashlight, Calculator
12	Finance	1,739	AliPay, MobileBanking	**27**	Game_other	11,481	SBgameHacker, GameKiller
13	Stock	661	Tonghuashun, Stock, Dazhihui	**28**	Launcher	287	NubiaLauncher, AndroidLauncher
14	Media_and_video	2,877	Youku, iQiYi, PPTV, TencentVideo	**29**	Others	1,627	Dabao, AndroidKing, BBML
15	Music_and_audio	1,869	KugouMusic, TTPODMusicPlayer				



(a) (b) (c)

Figure 1. (a) Quantity of users with respect to the total number of active days; (b) Quantity of users with respect to the number of active hours; (c) Quantity of users *vs.* total number of used apps.

the marketplace, had no categories. We manually categorized them based on their core functionality. We had a total of 29 app categories, shown in Table 4.

We manually categorized the 27,779 frequently used apps for which the sum of the usage weights is greater than 10, into the 29 categories. It would take us an enormous amount of time if we categorized all the apps manually. Thus, we trained an efficient category classifier which can automatically classify the remaining 49,906 more infrequently used apps into the 29 categories. We used the following features in our classifier: 1) average usage weight in 24 hours on holidays and weekdays separately; 2) category labels crawled from appstores; 3) app package names; 4) latent semantic topics learned from the description information of an app (from appstores) by leveraging Latent Dirichlet Allocation (LDA) [2]. As input to LDA, we used Jieba, a tool for Chinese text segmentation, to select the nouns to represent each app. We used an SVM classifier with a radial basis function and a five-fold cross-validation policy. The average accuracy for the 29 categories is 75.41%. After categorization, each category has about 52,293 users on average. Both the sparsity and dimensionality are reduced to a great extent by using app categories rather than apps. The usage weight of each category is calculated by summing the usage weight of the applications in the category. Intuitively, using these category labels for user clustering should obtain better performance than using the apps, and allow us to better reason about app usage and different kinds of users.

Data filtering

We focused our analysis on users who used their smartphones more frequently. We removed those with fewer than 20 active days, where an active day is one with at least 15 app usage records. We also computed the unique active hours in which users used at least one app, and we removed users (3,594) with fewer than 15 unique active hours, shown in Figure 1(b). Fewer unique active hours mean the users do not use their smartphones as frequently as others. We also removed outlier users: those that used fewer than 5 or more than 80 apps in total over the month of data collection, shown in Figure 1(c). After filtering, there are 89,926 users remaining, and 76,107 unique apps.

The demographics for these remaining users is shown in Table 5. There are more female than male users in our dataset (59% *vs.* 41%). Only 5% of users are in the age range of 0-17, with most users being adults. The income levels of the users are almost evenly distributed.

Table 5. The proportion of users in the dataset in each demographic

Proportion of users in each demographic attribute								
Gender		Age range				Income level		
Female	Male	0-17	18-24	25-34	35+	Low	Medium	High
0.59	0.41	0.05	0.37	0.36	0.22	0.31	0.38	0.31

CLUSTERING USERS

Users' app usage representation

We represent each user's app usage using the categories and average usage weight in different time periods. We divide each day into four time periods: *i.e.*, night (0:01am to 6:00), morning (6:01 to 12:00), afternoon (12:01 to 18:00), and evening (18:01 to 0:00). We chose the 6-hour time periods as a balance between dimensionality and meaningfulness. We treat holidays and weekdays separately. Thus, each user is represented by a vector of 29 (*categories*) $\times$ 4 (*time periods*) $\times$ 2 (*holidays and workdays*) for a total of 232 dimensions. For example, Stock_W06-12 means the usage of apps in the Stock category during morning hours on workdays.

In order to capture behavioral patterns of application usage, we normalize the value of each category-time interval by dividing by the sum of the usage in each of the 4 time periods. Then, the value of each dimension (usage weight) in our vector representation is the usage percentage of one category in a corresponding time period.

Clustering approach

Clustering methods are in general very useful at automatically finding clusters of data points with similar characteristics in an n-dimensional space. To cluster users into different app usage groups, a challenge is in determining the appropriate number of natural clusters. Some clustering methods require the number of clusters to be set a *priori* before the clustering takes place, but these methods are fast to execute. There are, however, some methods that automatically find the number of clusters, but they are usually computationally expensive. In this work, we use these standard clustering techniques but combine them in a novel way. We leverage the fast clustering methods to reduce the complexity of the clustering problem to a manageable size and then apply a second clustering method to those results that automatically determines the final number of clusters. More specifically, we make use of k-means with a pre-specified number of clusters and then cluster the centroids found using MeanShift. To select an appropriate number of clusters for k-means, we can simply select a number that is significantly larger than the number of natural clusters, since the MeanShift step will merge centroids generated by k-means to match the natural clusters. The more computationally complex MeanShift clustering step can be performed quickly as its input data is much smaller than the original dataset.

Measuring clustering performance

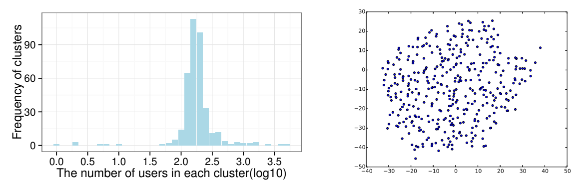
Using the k-means-MeanShift hybrid method, we clustered our dataset using different values for k (the number of clusters for k-means) and *bandwidth* (MeanShift parameter), respectively. In order to evaluate the quality of the clustering results from the different parameters used, we defined a clustering performance (cp) score by weighting four factors as shown in Equation (1). Although metrics for measuring clustering performance exist (*e.g.*, Silhouette coefficient [15], Gap Statistic [26], and Dunn's index [18]), we resort to our own as it captures properties of the clustering results that are not usually considered in more conventional metrics like penalizations both for complexity and non-uniform distribution of users across clusters. In our cp score,the first and the second factors are used to reward the clustering performance using two well-known metrics: Shannon's entropy (E) [20] and Dunn's index (D) [18]. The probability used for calculating Shannon's entropy score is the normalized number of users in each cluster. Thus, entropy assigns a high value to clustering results that have a uniform distribution of users across clusters (and we weight this factor slightly more heavily than the others). Dunn's index, on the other hand, measures the compactness and separation of the clusters obtained. The third and the fourth factors are for penalizing clustering results. In particular, we penalize complex results: those that do not improve over the k-means results where m (the number of final clusters from MeanShift) is close to k (the number of clusters found by k-means). We also penalize results that result in non-uniform distribution of users across clusters: particularly those in which the biggest cluster contains most of the users in the dataset where N is the number of users input to k-means, and n is the number of users in the biggest cluster after applying MeanShift.

$$cp = 0.3 \cdot E + 0.23 \cdot D + 0.23 \cdot \frac{k-m}{k} + 0.23 \cdot \frac{N-n}{N} \quad (1)$$

By combining these four factors, we guarantee that the resulting clusters are compact, well separated and the number of users in the biggest clusters are well distributed (*i.e.*, a single cluster does not contain most of the users).

RESULTS

Using the cp score shown in Equation (1), and trying several values for k and *bandwidth*, we obtained the best clustering result with the highest cp score (0.61) for: $E = 0.9275$, $D = 0.1988$, $m = 382$, and $n = 4,981$. This result was obtained for$k = 500$, and $bandwidth = 1.0$, resulting in 382 clusters. Figure 2(a) shows the quantity of clusters with respect to the number of users. As we can see, most clusters consist of 100-300 users (326 clusters). There are 9 clusters consisting of more than 1,000 users. The biggest cluster has 4,981 users.



(a) (b)

Figure 2. (a) Quantity of clusters with respect to number of users; (b) t-SNE representation of the centroids.

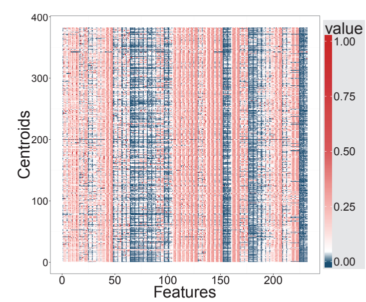


Figure 3. The heatmap of the 382 centroids with 232 features.

For each cluster, the value of each dimension is the average value of the corresponding feature for the users in the cluster. Due to the high number of dimensions of the resulting clusters, we used the t-SNE [28] transformation to visualize the 382 clusters with 232 features, shown in Figure 2(b). t-SNE is a data transformation method that calculates the distance between high dimensional data points, and uses the distance to plot the points on a 2-D plane. The centroids appear to be nicely separated without any visible clusters, giving a visual indication that our clustering approach was successful.

However, when we visualize the 382 clusters with the values for all 232 features (Figure 3), it is not clear how to appropriately describe the differences between the clusters. The dimensionality (232) is just too high to make sense and apply meaning to the different clusters. As we can see, the values of most features are strikingly similar across all the clusters, which makes it impossible to visually distinguish them. Additionally, although 232 features may not be considered a high dimensional dataset for machine learning algorithms, it is too large a set for characterizing a group of people in any succinct fashion. Thus, we performed an additional feature selection step to find the most salient *general* features (across all clusters) and *individual* features (what distinguishes a given cluster from all the others).

Feature selection

To better understand the clusters created by our method, we need to select a small subset of our 232 features from the resulting clusters. In general, we want to find distinctive features that distinguish the clusters from one another. To identify these distinctive features, we use two feature selection strategies: General features and Idiosyncratic features.

General features

General features are those that have the highest variation across the clusters and hence, may help to distinguish them. To capture this variation, for the clusters obtained, we compute Shannon's entropy [20] for each feature separately. The probability was calculated by discretizing the feature values and then calculating the probability for each value.

Idiosyncratic features

Idiosyncratic features are those that help distinguish each cluster from the "average user", a fictional user whose usage characteristics are equal to the average of all users in the dataset. The features were ranked using Equation (2)

$$r_{i,j} = \frac{c_{i,j} - \mu_j}{Max_{j \in J}(|c_{i,j} - \mu_j|)} \tag{2}$$

where $c_{i,j}$ is the ith centroid's value on the jth feature, and μ_j is the average value of the jth feature for all the users in the dataset.

Analyzing clusters by ranking features

For each cluster, we compared the cluster value of each selected feature to the average value in the dataset. The clusters were organized according to how many users are in them with Cluster1 representing the biggest cluster.

General features

We ranked the general features according to their entropy values, and selected the top 5 features with the highest entropy: *Stock_H06-12* (0.8563), *Shopping_H18-24* (0.8421), *Parent_and_child_W06-12* (0.8216), *Parent_and_child_H06-12* (0.8177), and *Business_H12-18* (0.8165). *Stock_H06-12* has the highest entropy, which means the users in the dataset have highly unpredictable usage behaviors of stock-related apps during the morning hours of holidays.

Idiosyncratic features

We ranked the idiosyncratic features in descending order according to $r_{i,j}$ (shown in Equation (2)) for each cluster. We listed the top 5 features with values higher than the average user, as well as the top 5 features with values lower than the average user (*i.e.*, lowest scoring features).

Analyzing the top 3 biggest clusters

Due to space constraints, we can only show our analysis of the 3 biggest clusters ranked by size, using the general and idiosyncratic features.

Cluster 1: Figure 4 shows our analysis of **Cluster1**, which has 4,981 users. There are slightly more females than males in this cluster, with about 40% of users being older than 35, and about 90% of users having low and medium income levels.

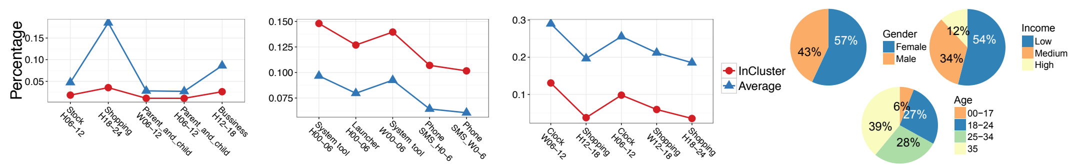
As we can see, the cluster values of *all* the general features are smaller than the average values, especially for the features of *Shopping_H18-24* (0.0354 *vs.* 0.1853), *Business_H12-18*, and *Stock_H06-12*.

Its top 5 highest idiosyncratic features are *System tool_H00-06*, *Launcher_H00-06*, *System tool_W00-06*, *Phone_and_SMS_H00-06*, and *Phone_and_SMS_W00-06*, respectively. Compared to all users in the dataset, the users in Cluster1 use categories of phone and SMS, launcher, and system tool more frequently during the night hours from 0am to 6am. Looking at all users in our dataset, we see they use phone and SMS more often during afternoon hours both on holidays and workdays, with the average values of *Phone_and_SMS_H12-18* and *Phone_and_SMS_W12-18* being 0.3414 and 0.3467, respectively. The values of these two features for Cluster1 users are smaller (0.2968 and 0.3055) than the average values. However, for the same app category, but for the night hours, *Phone_and_SMS_H00-06* (0.1070 *vs.* 0.0643) and *Phone_and_SMS_W00-06* (0.1016 *vs.* 0.0606), the values of these two features for Cluster1 users are greater than the average values. Thus, we can label these users as **Night communicators**. In addition, these users are distinguished by their relatively rare use of clock apps in the morning, shopping apps during the afternoon, and music and audio apps on holiday evenings (when compared to the average values across all users).

Cluster2: Figure 5 shows the analysis and summary for **Cluster2**, which has 3,814 users. This cluster has an even gender distribution, with the largest proportion of users being between 25-34 years old (41%) and half of the users having low income. The cluster values of the top 5 general features are smaller than the average values, especially *Shopping_H18-24* (0.0945 *vs.* 0.1853) and *Business_H12-18* (0.0289 *vs.* 0.0862).

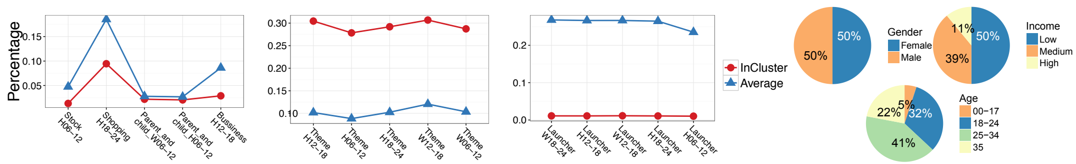
Among the top 5 highest idiosyncratic features shown in Figure 5, the greatest difference in usage behavior between these users and all users is the usage of *Theme* apps, in all time periods for both holidays and workdays. The Theme category contains screen locker, screen protector, desktop, background, and wallpaper apps. Using these apps, users can change their desktop or wallpaper, and lock or unlock the screen. Screen locker apps appear when users wake up the phone just to check the time, or notifications rather than unlock the screen. Users in Cluster2 rarely use the launcher category. The launcher process appears only after a user unlocks the phone screen and goes to the main interface/home screen. These users frequently wake up their smartphone but rarely unlock the screen and enter the main interface. We hypothesize that they are waking up their phones just to check the time or to see if there are any notifications, which has been studied in [8, 27]. Thus, we label these users as **Screen checkers**.

Cluster3: Figure 6 shows our analysis of **Cluster3**, which consists of 2,384 users. There is a greater proportion of female



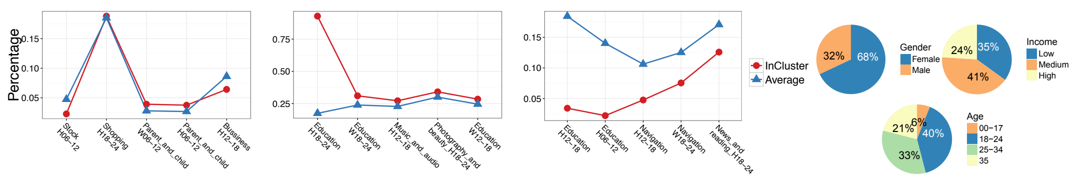
(a)Top 5 general features (b)Top 5 highest idiosyncratic features (c)Top 5 lowest idiosyncratic features (d) Proportion of users in each demographic attribute

Figure 4. Cluster1 (4,981 users)



(a)Top 5 general features (b)Top 5 highest idiosyncratic features (c)Top 5 lowest idiosyncratic features (d) Proportion of users in each demographic attribute

Figure 5. Cluster2 (3,814 users)



(a)Top 5 general features (b)Top 5 highest idiosyncratic features (c)Top 5 lowest idiosyncratic features (d) Proportion of users in each demographic attribute

Figure 6. Cluster3 (2,384 users)

users in this cluster (68%). The largest proportion of users is between 18-24 years old (40%) and 76% of users have low and medium income levels.

For the general features, the cluster values of *Stock_H06-12* and *Business_H12-18* are smaller than average values. However, the cluster values of *Shopping_H18-24*, *Parent_and_child_W06-12* and *Parent_and_child_H06-12* are greater than the average values, especially the usage of Parent_and_child apps during morning hours on workdays (0.0391 *vs.* 0.0279) and holidays (0.0375 *vs.* 0.0268).

The top 5 highest idiosyncratic features are *Education_H18-24*, *Education_W18-24*, *Music_and_audio_H12-18*, *Education _W12-18*, and *Photography_and_beauty_H18-24*. The value of *Education_H18-24* (0.9286) is substantially higher than the average value (0.1741). They also use education applications much more often during evening hours on workdays (0.3101 *vs.* 0.2386). We call the users in Cluster3 **Evening learners**. They rarely use education apps during afternoon hours (0.0342 *vs.* 0.1844) and morning hours (0.0221 *vs.* 0.1402) on holidays. They also use the categories of music_and_audio and photography_and_beauty more often during afternoon hours and evening hours of holidays, respectively. They are different from the average users, in their more infrequent use of navigation apps during afternoon hours of holidays and evening hours of workdays, and news_and_reading apps during evening hours of holidays and afternoon hours of workdays.

Analyzing some smaller clusters

In addition to these large clusters, we also found some smaller yet interesting clusters composed of mainly male or female users. We selected 3 smaller clusters (fewer than 300 users) by looking at the demographics and more specifically those that were composed of mostly one gender. Analyzing the app usage behaviors of these clusters is helpful for determining whether gender (and other demographics) impact app usage or not. Below we describe these clusters.

Cluster353 consists of 113 users, of which 75% of the 113 users are male. We label this cluster as **Financial users**. For the general features, the cluster values of *Stock_H06-12* and *Business_H12-18* are much greater than the average users, especially *Stock_H06-12* (0.4388 *vs.* 0.0474). It means these users have a strong preference in stock-related apps.

The top 5 highest idiosyncratic features are *Stock_H06-12* (0.4388 *vs.* 0.0474), *Navigation_W06-12* (0.4314 *vs.* 0.1062), *Game_casual_and_puzzle* (0.4548 *vs.* 0.1402), *Stock_W12-18* (0.3770 *vs.* 0.0808), and *Stock_W06-12* (0.3326 *vs.* 0.0474). These users use stock apps more often in the morning on holidays, and during morning hours and afternoon hours on workdays. They also use navigation apps more often on workday mornings (presumably to commute to work). They play casual and puzzle games for entertainment in the afternoons on holidays, and use weather apps more often in the mornings on holidays. The users in this clusters, do not use very often the categories of browser_and_search, education, theme, lifestyle and media_and_video during night hours of workdays.

Cluster219 consists of 164 users, of which 88% are female users, and 72% are in age range from 25 to 34. For the general features, the cluster values of *Stock_H06-12*, *Shopping_H18-24*, *Parent_and_child_W06-*

12, and *Parent_and_child_H06-12* are greater than the average values, especially *Parent_and_child_W06-12* (0.2205 *vs.* 0.1853) and *Parent_and_child_H06-12* (0.2576 *vs.* 0.0268). It means that the users in Cluster219 use the category of parent_and_child more frequently in the morning on both holidays and workdays.

Moreover, in the top 5 highest idiosyncratic features, there are 4 features related to the category of Parent_and_child. The applications in this category are about how to raise a baby, how to help pregnant women, *etc.* The users in Cluster219 use the category of parent and child much more frequently than others, especially during afternoon and morning hours on both holidays and workdays. Given the proportion of users that are 25-34 years old, we can label the users in this cluster as **Young parents**. They also use calendar apps and casual and puzzle apps more frequently during afternoon hours on holidays. However, the young parents rarely use clock applications, especially in the morning. They do not often use the categories of education and lifestyle during afternoon hours on workdays.

Cluster80 contains 221 users of which 86% of users are male, and 69% are in the age range from 25 to 34 years old. For the general features, the cluster values of *Stock_H06-12*, *Shopping_H18-24* and *Business_H12-18* are greater than the average values, especially *Stock_H06-12* (0.1032 *vs.* 0.0474).

Three of the highest idiosyncratic features are related to car apps. Thus, we can label the users in Cluster80 as **Car lovers**. They use applications about cars much more often during night hours on holidays (0.8039 *vs.* 0.0231). They also use the car apps more frequently during the 4 time periods of workdays, especially the night hours (0.5549 *vs.* 0.0331). They use navigation apps more often in the mornings on holidays and in the evenings on workdays and shopping apps more often during evening hours on holidays. However, they rarely play other_games in the afternoon on both workdays and holidays, and casual and puzzle games during evening hours on workdays. The users in this cluster are very different from the users in Cluster219 (**Young parents**), despite a similar age distribution. We can see that there is a strong correlation of gender on application usage behaviors.

In addition to gender, we also found that income level and age have a strong impact on the usage behaviors. For example, users with high-income levels use the categories of travel and health_and_fitness more frequently on holidays. Female users with ages between 25-34 years old use the category of parent_and_child more often in the daytime on both holidays and workdays while female users between 0-17 use education-related apps more frequently in the evening on workdays.

DISCUSSION

In this study, our goal was not to identify *all* user clusters that exist in the worldwide smartphone user population, but to show evidence that there actually exists several (and not just a small number of) different and diverse user groups. Although in our results we only discussed 6 clusters, the same analytic approach we took can be used to describe any of the remaining clusters using their most salient properties. We have shown through our 6 clusters, that each cluster has interesting properties that give us a much better picture about the users in that cluster.

Through our results, we have successfully identified that there are several differentiable groups of users that were identified solely from their application usage behaviors. Moreover, we found that demographics play a strong role on application usage behaviors, especially in some of the smaller clusters. As mentioned above, the female- and male-dominated clusters had very different behaviors despite other demographics being similar. We also found that income level and age have a strong impact on the usage behaviors.

The several distinct groups of users we discovered prove that the assumption or simplification that all smartphone users are similar and that they can be treated as a uniform group, is not true. Smartphone manufacturers, mobile carriers, app developers and anyone who impacts the kinds of apps that are placed on phones, what apps are provided on phones and how people select apps to execute, can no longer treat users like they all fit into one big group. This simplifying assumption should not be used in practice. Based on our findings about the selected clusters, we will discuss the implications for researchers, smartphone manufacturers, mobile carriers and app developers.

Implications for researchers

The high variety of application usage behaviors (382) found in our dataset is a clear indication that any given research study with a small user sample, may be inadvertently targeting one or a small set of user clusters. This may cause biased results that are not representative of the population and may only hold for a small set of people. This may explain the inability to replicate results across studies, as mentioned in [4]. To avoid this problem, researchers should either target specific kinds of users or at least acknowledge the kinds of users they ended up targeting in their studies. This will lead to more replicable studies and results. Moreover, we urge researchers, to recruit users outside of the general population of college students, who are more representative of different kinds users.

Implications for smartphone manufacturers

Smartphone manufacturers can build smartphones that are targeted towards improving the user experience of the different kinds of users by providing features that different users may value more than others. For example, **Financial users** may value an improved GPS sensor since they use the navigation category more frequently in the morning on holidays. Particularly with the new LG G5 phone and the proposed Aria phone that support pluggable sensor packages, this type of customization might become readily accessible. For **Screen checkers** who frequently wake up their phones just to check the time or see if there are any notifications, smartphone manufacturers can add an AMOLED or e-ink "always-on" display that lets the locked phone show time or unread notifications on what would otherwise be a black screen. Screen checkers could choose what to show according to their preferences, and would not need to manually turn on their phones and unlock it for much of the day.

Implications for mobile carriers

Mobile carriers that sell smartphones often pre-populate the phones with apps of their choosing. They could allow for the customization of what apps are made available for different kinds of users according to their interest or preferences. For example, mobile carriers can install stock-related apps on the phones for **Financial users**, who have a strong interest in stocks. Similarly, apps related to parent_and_child and cars can be installed on the phones for **Young parents** and **Car lovers**, respectively, since Young parents use the category of parent_and_child more frequently and Car lovers have a strong interest in cars. Of course, rather than pre-installing, the apps would be installed after some use of the phone, to allow for identification of which cluster a user belongs to without leaking users' privacy. Given there are so many apps in one category, the app which is the most popular could be selected for installation.

Implications for app developers

App developers should also be thinking about their target populations in terms of adaptations that their apps could support based on the cluster their users are a part of. For example, **Night communicators** use phone and SMS apps more frequently during night hours on both holidays and workdays. Guaranteeing that the battery level is high enough at night by reminding users to recharge their smartphones may be an invaluable feature for these users. Also, the home screen could be changed to highlight phone and SMS apps during night hours.

For **Evening learners**, notifications from other applications could be suppressed to help users better concentrate on their learning, when using education apps. In addition, they use education apps more frequently in the evening hours, especially on holidays. However, they rarely use education apps in the morning and afternoon on holidays. Based on this habit, the learning apps could be adapted to work with calendar or alarm apps to automatically make a schedule for the Evening learners, to remind them to use the apps.

In addition to adaptation, apps and phones could support better recommendations based on user clusters. For **Financial users**, news about stocks and business can be recommended to them. Casual and puzzle games can also be recommended to them in the mornings on holidays. For **Young parents**, news, forums or applications about raising a baby or pregnancy can be recommended. For **Car lovers**, similarly, some news, forum or applications about cars can be recommended.

Furthermore, our results could be used to curate, and better time, the delivery of better content and advertisements for different groups of users. Young parents may be interested in vaccines, parenting skills, school districts *etc.* Car lovers may not be as interested in those topics, but more interested in car sales, new car technology, *etc.*

Developers can also provide personalized services by identifying which user cluster a user belongs to. When a phone is first turned on, it can collect demographic and initial usage information through short surveys that help to initially place users in a particular cluster. Then the phone can be populated with apps appropriate for the identified cluster. Considering that users change their interests over time, their interests can be observed over different time windows. App usage data collected in these windows can be compared and used to refine the cluster choice, and even reassign the cluster choice. The optimal time window duration can be adjusted according to the observation. This assumes that users will agree to contribute their app usage data, but this may require compensation in some form.

All of these suggested adaptations are intended to improve the user experience for each cluster type, but need to be evaluated to confirm whether these are desired adaptations.

Study limitations

Although we can discover different kinds of users from the recent task lists, we must acknowledge the limitations of the dataset used in this study. First, our dataset is not a representative dataset of all users, as the users were all in China. Second, the dataset consisted of recent task lists that were collected once every hour. This low sampling rate can cause us to miss information about app usage. Third, from the recent task lists, we do not know how long each app is used, how often it is used, and in which order the task list changes. This kind of information could be very helpful to more precisely characterize usage behaviors. This was a known tradeoff of using an existing dataset *vs.* collecting our own, and is one we will address in a future data collection of our own.

Future work

In this work, we provided a method for finding user clusters, showed that there are several interesting user clusters that exist in our dataset, and discussed the implications of our findings. However, it would be valuable to see how well the clusters we identified generalize to other user populations. We plan to explore this through our own future data collections. In addition, we described a series of implications of the existence of different clusters, and proposed a set of changes to both research and professional practice. In our future work, we will implement these changes and determine what impact they have on the user experience. We could do this by recruiting and screening for a particular user group, and apply our proposed changes (adaptations, recommendations, *etc.*) to their phones, and evaluate how useful users find them to be.

CONCLUSION

In this work, we have successfully shown the existence of a much more diverse set of users than previously shown. We discovered 382 distinct kinds of users using our own clustering and feature ranking method. We selected general features and idiosyncratic features to identify the app usage behaviors of the users in each cluster. Then, we gave a meaningful label to the users in each cluster, such as Night communicators, Evening learners, Screen checkers, Financial users, Young parents and Car lovers. We showed that demographics have an important impact on how users use applications on their smartphones. Finally, we show how our findings, which break the simplifying assumption that all smartphone users are similar, can be used by researchers, smartphone manufacturers, mobile carriers and app developers, to tailor better research studies, smartphones and apps for different kinds of users.

ACKNOWLEDGEMENTS

This research was supported by the Zhejiang Provincial Natural Science Foundation of China (LR15F020001) and Program for New Century Excellent Talents in University (NCET-13-0521), funded by the U.S. National Science Foundation under Grant IIS-1407630 and supported by National Key Research and Development Plan under Grant No. 2016YFB1001200. This work was performed when Sha Zhao was visiting the Human-Computer Interaction Institute, Carnegie Mellon University, USA.

REFERENCES

1. Nikola Banovic, Christina Brant, Jennifer Mankoff, and Anind Dey. 2014. ProactiveTasks: the short of mobile device use sessions. In *Proceedings of the 2014 International Conference on Human-computer Interaction with Mobile Devices and Services*. ACM, 243–252.
2. David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of Machine Learning Research* 3 (2003), 993–1022.
3. Matthias Böhmer, Brent Hecht, Johannes Schöning, Antonio Krüger, and Gernot Bauer. 2011. Falling asleep with Angry Birds, Facebook and Kindle: a large scale study on mobile application usage. In *Proceedings of the 2011 International Conference on Human Computer Interaction with Mobile Devices and Services*. ACM, 47–56.
4. Karen Church, Denzil Ferreira, Nikola Banovic, and Kent Lyons. 2015. Understanding the Challenges of Mobile Phone Usage Data. In *Proceedings of the 2015 International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 504–514.
5. Trinh Minh Tri Do, Jan Blom, and Daniel Gatica-Perez. 2011. Smartphone usage in the wild: a large-scale analysis of applications and context. In *Proceedings of the 2011 International Conference on Multimodal Interfaces*. ACM, 353–360.
6. Hossein Falaki, Ratul Mahajan, Srikanth Kandula, Dimitrios Lymberopoulos, Ramesh Govindan, and Deborah Estrin. 2010. Diversity in smartphone usage. In *Proceedings of the 2010 International Conference on Mobile systems, Applications, and Services*. ACM, 179–194.
7. Denzil Ferreira, Jorge Goncalves, Vassilis Kostakos, Louise Barkhuus, and Anind K Dey. 2014. Contextual experience sampling of mobile application micro-usage. In *Proceedings of the 2014 International Conference on Human-computer Interaction with Mobile Devices and Services*. ACM, 91–100.
8. Daniel Hintze, Rainhard D Findling, Muhammad Muaaz, Sebastian Scholz, and René Mayrhofer. 2014. Diversity in locked and unlocked mobile device usage. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 379–384.
9. Ke Huang, Chunhui Zhang, Xiaoxiao Ma, and Guanling Chen. 2012. Predicting mobile application usage using contextual information. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 1059–1065.
10. Chakajkla Jesdabodi and Walid Maalej. 2015. Understanding usage states on mobile devices. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 1221–1225.
11. Simon L Jones, Denzil Ferreira, Simo Hosio, Jorge Goncalves, and Vassilis Kostakos. 2015. Revisitation analysis of smartphone app use. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 1197–1208.
12. Philip Leroux, Klaas Roobroeck, Bart Dhoedt, Piet Demeester, and Filip De Turck. 2013. Mobile application usage prediction through context-based learning. *Journal of Ambient Intelligence and Smart Environments* 5, 2 (2013), 213–235.
13. Huoran Li, Xuan Lu, Xuanzhe Liu, Tao Xie, Kaigui Bian, Felix Xiaozhu Lin, Qiaozhu Mei, and Feng Feng. 2015. Characterizing Smartphone Usage Patterns from Millions of Android Users. In *Proceedings of the 2015 ACM Conference on Internet Measurement Conference*. ACM, 459–472.
14. Zhung-Xun Liao, Yi-Chin Pan, Wen-Chih Peng, and Po-Ruey Lei. 2013. On mining mobile apps usage behavior for predicting apps usage in smartphones. In *Proceedings of the 2013 ACM International Conference on Information and Knowledge Management*. ACM, 609–618.
15. R Lletı, M Cruz Ortiz, Luis A Sarabia, and M Sagrario Sánchez. 2004. Selecting variables for k-means cluster analysis by using a genetic algorithm that optimises the silhouettes. *Analytica Chimica Acta* 515, 1 (2004), 87–100.
16. Anmol Madan, Manuel Cebrian, David Lazer, and Alex Pentland. 2010. Social sensing for epidemiological behavior change. In *Proceedings of the 2010 ACM International Conference on Ubiquitous Computing*. ACM, 291–300.
17. Anmol Madan, Manuel Cebrian, Sai Moturu, Katayoun Farrahi, and others. 2012. Sensing the" health state" of a community. *IEEE Pervasive Computing* 4 (2012), 36–45.
18. Ujjwal Maulik and Sanghamitra Bandyopadhyay. 2002. Performance evaluation of some clustering algorithms and validity indices. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 12 (2002), 1650–1654.
19. Norbayah Mohd Suki and Norazah Mohd Suki. 2007. Mobile phone usage for m-learning: comparing heavy and light mobile phone users. *Campus-Wide Information Systems* 24, 5 (2007), 355–365.

20. Steven M Pincus. 1991. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences* 88, 6 (1991), 2297–2301.

21. Ahmad Rahmati and Lin Zhong. 2010. A longitudinal study of non-voice mobile phone usage by teens from an underserved urban community. *arXiv preprint arXiv:1012.2832* (2010).

22. Choonsung Shin and Anind K Dey. 2013. Automatically detecting problematic use of smartphones. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 335–344.

23. Choonsung Shin, Jin-Hyuk Hong, and Anind K Dey. 2012. Understanding and prediction of mobile application usage for smart phones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 173–182.

24. Vijay Srinivasan, Saeed Moghaddam, Abhishek Mukherji, Kiran K Rachuri, Chenren Xu, and Emmanuel Munguia Tapia. 2014. Mobileminer: Mining your frequent patterns on your phone. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 389–400.

25. statista. 2016. Number of available applications in the Google Play Store from December 2009 to November 2015. statista. (2016). Retrieved March 26, 2016 from `http://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/`.

26. Robert Tibshirani, Guenther Walther, and Trevor Hastie. 2001. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63, 2 (2001), 411–423.

27. Niels van Berkel, Chu Luo, Theodoros Anagnostopoulos, Denzil Ferreira, Jorge Goncalves, Simo Hosio, and Vassilis Kostakos. 2016. A Systematic Assessment of Smartphone Usage Gaps. In *Proceedings of the 2016 SIGCHI Conference on Human factors in Computing Systems*. ACM, 4711–4721.

28. Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9, 2579-2605 (2008), 85.

29. Qiang Xu, Jeffrey Erman, Alexandre Gerber, Zhuoqing Mao, Jeffrey Pang, and Shobha Venkataraman. 2011. Identifying diverse usage behaviors of smartphone apps. In *Proceedings of the 2011 ACM Conference on Internet Measurement Conference*. ACM, 329–344.

30. Tingxin Yan, David Chu, Deepak Ganesan, Aman Kansal, and Jie Liu. 2012. Fast app launching for mobile devices using predictive user context. In *Proceedings of the 2012 international conference on Mobile systems, Applications, and Services*. ACM, 113–126.

31. Chunhui Zhang, Xiang Ding, Guanling Chen, Ke Huang, Xiaoxiao Ma, and Bo Yan. 2012. Nihao: A predictive smartphone application launcher. In *Mobile Computing, Applications, and Services*. Springer, 294–313.

32. Sha Zhao, Gang Pan, Yifan Zhao, Jianrong Tao, Jinlai Chen, Shijian Li, and Zhaohui Wu. 2016. Mining user attributes using large-scale app lists of smartphones. *IEEE SYSTEMS JOURNAL* 99 (2016), 1–9.