Demographics of mobile app usage: long-term analysis of mobile app usage

Tu, Zhen

2021

Tu , Z , Cao , H , Lagerspetz , E , Fan , Y , Flores , H , Tarkoma , S , Nurmi , P & Li , Y 2021 , ' Demographics of mobile app usage: long-term analysis of mobile app usage ' , CCF þÿ T r a n s a c t i o n s o n P e r v a s i v e C o m p u t i n g a n d I n t e r a c t i o n , v o l . 3 , p p . 2 3 5

http://hdl.handle.net/10138/333795
https://doi.org/10.1007/s42486-020-00041-3

acceptedVersion

CCF TPCI manuscript No.
(will be inserted by the editor)

Demographics of Mobile App Usage: Long-Term Analysis of Mobile App Usage

Zhen Tu · Hancheng Cao · Eemil Lagerspetz · Yali Fan · Huber Flores · Sasu Tarkoma · Petteri Nurmi · Yong Li

Received: date / Accepted: date

Abstract In the past decade, mobile app usage has played an important role in our daily life. Existing studies have shown that app usage is intrinsically linked with, among others, demographics, social and economic factors. However, due to data limitations, most of these studies have a short time span and treat users in a static manner. To date, no study has shown whether changes in socioeconomic status or other demographics are reflected in long-term app usage behavior. In this paper, we contribute by presenting the first ever long-term study of individual mobile app usage dynamics and how app usage behavior of individuals is influenced by changes in socioeconomic demographic factors over time. Through a novel app dataset we collected, from which we extracted records of 1,608 long-term users with more than 3-year app usage and their detailed socioeconomic attributes, we verify the stable correlation between user app usage and user socioeconomic attributes over time and identify a number of representative app usage patterns in connection with specific user attributes. On the basis, we analyze the long-term app usage dynamics and reveal that there is significant evolution in long-term app usage that 60%-70% of users change their app usage patterns during the duration of more than 3 years. We further discover a variety of app pattern change modes and demonstrate that the long-term app usage behavior change reflects

Z. Tu, Y. Li
State Key Laboratory on Microwave and Digital Communications, Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China.
E-mail: liyong07@tsinghua.edu.cn

E. Lagerspetz, H. Flores, S. Tarkoma, P. Nurmi
Department of Computer Science, University of Helsinki, Helsinki, Finland.

H. Cao
Department of Computer Science, Stanford University, California, United States.

Y. Fan
Department of Sun Yat-Sen University, Guangzhou, China.

corresponding transition in socioeconomic attributes, such as change of civil status, family size, transition in job or economic status.

Keywords App usage · Long-term analysis · Economic attributes · user study

1 Introduction

In the past decade, mobile app usage has played an important role in our daily life. Recent estimates suggest that people spend over 3 hours each day on mobile phones [41], with approximately 90% of this time spent with mobile apps [39]. This ubiquity of mobile app usage provides an opportunity to study and understand human behavior at an unprecedented scale through information garnered from mobile devices [3, 40, 45, 51]. Besides offering information on human behavior, understanding mobile app usage is also essential for mobile developers, service providers and other stakeholders [2, 56, 72]. Having a better understanding of mobile app usage is also paramount to delivering better personalized services, advertisements, and recommendations [7, 37, 44, 54], and serves as essential input for mobile analytics [56, 72].

The growing importance of mobile app usage has been driving research into understanding factors that govern it. Several studies have investigated the relationship between mobile app usage and demographic attributes, demonstrating that factors ranging from cultural dimensions to socioeconomic attributes affect app usage [13, 18, 45, 48]. While these studies offer insights into the relationships between demographics and app usage, a significant limitation is that they have only considered short time spans. As a matter of fact, most studies consider data collected from periods shorter than a year, thus unable to investigate how socioeconomic demographics influence a user's app usage behavior in the long term. Specifically, when a user changes some of his/her demographics over time, will his/her app usage behavior change correspondingly? For the long-term evolution of individual's app usage behavior and corresponding socioeconomic demographics, so far rare study has explored this important issue. The only exceptions have been studies aiming at understanding evolution of individual apps [4, 38, 53], but no study has examined the long-term characteristics of mobile app usage as a whole.

To shed lights on this question, in this paper we contribute by conducting the first ever *long-term study* on mobile app usage based on more than 3 years of app usage data from 1,608 users from more than 40 countries all around the world. In terms of data scale, it's several orders of magnitude larger than that in previous works[28, 36]. In addition, our dataset is complemented with an online survey of user demographics that 3,303 app users have reported their detailed demographics, i.e., age, gender, household, education, occupation, salary, savings and debt. Based on the dataset, we first analyze the stability of user behavior over time, demonstrating that most users (approximately 70%) indeed change their app usage in the long-term, but the relationship between demographic attributes and app usage remains stable over time. To understand dynamics in mobile app usage, we develop a novel methodology for identifying *change modes* in mobile app usage. By correlating extracted change modes with varying user attribute status, we demonstrate that these app usage changes can be (largely) explained by changes in demographic attributes, such as changes in civil status, job, family size, or income. Further, we use our model to explain changes occurring during the 3+ year period our data spans, offering new insights into how mobile app usage evolves in the long-term.

We summarize our work's contribution in the following three aspects:

- We present the first ever long-term study (more than 3 years) of mobile app usage behavior, and reveal how usage patterns are influenced by changes in demographics, and particularly changes in socioeconomic status.
- We develop a novel framework to capture and model correlation between app usage patterns and specific user groups with over 80% prediction accuracy, which enables us to further identify representative change modes of app patterns in connection with changes of socioeconomic status.
- Analysis of our long-term app dataset reveals some interesting phenomena. First, 60-70% of users do change their category-level app usage patterns in the long run, though a user usually keeps a stable pattern for a period of time with the socioeconomic status. Second, long-term changes predominantly reflect corresponding changes in socioeconomic status, such as changes in civil status, family size, transitions in job or economic status.

The structure of this paper is as follows. Section 2 describes the background of our research. Section 3 introduces the dataset and introduces methodology. Section 4 demonstrates long-term app usage behavior and its relationship with user social economic demographics. Section 5 reveals the dynamics of long-term app usage and the drivers behind. Finally, Section 6 discusses the implication and application of our work and Section 7 provides the conclusion of the paper.

2 Background

In this section, we introduce the background of our study and define our research question and hypotheses.

2.1 App Usage Modeling

To understand smartphone app usage patterns, existing work focus on different aspects of them from different angles. For example, Falaki et al. [18] studied the number of the daily interactions of individuals with various apps, Li et al. [33] investigated the different download and usage modes of different apps. Besides, Ferreira et al. [20] and Do et al. [13] investigated how app usage varies with different contexts. Tu et al. [61] measured the uniqueness of mobile app usage behavior considering spatial and temporal context information. Other works

[20, 24, 32, 35, 55] focused more on the prediction of app usage by studying how to predict which apps users are likely to install or visit. Xu et al. [64] identified traffic from distinct marketplace apps and presented statistic results on their spatio-temporal prevalence and correlation. In addition, existing works have also payed attention to identify various app usage patterns: [6, 26] studied how often individuals revisit a specific app and Rachuri et al. [48] discovered the pairwise apps that were frequently used together. Furthermore, existing works also tried to understand the different characteristics of different user groups on app usage. Zhao et al. [70] grouped smartphone users by their different app usage behaviors, and Blaszkiewicz et al. [5] differentiated smartphone users by their installed Top-60 apps.

However, all these existing works leverage short-term data, which makes it hard to study long-term app usage patterns of mobile users. In this study, taking advantage of a novel dataset that tracks user app usage for more than 3 years, we provide the first glimpse into the long-term dynamics of user app usage behavior. Specifically, we aim to answer the following research question: What patterns do individual's app usage behavior demonstrate in the long term? And what are the drivers behind them?

2.2 Long-term Behavioral Pattern Analytics

There has been a rich literature on analyzing long-term human behavioral patterns. Contrary to short-term phenomena, human behavior in the long run are found to be dynamically evolving [10] rather than stable, supported by theories and empirical evidence in various settings. For instance, in online communities, users are found to demonstrate linguistic variation [12, 49], changes of roles [47, 65], emergence of convention [30], churn behavior [15], and forming activity lifespans [16, 22, 66]. Longitudinal studies in medical science [21, 60] and social science [9, 43] have also demonstrated the evolving nature of long-term human behavior.

However, there are only a few works analyzing long-term behaviors of smartphone users to the best of our knowledge. Lin et al. [38] studied how users engaged on a health app for about 31 months and demonstrated user's "multiple lives", where most of users turn back to use this app after a long period of inactivity. Shameli et al. [53] analyzed one-year user records on an activity tracking app, and found the app's gamification affects user physical activities. Althoff et al. [4] discovered that the social networking behavior on the same app promotes the increase of user online and offline activities. A recent work [34] studied evolution of user app usage diversity over time. However, these line of works are generally limited to analyzing user behavior on a single app. Different from them, we aim at providing a comprehensive view on how individuals use different kinds of apps over time.

Based on existing findings of user long-term behavior in these various settings and observed user behavior on single app, we make the following hypothesis:

[H1] User app usage behavior evolves in the long run.

2.3 Social Demographic Determinants of Human Behavior

Social science theories have identified social demographic factors as important determinants of human behavior [58, 68], and empirical studies in diverse fields have confirmed the close link between social demographic attributes and user behavior [1, 14, 27, 31]. Based on user behavior data from different sources (e.g., communication, credit card transaction, traffic, online social networks, wearable sensors), numerous attempts have been made to predict user gender [71], ethnicity [46], age [19], profession [63], living pattern [8], income [59], employment status [3], financial well-being [57], social relationship [17], etc. Studies have also shown that user behavior change is linked with change in his/her social demographic status [25, 42, 43].

Previous works have also demonstrated the strong correlations between user app use behavior and user demographic attributes. Malmi et al. [40] found that app usage behaviors between user groups with distinguished demograpghics had some differences, and further proposed a method to predict users' race, age and income, extending earlier works by Seneviratne et al. [51, 52]. The impact of country and demographic factors' on app category usage has also been studied recently [45], where the authors found that geographic boundaries and users' language and profession, affect app category usage patterns significantly more than commonly used attributes, such as gender and age. However, while these studies have explored the link between mobile app usage and demographics, no study has examined the stability of this relationship in long-term app usage. We hypothesize that this link remains stable over time, unless the user transitions from social group to another (i.e., socioeconomic demographics change). Specifically, we hypothesize:

[H2] App usage patterns reflect, and are influenced by socioeconomic status, and remain stable unless socioeconomic status also changes.

[H3] Significant changes in long-term app usage behavior occur in correspondence with changes in user social and economic demographics.

To test our hypotheses, we collected a long-term app usage dataset of more than 3 years and carried out an online survey about user socioeconomic attributes. To test hypothesis H1, we study whether individuals change their usage pattern in the long term. While for hypothesis H2 and H3, we propose a framework to capture and model correlation between app usage patterns and specific user groups, which enables us to identify representative change modes in connection with socioeconomic status.

3 Dataset and Method

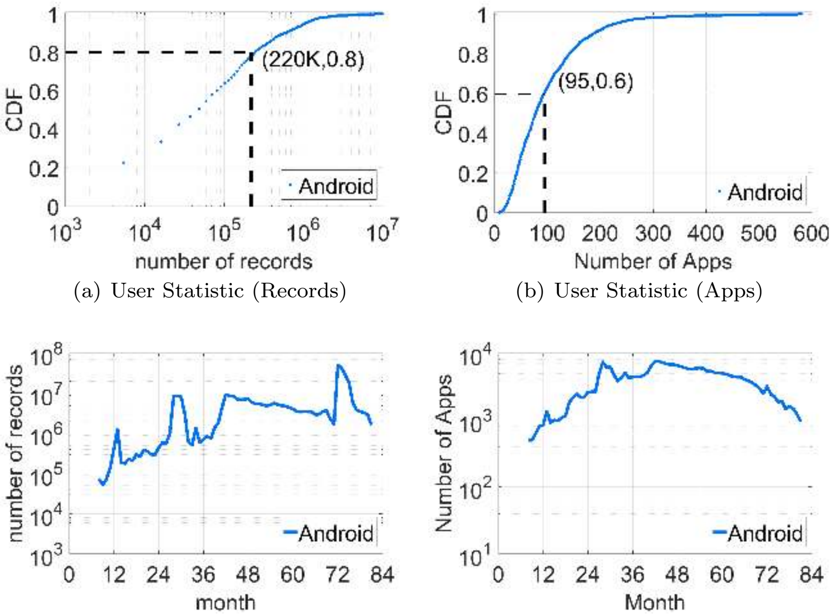
3.1 Data Collection

To study the long-term individual app usage behavior, we have instrumented a popular mobile energy-awareness app to continuously collect app usage data for almost seven years from 1st January 2012 to 25th August 2018. The app collects measurements every time the phone battery increases or decreases by 1% of the whole power. Each record contains the battery level, complete list of running apps, a randomly generated user identifier, timestamp, and time zone. Up to now, we have gathered users from around 200 countries or geographic areas and more than 100,000 apps in 36 categories.

Users are identified by a unique, randomly generated identifier assigned to each installation of the app at install time. Therefore, users who uninstall and reinstall the app are considered new users. Long-term users that we study were therefore users of the same device with the same installation of the app for a prolonged time. In addition, answers from *survey users* described below can be used to group those users who changed devices or reinstalled the app, but answered the survey on both devices. Since different users begin to use this app at different times and keep it for different time periods (i.e., ranging from a couple months to years), to facilitate our long-term study, we select user records with **more than 3-year record history** (because users may join in and leave in), and obtain a large-scale usage data with 1,608 users globally (we call them ***long-term users***) with app usage history of ranging from 3 to 7 years.

We first show the basic statistical information of users and periods to illustrate data quality. In Fig. 1 (a) and (b), we plot the Cumulative Distribution Function (CDF) of the number of records and apps per user. From the figure, we can observe that top 20% users have more than 220,000 records and top 40% users have more than 95 distinct apps. In addition, the number of records and apps per month are shown in Fig. 1 (c) and (d). From the results, we can observe that there are a million records and 1,000 apps on average, and the two curves are relatively stable. This demonstrates that our data is continuous, and record coverage is high over the entire duration of the long-term dataset.

Apart from user app usage records, we also conducted an online survey to collect the socioeconomic attributes of the users. The survey started in July 2016 and ended in February 2017 (i.e., 8-month duration). In total, 3303 users (we call them *survey users*) participated in our online survey and provided eight attributes: *gender, age, household, education, occupation, salary, savings, debt*. In addition, the app usage data of these users has been also collected, though the durations of these app usage records are not as long as long-term users'. According to our statistical result, among the durations of the *survey users'* app records, the largest is 6 years, and 50% of them are more than 10 months. Fortunately, there are 324 users (*i.e.*, 20% of *long-term users*) existing in both *long-term users* and *survey users* simultaneously, which helps us to



(a) User Statistic (Records) (b) User Statistic (Apps)

(c) Period Statistic (Records) (d) Period Statistic (Apps)

Fig. 1 Basic statistical information of users and periods in our long-term app usage dataset.

better study the relationship between users' app usage behaviors and their socioeconomic attributes.

In summary, our long-term app usage dataset is very dense and of high quality, and the online survey collects user fine-grained socioeconomic attributes, which provides us a unique opportunity to explore and understand long-term app usage dynamics with social and economic factors.

Ethical Considerations. We are aware of the privacy implications of using individual app usage data for this work. When dealing with such sensitive dataset, we have taken careful actions due to ethical considerations. User data was identified by a randomly generated identifier when the app was first installed. We have informed the mobile users of the data collection procedures and received their consent before this study. In addition, the online survey was conducted through the app, without revealing personally identifying information. Participants are voluntary to submit the online survey after being informed about the data collection purpose for only academic research. As for data storage, all the data are stored in a local and secure server behind a firewall, which is protected by strict authentication mechanisms. In other words, only researchers after signing a strict non-disclosure agreement can have the right to process this data. At last, this study has been approved by our local institution.

3.2 Identifying the Long-term App Usage Changes

In this paper, we aim to answer the research question of what patterns app usage behavior demonstrates in the long term and what's the drivers behind. In order to answer it, we need to know whether individuals change their app usage pattern in the long term, and how the changes look like. Therefore, one natural solution is to learn the usage pattern in each period (i.e., short term) and then compare the same user's patterns at different period (i.e., long term) to detect whether changes happen. To achieve this goal, we propose a framework to first model typical app usage patterns based on user attributes, and then predict individual app usage patterns and corresponding attributes in all periods and detect their changes in the long term, and finally identify representative long-term change modes and explore the underlying reasons in social and economic aspects.

Our proposed framework is shown in Fig. 2, which contains four modules: *data pre-processing, user division, pattern prediction, change detection.* After receiving the inputs of mobile app usage data from survey users and long-term users, our framework passes through all the modules and finally outputs the discovered individual's change modes, including both corresponding app usage and demographic changes. The details of each module are as follows:

Data Pre-processing. This module transforms the original app usage records into weekly usage vectors, which can be easily utilized by other modules. Formally, assume there are N app users and the app usage records of user u_i in the time order are denoted as $\{r_{ij}\}$, where r_{ij} represents the j-th records of user u_i and each record contains timestamp, app and category information. Since our dataset covers about almost 7 years from 1st January 2012 to 25th August 2018, we divide the whole time duration into 347 weeks and collect weekly usage vectors for all the users. Specifically, since each user only uses a small part of the whole app sets (i.e., more than 100k), we adopt the app category information to form the app usage vectors to avoid them to be too sparse. Again, for user u_i, his app usage vector for the k-th week is denoted as $V_{ik} = [v_{ik}^1, ..., v_{ik}^{36}]$, which v_{ik}^m represents the usage frequency of apps from the m-th category in that week. In addition, $\widetilde{V}_{ik}$ is the normalized usage vector.

In this way, we process the app usage records from long-term users and survey users into weekly usage vectors, respectively. For 1,608 long-term users, we obtain the app usage tensor $V^L = \{V_{ik}^L, k = 1, ..., 347\}$, which serves as the input of the third module *pattern prediction.* Since long-term users have more than 3-year history, users may have some empty weeks without any record (e.g, before joining the app or after leaving the app). Thus we skip these weeks in the following analysis. While for survey users, they answered the online survey about their attributes during the period from July 2016 to February 2017. In order to ensure an accurate matching of specific demographic users and their app usage behavior at that time, we check the answering time of each user and only utilize the nearest half-year app usage records, then we obtain the weekly app usage vectors for each user to form the app usage tensor $V^S = \{V_{ik}^S, k = 1, ..., 24\}$ as the input for next module *user division.* In addition, considering

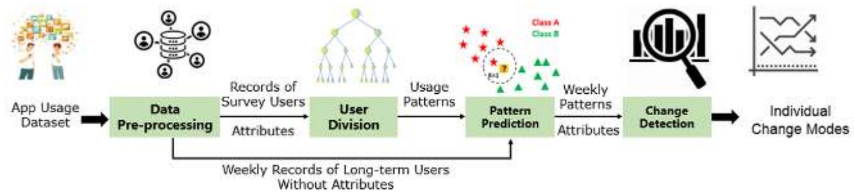


Fig. 2 Our framework for identifying individual long-term app change modes.

social and economic attributes (i.e., household, education, occupation, salary, savings and debt), we also form a user attribute matrix $D = \{d_{ij}\}$, where d_{ij} represents the value for the j-th attribute for user u_i.

User Division. Based on app usage behavior and attributes of survey users, this module aims to measure how social and economic attributes influence the behavior, and further extract typical app usage patterns and specific user groups by appropriate user division.

First, we measure whether there are significant differences between user groups with different attribute status. According to each attribute, we divide users into different groups and compare their difference in app usage behavior by conducting pairwise significance testing experiments [23]. For example, for the j-th attribute with A_j types of status, we calculate the average app usage vectors weekly for group 1 and group 2, denoted as $\{V_{i,1}^{ave}, ..., V_{i,24}^{ave} | \forall i, A_{ij} = m_1\}, \{V_{i,1}^{ave}, ..., V_{i,24}^{ave} | \forall i, A_{ij} = m_2\}, m_1 < m_2, 0 \leqslant m_1 < A_j, 0 < m_2 \leqslant A_j$, where V_{ik}^{ave} means the average app usage vector of that group for the k-th week, then perform a *paired t-test* between them and by the output p-value. By setting the significant level at 0.05, we measure whether pairwise groups are significantly different.

Second, we extract typical app usage patterns from survey users and obtain the corresponding user groups by appropriate user division. To make full use of user attribute information, we design a decision tree [50] based user division method to group users by maximizing inter-group differences as much as possible. Originally, the decision tree is used for classification task by differentiating samples by choosing the most effective feature with the largest information gain in every division step [29]. In our case, we want to differentiate users by app usage behavior (a vector) rather than sample labels (a value) and the features are user attributes. After each division, users in the same subgroup become more similar and different subgroups become more distinct. Step by step, we can get more and more distinctive subgroups and extract typical app usage patterns by the end of iteration. Compared with classical decision tree model, here we adopt a new "entropy" metric for vectors and compute the information gain in a different way. Specifically, we utilize *KL-distance* [62], i.e., a common distance function comparing the difference between a target probability distribution Q and a true probability distribution P, to compute the entropy of vectors. The average distance between a number of vectors is measured by the mean *KL-distance* between each vector and the central vector. Accordingly, the information gain is calculated as follows:

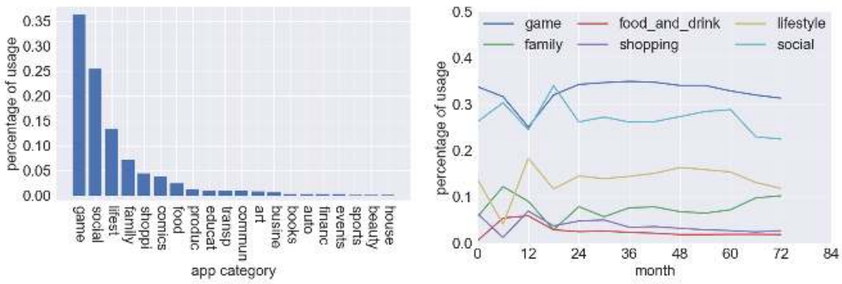
$$\begin{aligned} KL(P,Q) &= \sum_i P_i log \frac{P_i}{Q_i}, \\ g(V,C) &= \frac{1}{|C|}\sum_{i\in C} KL(\widetilde{V_i}, \widetilde{V_i}^{ave}) - \sum_j \frac{|C_j|}{|C|}\sum_{i\in C_j} KL(\widetilde{V_i}, \widetilde{V_i}^{ave}), \end{aligned} \tag{1}$$

where $\widetilde{V_i}$ is the normalized average app usage vector for user u_i, $\widetilde{V_i}^{ave}$ means the normalized average app usage vector of that group, C_j contains the user set of the j-th group divided by the corresponding attribute, $|*|$ represents the size of set $*$. By adopting this modified metric, we compute the information gain after introducing each new attribute to divide users before a new division. Then decide to divide users by choosing the attribute with the maximum information gain, unless the maximum information gain is negative or there are too many small subgroups (i.e., more than 50% of subgroups with a size smaller than the threshold 10). When the division finally stops, we obtain the central app usage vector for each group as the typical usage pattern after filtering out abnormal users. Further, we guarantee the accurate mapping relationship of app usage patterns and corresponding attributes by performing an external prediction task to evaluate its performance.

Pattern Prediction. Based on extracted app usage patterns and user groups, this module aims to label weekly app records of long-term users (without attributes) with the most similar usage pattern and the most likely attribute. Specifically, for user u_i in the k-th week, we utilize the normalized app usage vector $\widetilde{V_{ik}}$ to compute the *KL-distance* with different usage patterns, and assign it the nearest pattern and attribute status accordingly. In this way, for each user, we continuously predict his/her app usage and demographic status in different weeks, i.e., the long-term individual app usage patterns, which are the inputs of next module *change detection*.

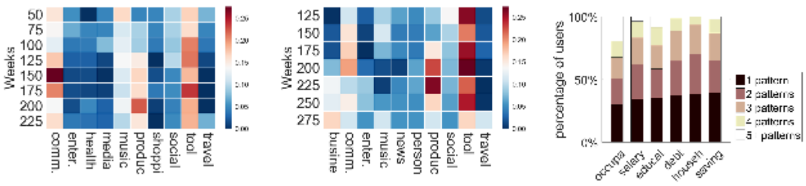
Change Detection. This module tries to detect long-term app usage change modes and examines the correlation with changes of social and economic attributes. Different from short-term app usage changes, long-term app usage changes mean very stable changes of app usage patterns, i.e., keeping one usage pattern for a long time and then changing to another and keeping it again. Therefore, the most fundamental and meaningful template of change pattern is $AABB$, showing that the user both keeps the previous and present app usage pattern for more than one period (i.e., several months). Adopting this template, we utilize the fundamental four-period sliding window to detect whether there exist any pattern changes for long-term app users. Furthermore, we focus on a small number of change modes with the highest occurrence frequency. For those representative app change modes, we try to understand what kinds of change do they have and why they change in connection with the social and economic attributes.

In conclusion, to explore and understand long-term app usage changes, we collect a long-term app usage data and fine-grained online survey about user



(a) Overall App Usage (b) Long-term Usage Changes

Fig. 3 (a) The overall usage frequency of top 20 app categories in the long run. (2) The long-term changes of usage frequency for 6 popular app categories.



(a) Case Study of User 1 (b) Case Study of User 2 (c) Pattern Numbers/User

Fig. 4 (a)-(b) Case study: two typical users' app category usage frequency over the long term. (c) Distribution of users with different numbers of app usage patterns in the long term, when considering different attributes.

attributes, and propose a framework with four modules to identify long-term app usage behavior changes and correlate them with changes of social and economic status. Based on the data and proposed method, next we present our findings on long-term individual app usage behavior.

4 Long-term App Usage and Demographics

Studies on online behaviour suggest that human behavior is dynamically evolving in the long run [12, 15, 30, 47, 65]. However, the extent at which this holds for mobile app usage is less known due to the short time span of earlier studies [37, 54, 69]. In this section, we demonstrate that mobile app usage indeed evolves in the long-term, and that the relationship between mobile app usage and social-economic demographics remains stable across time.

4.1 User App Usage Behavior Evolves in the Long Term.

First of all, we test our hypothesis H1, *user app usage behavior evolves in the long term*, at three levels of aggregated behavior, individual behavior, and usage patterns.

First, based on our long-term app dataset, we analyze the overall usage frequencies and monthly usage frequencies of all the app categories. In Fig. 3 (a), we plot the overall usage frequencies of top 20 app categories. From the figure, we can observe that top 5 categories are *game, social, lifestyle, family* and *shopping*, showing that mobile apps are used for many user needs, such as communicating with others, online shopping, having fun, and enjoying family life. Note that the power-consuming app category *game* has the highest usage percentage, probably due to our data collection method (*i.e.*, taking a sample every time the battery level of the device changes by 1%). However, such method has little biased effect to the observation of the other app categories which consume little power when being used. In addition, we plot the monthly usage frequencies of 6 popular app categories in Fig. 3 (b). From the figure, we can observe that the usage frequency of *social* apps decrease while the usage frequency of *lifestyle* apps increase over time, showing diverse trends of app usage behavior. More importantly, the usage frequencies of all the categories change often, suggesting users change their category-level usage behavior in the long-term.

Second, by conducting case studies of two randomly selected users, we show the long-term app usage also changes on the individual level. In Fig. 4 (a) and (b), we plot the app usage frequencies of ten most notable categories in different periods (i.e., neighboring periods are about half year apart) of two users. For the first user, we can observe that there are apparent usage changes in communication, music, productivity and tools apps. For communication apps, the user first increases then decreases its usage frequency, while productivity apps shows an obvious rise trend of usage frequency. Moreover, the other user also shows various category-level app usage behavior in different period, such as business, communication and entertainment. Interestingly, we can observe that the trends of business and communication apps are opposite, showing that if the user uses the business apps more frequently, he/she also will decrease the usage of communication apps in the long term. Such changes could be the result of getting a new job and becoming busier, thus decreasing the communication with others. As the two case studies show, user app usage behavior evolves rather than remain stable in the long run.

Third, we demonstrate long-term app usage patterns are changing by counting the number of distinct usage patterns each user demonstrates in the long term, obtained from our framework. Specifically, by considering different attributes, we obtain typical app usage patterns from survey users. We predict long-term users' app usage patterns in different periods (note that we will show the prediction's reliability by evaluation in subsection 4.2), then count the number of distinctive usage patterns per user and plot the statistic information in Fig. 4 (c). As shown in the result, considering occupation related usage patterns, we can observe that only 29.75% of users have one usage pattern, which means that about 70% of users have changed their usage patterns in the long term. In addition, we can find that only a small amount (19.38%) of users have more than 5 different patterns, suggesting that changes over the long term stay stable for a significant amount of time. As for savings correlated patterns, we see that only 39.68% of users have one pattern and 60% of users evolve their app usage pattern in the long run. Other attributes (salary, education, debt and household) also reveal similar patterns. All the results clearly demonstrate that 60-70% of users do change their app usage pattern in the long term (e.g., considering a long period of several years). On the other hand, it's worth noting that only a small number of users change their patterns more than four times, indicating that app usage pattern remains largely stable between changes, and these changes are significant events in the dataset.

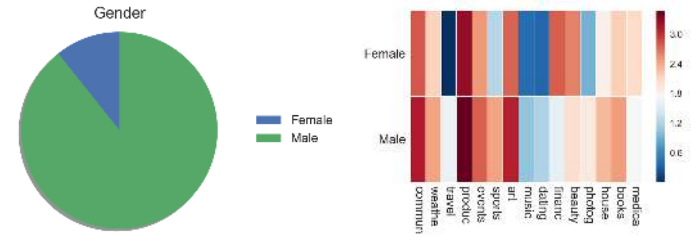
In short summary, through empirical measurement and case study, we have clearly demonstrated that individuals are changing their app usage patterns in the long term: different app categories demonstrate monthly usage frequency variation, and typical users drastically change their category-level usage behavior in different periods. More importantly, 60-70% of users are found to change their app usage patterns during the time span of the dataset. Clearly, consistent results have verified our hypothesis H1, *user app usage behavior evolves in the long term*. Nevertheless,s only a small number of users change their patterns more than four times, indicating that app usage behavior remains stable between changes, and that these changes are significant events in the dataset.

4.2 Relationship between App Usage and Socioeconomic Demographics

Based on app usage data and fine-grained demographic labels from survey users, we further study the relationship between app usage and different social economic attributes so as to test hypothesis H2: *app usage patterns reflect, and are influenced by socioeconomic status, and remain stable unless socioeconomic status also changes.*

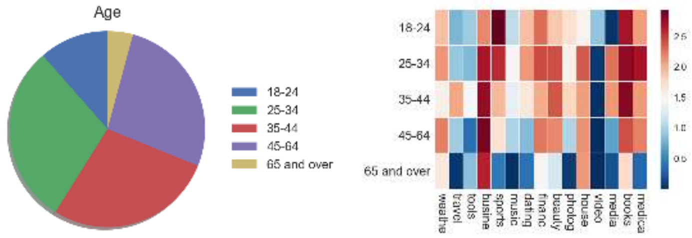
First, we compare the social economic attribute of users across different app usage patterns. In Fig. 5 - Fig. 12, we plot the distribution of user attribute and app usage behavior across different user groups for each of the eight attributes collected from online survey.

- **Gender**: As shown in Fig. 5 (a), the percentages of male and female users are 89.26% and 10.74%. In Fig. 5 (b), for male group and female group (the same user size after random sampling to avoid bias from unbalanced distribution), we plot the average logarithmic usage frequency of different app categories in a week. Different colors represent different usage frequency, a blue color means a low frequency and a red color means a high frequency. From the results, we can observe that female and male users have very different app usage preferences. For example, female users are more likely to use beauty and finance apps, while male users have higher usage frequencies in sports and dating apps.
- **Age**: Fig. 6 (a) shows the distribution of users with different ages. In our dataset, there are five age groups: *18-24, 25-34, 35-44, 45-64, 65 and over*, and the largest group is *25-34* with 29.52% of users. Also, Fig. 6 (b)



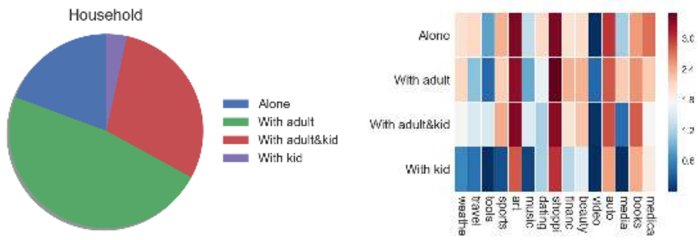
(a) User Distribution (b) App Usage Comparison

Fig. 5 User distribution and app usage comparison between user groups with different genders.



(a) User Distribution (b) App Usage Comparison

Fig. 6 User distribution and app usage comparison between user groups with different ages.

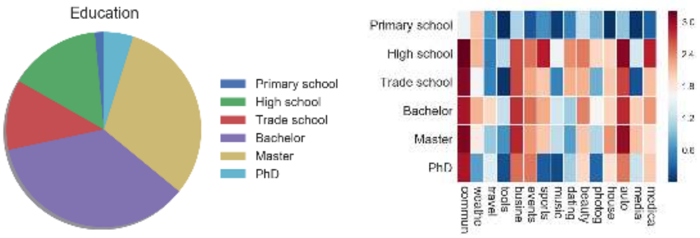


(a) User Distribution (b) App Usage Comparison

Fig. 7 User distribution and app usage comparison between user groups with different household attributes.

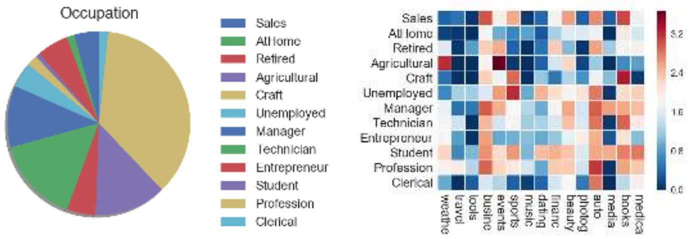
demonstrates the comparison of average category-level app usage frequency for those five groups. Clearly, we can find that young users use dating and sports app more frequently than old users, showing that users of different ages tend to have distinctive app usage behaviors.

– **Household**: We have four user groups considering different household status: *living alone, living with adult(s), living with adult(s) and kid(s), living only with kid(s).* As shown in Fig. 7 (a), the largest group is living with adult with 47.68% of users. In addition, Fig. 7 (b) plots the app usage



(a) User Distribution (b) App Usage Comparison

Fig. 8 User distribution and app usage comparison between user groups with different education levels.

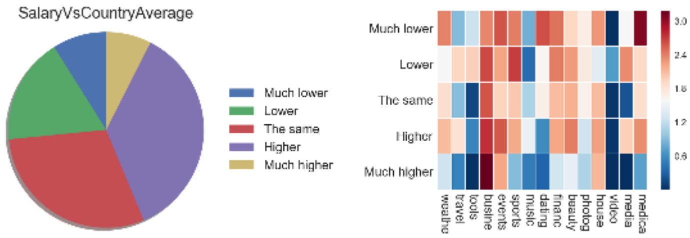


(a) User Distribution (b) App Usage Comparison

Fig. 9 User distribution and app usage comparison between user groups with different occupations.

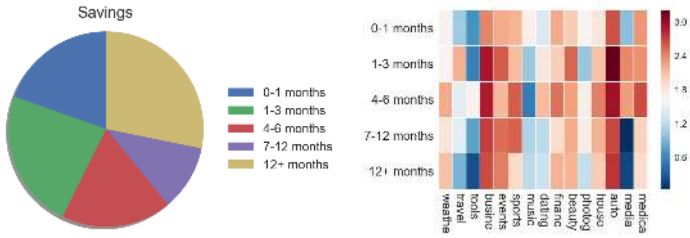
differences of these four user groups. From the figure, we can observe that users living alone will use dating apps more frequently to find for partners, and users living with adult(s) and further living with adult(s) and kid(s) will decrease the travel app usage a lot, for it is harder to make travel plans for two users or even a family with kids.

- **Education**: Considering education level, we have six user groups: *primary school, high school, trade school, bachelor, master, PhD*, and the largest group is bachelor degree with 36.69% of users. Moreover, in Fig. 8 (b), users of different education levels seems to have different app usage patterns, such as PhD users use more communication apps but less sports apps than other groups.
- **Occupation**: There are twelve groups with different occupations. As shown in Fig. 9 (a), the largest three groups are *professional, technician, student* with 36.31%, 15.08% and 12.65% of users, respectively. In addition, from Fig. 9 (b), we can find distinctive app usage patterns for users of different occupations, like sales use business apps more frequently for the need of his work.
- **Salary**: We have five salary levels: *much lower, lower, the same, higher, much higher*, compared with the average salary level in the country where the user lives in. As shown in Fig. 10 (a), the largest group is *Higher* with



(a) User Distribution (b) App Usage Comparison

Fig. 10 User distribution and app usage comparison between user groups with different salary levels.



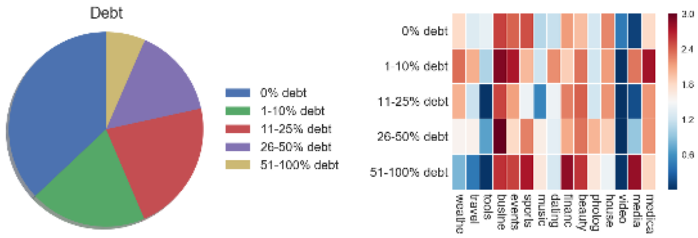
(a) User Distribution (b) App Usage Comparison

Fig. 11 User distribution and app usage comparison between user groups with different savings.

36.24% of users. In Fig. 10 (b) we can find that users with a high salary seem to use more business apps and less dating and video apps for they are busy with their work and have less time to enjoy life.

- **Savings**: We obtain the savings level of users by asking them how many months can you live on your savings if you lose your job now? As shown in Fig. 11 (a), there are five user groups: *0-1 months, 1-3 months, 4-6 months, 7-12 months, 12+ months*, and the largest group is *12+ months* with 28.19% of users. In Fig. 11 (b), we can see that there exists some differences in app usage between groups of varying savings level.
- **Debt**: In Fig. 12 (a), we have five levels of debt as proportion of monthly income of users: *0% debt, 1-10% debt, 11-25% debt, 25-50% debt, 51-100% debt*, among which the the largest group is *0% debt* with 37.03% of users. In addition, from Fig. 12 (b), we can observe that users with debt will use less travel apps than users without debt in order to reduce daily expenses.

Second, we conduct significance analysis [23] of app usage differences among different user groups, i.e., to test whether one attribute has an impact on app usage behavior. The results are shown in Table 1. For eight attributes, we conduct all the pairwise significant difference tests by setting the significant level at 0.05. We find that more than half of pairwise groups have significant app usage differences with p-value<0.05. For example, male and



(a) User Distribution (b) App Usage Comparison

Fig. 12 User distribution and app usage comparison between user groups with different debt levels.

Table 1 Significance analysis of app usage differences among different user groups.

Attributes	**Groups**	**Pairs**	**Significant Difference (p-value<0.05)**	**The most different pairs (p-value)**
Gender	2	1	100%	Female - Male: 0.0003
Debt	5	10	90%	11-25% debt - 26-50% debt: 0.2e-4 0% debt - 26-50%: 0.0001 0% debt - 51-100% debt: 0.0009
Education	6	15	86%	AtHome - Entrepreneur: 0.04e-4 Craft - Manager: 0.3e-4 Craft - Professional: 0.3e-4
Occupation	12	66	83%	AtHome - Entrepreneur: 0.04e-4 Craft - Manager: 0.3e-4 Craft - Professional: 0.3e-4
Age	5	10	60%	45-64 - 65 and over: 0.0067 35-44 - 45-64: 0.0094 18-24 - 25-34: 0.0118
Household	4	6	50%	With adult&kid - With kid: 0.0004 Alone - With kid: 0.0005 With adult - With kid: 0.0035
Salary	5	10	50%	Much lower - The same: 0.0008 The same - Higher: 0.0013 Lower - Higher: 0.0062
Savings	5	10	50%	4-6 months - 12+ months: 0.0006 4-6 months - 7-12 months: 0.0050 0-1 months - 4-6 months: 0.0251817

female users have distinctive usage behavior, which is consistent with above analysis of app usage comparison. In addition, three attributes (i.e., debt, education, occupation) are very powerful (more than 80% significance ratio) in influencing user app usage behaviors. What's more, take household for example, about 3 pairwise groups are shown to be significant different in app usage behavior, and users living only with kid(s) are quite different from users living alone (p-value=0.0005), living with adult(s) (p-value=0.0035) and living with adult(s) and kids(s) (p-value=0.0004), showing that household status does impact mobile app usage behavior a lot. Obviously, all the significance analysis has demonstrated that user app usage pattern is influenced by the social and economic attributes.

Table 2 Attribute prediction accuracy of our method.

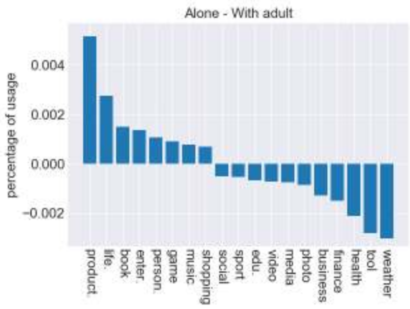
Prediction Accuracy	Debt	Educat	Househ	Occupa	Salary	Saving
SVM	0.371	0.347	0.477	0.340	0.362	0.281
Decision Tree	0.619	0.663	0.672	0.601	0.696	0.607
Our Method	0.832	0.832	0.835	0.830	0.860	0.823

Third, in order to model the correlation between app usage pattern and attributes, as mentioned before, we use the decision tree based model to extract typical app usage patterns and specific demographic groups accordingly. In order to evaluate the distinguishability of those usage patterns, we perform an external prediction task for labelled survey users, i.e., predicting user attributes by finding the most similar app usage pattern and adopting its attributes, then evaluating its prediction performance by comparing with prevalent methods Decision Tree[50] and SVM (Support Vector Machine)[11]. Note that the latter one was also utilized by Seneravithe et al. to predict the user's family size according to his apps installed on the smartphone [51]. Based on our test result, as shown in Table 2, our method can achieve over 80% accuracy in predicting different attributes, which is higher than the accuracy of SVM and Decision Tree methods. This further demonstrates the stable correlation between category-level app usage pattern and socioeconomic attributes. Therefore, all the results clearly verify that user app usage pattern can reflect user social and economic status.

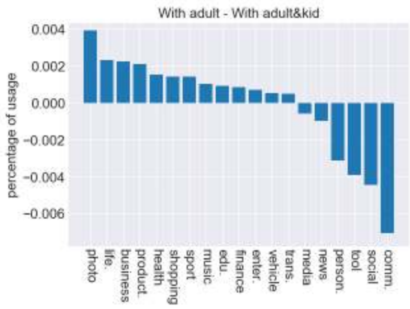
In summary, through difference significance test among different demographic groups and empirical performance evaluation of prediction task, we have found that user app usage behavior has a close relationship with demographics, social and economic attributes. According to our study, social and economic attributes have an impact on app usage that differences are found between user groups with different demographics. In addition, we have demonstrated that our extracted app usage patterns are able to distinguish different attributes by over 80% prediction accuracy, showing a stable correlation between app usage and user attributes. Therefore, we verify hypothesis: *app usage patterns reflect, and are influenced by socioeconomic status, and remain stable unless socioeconomic status also changes.*

5 Dynamics of Long-Term App Usage

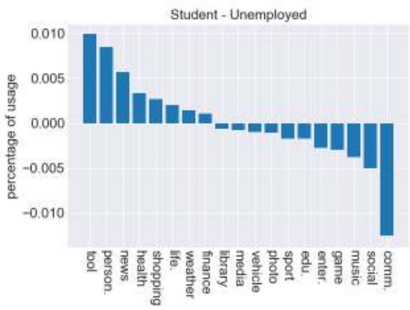
The results in the previous section have demonstrated that user app usage indeed evolves in the long run, but also remains closely connected with social and economic demographics. In this section we analyze reasons for changes in app usage and demonstrate that they mirror changes in corresponding demographics (i.e., social or economic status). More specifically, we demonstrate that we are able to identify representative change in app usage patterns and corresponding changes of demographic attributes. By analyzing category-level app usage changes and correlating them with attributes, we aim to under-



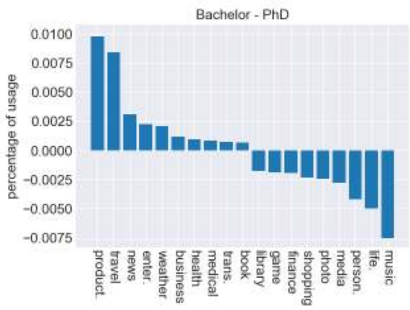
(a) Household: Change Mode 1



(b) Household: Change Mode 2



(c) Occupation: Change Mode 3



(d) Education: Change Mode 4

Fig. 13 Illustration of representative long-term app usage change modes correlated with changes of social attributes.

stand user long-term app usage changes and explore the impact of different demographic attributes.

5.1 App Pattern Changes Reflect Social Demographic Changes.

We present representative long-term app behavior change modes that are correlated with social demographic changes, including household, occupation and education attributes. For each attribute, we choose one or two typical change modes with the highest number of occurrences and show their category-level app usage changes in Fig. 13. We term these change modes 1-4.

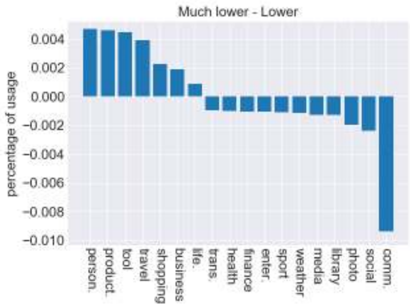
First, change modes 1 and 2 in Fig. 13(a)-(b) are correlated with changing household status. As shown in Fig. 13 (a), it shows the usage changes of different app categories when the user undergoes a household transition from living alone to living with adult(s), i.e. from single to being in a relationship. After the change, we can observe that the user uses online shopping apps more frequently but uses social apps less, possibly due to purchasing goods for two people and more face-to-face communication with his partner. Further, when a user living with other adult(s) starts to live with kids(s), i.e. having a baby, the usage of health, shopping and education apps all increased, typical for situations such as the new parent learning more knowledge from those apps to educate and care for his kid(s) as well as buying more goods. Patterns 1 and 2 therefore seem to indicate (some of the) typical behaviors associated with the corresponding social demographic changes, such as changing civil status and growing family.

We also demonstrate change modes 3 and 4 in Fig. 13(c)-(d), which have a close relationship with changes of occupation and education level. In Fig. 13 (c), reflects typical behavior when a student graduates from school and cannot find a suitable job at first, his app usage may change a lot, *i.e*, using the news apps more frequently to look for a job and turning to finance apps for economic planning, while spending less time on entertainment and social activity such as game and social apps. Finally, change mode 4 presents the app usage changes when a bachelor student becomes a PhD candidate, as shown in Fig. 13 (d), the user increases the usage of travel apps and decreases the usage of entertainment apps such as music and lifestyle category. This change is exemplified by a PhD student busy with starting a new research project and traveling more for the need of attending conferences.

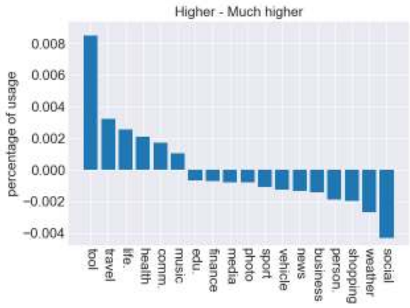
From these four long-term change modes, we clearly demonstrate that changing civil status, growing family, unemployment or pursuing a higher degree are associated with distinct app usage changes in the long term. All the results and analysis have shown that long-term app usage patterns vary in correspondence with the change of social demographics, including household, occupation and education level.

5.2 App Pattern Changes Reveal Changes in Economic Conditions.

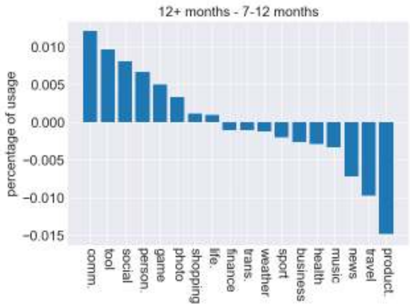
After analyzing the impact of social demographics, we discuss representative change modes with a close relationship with economic attributes, including salary, saving and debt. Four change modes 5-8 are shown in Fig. 14. To begin with, Fig. 14 (a)-(b) present the changes with a salary rise, while the main difference is the previous salary level. In other words, we want to distinguish the impact of increasing salary from a low level and increasing salary from a high level towards long-term app usage changes. change modes 5 and 6 show the app usage differences associated with this change. In Fig. 14 (a), when a user increases his salary from a low level, he uses more business and travel apps and less sports and health apps, probably due to harder working and longer working time associated with a promotion. However, in Fig. 14 (b), when a user increases his salary from a higher level, he will focus more on health, travel and lifestyle apps, showing that the user cares more about his health and enjoy life after further financial freedom. Therefore, even with the same trend, different attribute status transition may provide different influences in long-term app usage behavior, where the comparison between change mode 5 and 6 is a typical case.



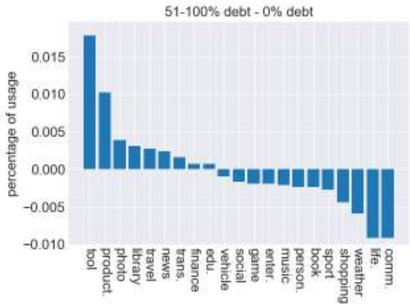
(a) Salary: Change Mode 5



(b) Salary: Change Mode 6



(c) Savings: Change Mode 7



(d) Debt: Change Mode 8

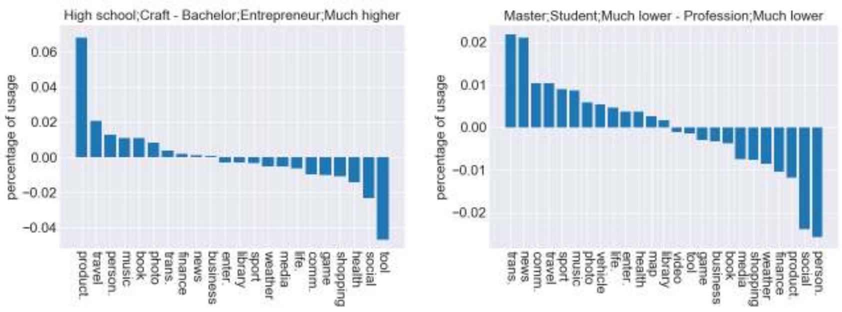
Fig. 14 Illustration of representative long-term app usage change modes correlated with changes of economic attributes.

Besides, we also investigate into the impact of different savings and debt levels on the long-term app usage patterns. Fig. 14 (c)-(d) represent weaker economic guarantee (i.e., savings decreasing) and improved economic status (i.e., paying off debt), respectively. From both figures, we can observe that travel app are used more frequently in better economic condition while game apps are used more frequently in worse economic condition. Travel may be more easily available to users with a good economic status, while mobile games are a cost-effective way to enjoy free time.

The above four long-term change modes demonstrate that increasing family income, savi gs decreasenand paying off debt can change user's app usage behaviors in the long term. All the results suggest that long-term app usage patterns reveal changes in economic conditions, including household, occupation and education level.

5.3 App Pattern Changes Are Influenced by Social and Economic Factors.

In this part, we consider both social and economic attributes, and present two representative change modes with the highest occurrence, shown in Fig. 15



(a) Multiple Attributes: Change Mode 9 (b) Multiple Attributes: Change Mode 10

Fig. 15 Illustration of representative long-term app usage change modes correlated with changes of socioeconomic attributes.

which we name as change mode 9 and 10. change mode 9 demonstrates the app usage changes for users who achieve a *learn-more-and-earn-more* change in the long term, e.g., a craft operator with high school diploma turns to earn a bachelor degree and become a entrepreneur with a much higher salary than average level in the living country. After such transformation, the user uses books, travel and business apps more frequently, and spend less time on game and shorts apps.

Moreover, change mode 10 shows a typical *early-career-after-graduation* change in the long run, e.g., a master student graduates from school and becomes a professional worker with not much salary rise in the early stage. As for app pattern changes, we can observe that the user uses more transportation and map apps typically used for commuting between office and home, and spends more time on news apps.

From these two cases, we can find that the long-term app usage changes are more complex when considering multiple attributes, but still we can discover interesting and meaningful change modes (e.g., *learn-more-and-earn-more, early-career-after-graduation*) and promote better understanding about the drivers behind in connection with social and economic factors. Taken together, the ten representative change modes shed light on the underlying reasons of long-term app usage changes by correlating with varying user attributes. This verifies hypothesis H3: *Significant changes in long-term app usage behavior occur in correspondence with changes in user social and economic demographics.*

In conclusion, through all of our analysis, we have sought to validate three hypotheses. For each of the hypotheses, we found strong evidence to support them. Below we summarize our hypotheses and the evidence supporting their validity:

H1. *User app usage behavior evolves in the long run.* As seen in Section 4.1, significant changes occur in app usage behavior over the long term in our 7-year dataset. Users not only change their applications, but also use applications of different categories, representing significant changes of user behavior. Thus, our analysis strongly supports that **H1 holds**.

H2. *App usage patterns reflect, and are influenced by socioeconomic status, and remain stable unless socioeconomic status also changes.* Section 4.2 examined the interaction between socioeconomic status and app usage patterns, and found app usage patterns to be stable with the socioeconomic status of users. In particular, app usage patterns are a good predictor (over 80% accuracy) of socioeconomic variables, and has almost double accuracy when compared against always predicting the most common value (e.g. that all users would have a bachelor level education, 36% accuracy vs our 83.2%). This shows that **H2 holds**.

H3. *Significant changes in long-term app usage behavior occur in correspondence with changes in user social and economic demographics.* In Section 5, we showed that long-term changes in app usage patterns reflect corresponding changes of socioeconomic factors. In particular, a change that took a user from one socioeconomic group to another, such as going from living alone to living with other adult(s), was accompanied by significant changes in app usage patterns. This shows that **H3 holds**.

6 Discussion

6.1 Research Implications and Potential Applications

Our study has many valuable research implications. First, our work reveals correlation between app usage and user social demographic features in a way much more fine-grained than existing literature. Specifically, we analyze the role of gender, age and various other socio economic factors (e.g., education level, occupation, salary, savings status) in shaping users' app usage behavior. As demonstrated by our experiments, it is possible to predict fine-grained user social demographic features via mobile app use data using our proposed framework, which potentially enables better user profiling in mobile computing.

Second, we present the most comprehensive study to date aiming at understanding the long-term individual app use dynamics. As we demonstrate, individual category-level app usage pattern is not static but evolving, reflecting the social demographic status change of the user, which echoes previous works on longitudinal studies of user behavior pattern [25, 42]. In essence, our findings demonstrate that long-term app use patterns, even at an aggregated category level, provide enough signals to detect user transitions in social demographic status, which point to the great potentials of user app usage tracking in context-aware social computing. In the future, researchers could potentially leverage long-term app usage data to enable real-time sensing of individual/society status, as a complement of traditional survey/census approaches in social science, which are generally expensive and time-consuming. Specifically, mobile app use could be useful to study individuals' social roles and social trajectories.

- *Social Roles:* Social science theories stipulate that humans act in various *(social) roles* throughout their life, with each role being characterized by personal and societal expectations. These roles are reflected by different social demographic factors, with particularly socioeconomic factors being prominent in determining the relevant roles of an individual. Through our analysis, we have demonstrated that mobile app usage and socioeconomic demographics have a strong and stable relationship with changes in mobile app usage reflecting changes in the individual's social categories. The stability of this relationship suggests that mobile app usage could be used as proxy to infer the social roles of an individual and to support scientific studies on social categories.
- *Social Trajectories:* Our work shows the relationship between app usage behavior and socioeconomic factors is stable over time. As a result, app usage data could be used to study social circles, e.g. friends and coworkers. With significant life events, job changes and promotions, we can determine a user's *social trajectory*, either positive or negative progression as time goes by. App usage data could be a powerful tool for identifying and helping individuals struggling with life. It could also allow us to give advice based on what app behavior changes have been seen in the past in a similar situation, that have led to an improved social trajectory.

Our work has important implications for product designers as well. One direct application of our findings is for the design of personalized recommendation system. Unlike existing approaches to smartphone user profiling, which generally assign a static label (e.g., male student) to each user, our proposed framework enables dynamic tracking of user social demographic status via his/her app usage pattern. The framework is able to detect 'subtle' changes of users, e.g., household status transition from single to married, profession transition from student to white collar, which makes possible better user profiling. In this way, more appropriate apps could be recommended to the user by app stores, so as products of most interest on e-commerce platforms.

6.2 Limitations and Future Work

Similar to other mobile data collection campaigns, our measurements inevitably suffer from some biases. The online survey data has unbalanced gender distribution with male users occupying a large part, and we only study long-term Android users (i.e., lacking iOS users) to guarantee the same mobile system platform with the survey users, since we conducted the online survey on the Android platform. On the other hand, compared with app usage data collected from mobile operators [61, 67], our data collection system takes a sample every time the battery level of the device changes by 1%. It is therefore unable to capture all the app usage information, and may be biased towards power-hungry apps (e.g., games). However, compared with existing works, our data also has unique advantages: it contains very long-term app usage behavior, and a high number of users and samples over time, resulting in a high statistical likelihood to capture the most popular applications. The online survey also complements our dataset with fine-grained and comprehensive socioeconomic information of users.

In the future, we plan to improve our data collection procedure and carry out more surveys that tracks the same user over time to better support findings of long-term user app usage pattern evolution and its relationship with individual social economic demographics change. We will also take a further study to investigate how to apply our findings to promote various applications such as user attribute identification and time-variant personalized app recommendation system.

7 Conclusion

In this paper, we provide the first ever long-term study of mobile app usage dynamics and how it is influenced by changes in socioeconomic demographics. Based on a large-scale long-term app usage dataset with an online survey about user attributes, we propose a framework to identify representative pattern changes for individuals and the corresponding changes in social and economic status. Our findings reveal that 60-70% of users change their category-level app usage patterns, but the relationship between socioeconomic demographic attributes and mobile usage patterns remains stable in the long term that our methodology can achieve high accuracy (over 80%) in predicting user demographic attributes. To better understand reasons for changes in app usage, we develop a methodology for modeling and capturing pattern changes in mobile usage data. By correlating change modes identified by our method with socioeconomic demographics, we were able to demonstrate that the changes in app usage, to a large degree, reflect corresponding changes in socioeconomic attributes, such as changing civil status, family size, changes in job or economic status.

References

1. Abrahamse W, Steg L (2009) How do socio-demographic and psychological factors relate to households direct and indirect energy use and savings? Journal of economic psychology 30(5):711–720
2. Aggarwal V, Halepovic E, Pang J, Venkataraman S, Yan H (2014) Prometheus: toward quality-of-experience estimation for mobile apps from passive network measurements. In: Proceedings of the 15th ACM Workshop on Mobile Computing Systems and Applications (HotMobile 2014), ACM, Santa Barbara, California, US
3. Almaatouq A, Prieto-Castrillo F, Pentland A (2016) Mobile communication signatures of unemployment. In: International conference on social informatics, Springer, pp 407–418

4. Althoff T, Jindal P, Leskovec J (2017) Online actions with offline impact: How online social networks influence online and offline user behavior. In: Proceedings of the tenth ACM international conference on web search and data mining, ACM, pp 537–546
5. Blaszkiewicz K, Blaszkiewicz K, Blaszkiewicz K, Markowetz A (2016) Differentiating smartphone users by app usage. In: Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp), ACM, pp 519–523
6. Cao H, Chen Z, Xu F, Li Y, Kostakos V (2018) Revisitation in urban space vs. online: A comparison across pois, websites, and smartphone apps. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2(4):156
7. Cao H, Feng J, Li Y, Kostakos V (2018) Uniqueness in the city: Urban morphology and location privacy. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2(2):1–20
8. Cao H, Xu F, Sankaranarayanan J, Li Y, Samet H (2019) Habit2vec: Trajectory semantic embedding for living pattern recognition in population. IEEE Transactions on Mobile Computing
9. Coie JD, Dodge KA (1983) Continuities and changes in children's social status: A five-year longitudinal study. Merrill-Palmer Quarterly (1982-) pp 261–282
10. Cooper JO, Heron TE, Heward WL, et al. (2007) Applied behavior analysis. Pearson/Merrill-Prentice Hall Upper Saddle River, NJ
11. Cortes C, Vapnik V (1995) Support vector networks. Machine Learning 20(3):273–297
12. Danescu-Niculescu-Mizil C, West R, Jurafsky D, Leskovec J, Potts C (2013) No country for old members: User lifecycle and linguistic change in online communities. In: Proceedings of the 22nd international conference on World Wide Web, ACM, pp 307–318
13. Do TMT, Blom J, Gatica-Perez D (2011) Smartphone usage in the wild: a large-scale analysis of applications and context. In: Proceedings of the 13th International Conference on Multimodal Interfaces (ICMI), pp 353–360
14. Dong X, Jahani E, Morales AJ, Bozkaya B, Lepri B, Pentland A (2016) Purchase patterns, socioeconomic status, and political inclination. In: International conference on computational social science
15. Dror G, Pelleg D, Rokhlenko O, Szpektor I (2012) Churn prediction in new users of yahoo! answers. In: Proceedings of the 21st International Conference on World Wide Web, ACM, pp 829–834
16. Ducheneaut N, Yee N, Nickell E, Moore RJ (2007) The life and death of online gaming communities: a look at guilds in world of warcraft. In: Proceedings of the SIGCHI conference on Human factors in computing systems, ACM, pp 839–848
17. Eagle N, Pentland AS, Lazer D (2009) Inferring friendship network structure by using mobile phone data. Proceedings of the national academy of sciences 106(36):15274–15278

18. Falaki H, Mahajan R, Kandula S, Lymberopoulos D, Govindan R, Estrin D (2010) Diversity in smartphone usage. In: Proceedings of the 8th International Conference on Mobile Systems, Applications and Services (MobiSys), pp 179–194
19. Felbo B, Sundsøy P, Lehmann S, de Montjoye YA, et al. (2016) Using deep learning to predict demographics from mobile phone metadata. In: International Conference on Representation Learning (ICLR) 2016 workshop
20. Ferreira D, Goncalves J, Kostakos V, Barkhuus L, Dey AK (2014) Contextual experience sampling of mobile application micro-usage. In: Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI), pp 91–100
21. Gordon-Larsen P, Nelson MC, Popkin BM (2004) Longitudinal physical activity and sedentary behavior trends: adolescence to adulthood. American journal of preventive medicine 27(4):277–283
22. Hata K, Krishna R, Fei-Fei L, Bernstein M (2017) A glimpse far into the future: Understanding long-term crowd worker quality. In: CSCW: Computer-Supported Cooperative Work and Social Computing
23. Hochberg Y, Benjamini Y (1990) More powerful procedures for multiple significance testing. Statistics in medicine 9(7):811–818
24. Huang K, Zhang C, Ma X, Chen G (2012) Predicting mobile application usage using contextual information. In: Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp), pp 1059–1065
25. Jessor R, Jessor SL (1977) Problem behavior and psychosocial development: A longitudinal study of youth. New York New York Academic Press 1977.
26. Jones SL, Ferreira D, Hosio S, Goncalves J, Kostakos V (2015) Revisitation analysis of smartphone app use. In: Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp), ACM, pp 1197–1208
27. Kalmus V, Realo A, Siibak A (2011) Motives for internet use and their relationships with personality traits and socio-demographic factors. Trames: A Journal of the Humanities and Social Sciences 15(4):385
28. Kang S, Jung J (2014) Mobile communication for human needs: A comparison of smartphone use between the us and korea. Computers in Human Behavior 35:376–387
29. Kent JT (1983) Information gain and a general measure of correlation. Biometrika 70(1):163–173
30. Kooti F, Yang H, Cha M, Gummadi KP, Mason WA (2012) The emergence of conventions in online social networks. In: Sixth International AAAI Conference on Weblogs and Social Media
31. Lenormand M, Louail T, Cantú-Ros OG, Picornell M, Herranz R, Arias JM, Barthelemy M, San Miguel M, Ramasco JJ (2015) Influence of sociodemographic characteristics on human mobility. Scientific reports 5:10075

32. Leroux P, Roobroeck K, Dhoedt B, Demeester P, Turck FD (2013) Mobile application usage prediction through context-based learning. Journal of Ambient Intelligence and Smart Environments 5(2):213–235
33. Li H, Lu X, Liu X, Xie T, Bian K, Lin FX, Feng F, Feng F (2015) Characterizing smartphone usage patterns from millions of android users. In: Proceedings of the Conference on Internet Measurement Conference (IMC), pp 459–472
34. Li T, Zhang M, Cao H, Li Y, Tarkoma S, Hui P (2020) what apps did you use?: Understanding the long-term evolution of mobile app usage. In: Proceedings of The Web Conference 2020, pp 66–76
35. Liao, ZhungXun, YiChin, Peng, WenChih, Lei, PoRuey (2013) On mining mobile apps usage behavior for predicting apps usage in smartphones. In: Proceedings of the 22nd International Conference on Information and Knowledge Management (CIKM), pp 609–618
36. Lim SL, Bentley PJ, Kanakam N, Ishikawa F, Honiden S (2015) Investigating country differences in mobile app user behavior and challenges for software engineering. IEEE Transactions on Software Engineering 41(1):40–64
37. Lin J, Sugiyama K, Kan MY, Chua TS (2013) Addressing cold-start in app recommendation: latent user models constructed from twitter followers. In: Proc. ACM SIGIR, pp 283–292
38. Lin Z, Althoff T, Leskovec J (2018) I'll be back: On the multiple lives of users of a mobile activity tracking application. In: Proceedings of the 2018 World Wide Web Conference on World Wide Web, International World Wide Web Conferences Steering Committee, pp 1501–1511
39. Lipsman A, Lella A (2017) The 2017 us mobile app report. Retrieved from comScore: http://www. comscore. com/Insights/Presentations
40. Malmi E, Weber I (2016) You are what apps you use: Demographic prediction based on user's apps. In: Tenth International AAAI Conference on Web and Social Media, AAAI
41. Meeker M, Wu L (2018) Internet trends 2018. In: Kleiner Perkins
42. Moffitt TE (2017) Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. In: Biosocial Theories of Crime, Routledge, pp 69–96
43. Ollendick TH, Weist MD, Borden MC, Greene RW (1992) Sociometric status and academic, behavioral, and psychological adjustment: a five-year longitudinal study. Journal of consulting and clinical psychology 60(1):80
44. Park MH, Hong JH, Cho SB (2007) Location-based recommendation system using bayesian user's preference model in mobile devices. In: Proceedings of the 4th International Conference on Ubiquitous Intelligence and Computing (UIC), pp 1130–1139
45. Peltonen E, Lagerspetz E, Hamberg J, Mehrotra A, Musolesi M, Nurmi P, Tarkoma S (2018) The hidden image of mobile apps: Geographic, demographic, and cultural factors in mobile usage. In: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services, ACM, New York, NY, USA, MobileHCI '18, pp 10:1–10:12, DOI 10.1145/3229434.3229474, URL http://doi.acm.org/

10.1145/3229434.3229474
46. Pennacchiotti M, Popescu AM (2011) A machine learning approach to twitter user classification. In: Fifth International AAAI Conference on Weblogs and Social Media
47. Preece J, Shneiderman B (2009) The reader-to-leader framework: Motivating technology-mediated social participation. AIS transactions on human-computer interaction 1(1):13–32
48. Rachuri KK, Rachuri KK, Rachuri KK, Rachuri KK, Tapia EM, Tapia EM (2014) Mobileminer: mining your frequent patterns on your phone. In: Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp), pp 389–400
49. Ros M, Pegalajar M, Delgado M, Vila A, Anderson DT, Keller JM, Popescu M (2011) Linguistic summarization of long-term trends for understanding change in human behavior. In: 2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011), IEEE, pp 2080–2087
50. Safavian SR, Landgrebe D (1991) A survey of decision tree classifier methodology. IEEE transactions on systems, man, and cybernetics 21(3):660–674
51. Seneviratne S, Seneviratne A, Mohapatra P, Mahanti A (2014) Predicting user traits from a snapshot of apps installed on a smartphone. ACM SIGMOBILE Mobile Computing and Communications Review 18(2):1–8
52. Seneviratne S, Seneviratne A, Mohapatra P, Mahanti A (2015) Your installed apps reveal your gender and more! ACM SIGMOBILE Mobile Computing and Communications Review 18(3):55–61
53. Shameli A, Althoff T, Saberi A, Leskovec J (2017) How gamification affects physical activity: Large-scale analysis of walking challenges in a mobile application. In: Proceedings of the 26th International Conference on World Wide Web Companion, International World Wide Web Conferences Steering Committee, pp 455–463
54. Shi K, Ali K (2012) Getjar mobile application recommendations with very sparse datasets. In: Proc. ACM KDD, pp 204–212
55. Shin C, Hong JH, Dey AK (2012) Understanding and prediction of mobile application usage for smart phones. In: Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp), pp 173–182
56. Sigg S, Peltonen E, Lagerspetz E, Nurmi P, Tarkoma S (2019) Exploiting usage to predict instantaneous app popularity: Trend filters and retention rates. ACM Transactions on the WEB
57. Singh VK, Bozkaya B, Pentland A (2015) Money walks: implicit mobility behavior and financial well-being. PloS one 10(8):e0136628
58. Skinner BF (1953) Science and human behavior. 92904, Simon and Schuster
59. Sundsøy P, Bjelland J, Reme BA, Iqbal AM, Jahani E (2016) Deep learning applied to mobile phone data for individual income classification. In: 2016 International Conference on Artificial Intelligence: Technologies and Applications, Atlantis Press

60. Telama R, Yang X, Viikari J, Välimäki I, Wanne O, Raitakari O (2005) Physical activity from childhood to adulthood: a 21-year tracking study. American journal of preventive medicine 28(3):267–273
61. Tu Z, Li R, Li Y, Wang G, Wu D, Hui P, Su L, Jin D (2018) Your apps give you away: distinguishing mobile users by their app usage fingerprints. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2(3):138
62. Veldhuis R (2002) The centroid of the symmetrical kullback-leibler distance. IEEE signal processing letters 9(3):96–99
63. Xu F, Xia T, Cao H, Li Y, Sun F, Meng F (2018) Detecting popular temporal modes in population-scale unlabelled trajectory data. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2(1):46
64. Xu Q, Erman J, Gerber A, Mao Z, Pang J, Venkataraman S (2011) Identifying diverse usage behaviors of smartphone apps. In: Proceedings of the Conference on Internet Measurement Conference (IMC), pp 329–344
65. Yang D, Kraut R, Smith T, Mayfield E, Jurafsky D (2019) Seekers, providers, welcomers, and storytellers: Modeling social roles in online health communities. In: ACM CHI Conference on Human Factors in Computing Systems, ACM
66. Yang J, Wei X, Ackerman MS, Adamic LA (2010) Activity lifespan: An analysis of user survival patterns in online knowledge sharing communities. In: Fourth International AAAI Conference on Weblogs and Social Media, AAAI
67. Yu D, Li Y, Xu F, Zhang P, Kostakos V (2018) Smartphone app usage prediction using points of interest. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1(4):174
68. Zastrow C, Kirst-Ashman K (2006) Understanding human behavior and the social environment. Cengage Learning
69. Zhang M, Tang J, Zhang X, Xue X (2014) Addressing cold start in recommender systems: a semi-supervised co-training algorithm. In: SIGIR
70. Zhao S, Ramos J, Tao J, Jiang Z, Li S, Wu Z, Pan G, Dey AK (2016) Discovering different kinds of smartphone users through their application usage behaviors. In: Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp), pp 498–509
71. Zhu L, Gonder J, Lin L (2017) Prediction of individual social-demographic role based on travel behavior variability using long-term gps data. Journal of Advanced Transportation 2017
72. Zuniga A, Flores H, Lagerspetz E, Tarkoma S, Manner J, Hui P, Nurmi P (2019) Tortoise or hare?: Quantifying the effects of performance on mobile app retention. In: Proceedings of the 2019 World Wide Web Conference (WWW '19)