Gender Profiling from a Single Snapshot of Apps Installed on a Smartphone: An Empirical Study

Sha Zhao, Yizhi Xu, Xiaojuan Ma, Ziwen Jiang, Zhiling Luo, Shijian Li, Laurence Tian-ruo Yang, Anind Dey, and Gang Pan

***Abstract*—The integration of 5G networks and AI benefits to create a more holistic and better connected ecosystem for industries. User profiling has become an important issue for industries to improve company profit. In the 5G era, smartphone applications have become an indispensable part in our everyday lives. Users determine what apps to install based on their personal needs, interests, and tastes, which is likely shaped by their genders - the behavioral, cultural, or psychological traits typically associated with one sex. It is possible to profile users' gender based simply on a single snapshot of apps installed on their smartphones. With this inference based on easy to access data, we can make smartphone systems more user-friendly, and provide better personalized products and services. In this paper, we explored such possibility through an empirical study on a large-scale dataset of installed app lists from 15,000 Android users. More specifically, we investigated the following research questions: *1) What differences between females and males can be explored from installed app lists? 2) Can user gender be reliably inferred from a snapshot of apps installed? Which snapshot feature(s) are the most predictive? What is the best combination of features for building the gender prediction model? 3) What are the limitations of a gender prediction model based solely on a snapshot of apps installed on a smartphone?* We found significant gender differences in app type, function, and icon design. We then extracted the corresponding features from a snapshot of apps installed to infer the gender of each user. We assessed the gender predictive ability of individual features and combinations of different features. We achieved an accuracy of 76.62% and AUC of 84.23% with the best set of features, outperforming the existing work by around 5% and 10%, respectively. Finally, we performed an error analysis on misclassified users and discussed the implications and limitations of this work.**

***Index Terms*—Installed app lists, smartphones, gender, user studies.**

I. Introduction

THE evolving of the fifth generation (5G) networks is becoming more readily available nowadays, providing new connectivity interfaces for the future IoT applications [1]. As with the advent of 5G, a much larger amount of data is captured and stored on dependable communication networks. From the abundant data, advanced AI techniques benefit to learn valuable knowledge to serve for IoT applications. The integration of 5G and AI can lead to the expanding of industrial automation, by directly contributing to revolutionizing smart business solutions. Users play an important role in improving the smart business solutions, since most services are user-centric and provided depending on users' needs and interests. Thus, user profiling has been becoming an important issue in the era of industrial automation, which is a process of identifying users' attributes, such as needs and preferences, from user sensing data.

User profiling helps personalization of applications to improve company profit, such as targeted advertising. For example, in 2017, advertising accounts for the vast majority of Feacebook's revenue, and about 98 percent of its global revenue was generated from advertising. [2] There are various kinds of user attributes, such as occupation, income and preferences. Among these user attributes, gender has shown enormous significance [1]. Gender plays a fundamental role in our daily lives, shaping identities and perception, interactional practices, and the very forms of social institutions [2]. People respond and behave differently according to gender [3]. Hence, taking gender into consideration could have great impact in improving smart applications or systems, especially most systems with an interface, in such a way as to make them more user-friendly and act more human-like [3]. The knowledge can be further used for improving company profit by facilitating appropriate and efficient marketing of products and services, such as personalized recommendation and personalized searching. Appropriate services can be recommended to females or males according to their differences. Thus, automatically inferring gender is important for industries.

In the 5G era, the prevalence of smartphones [4] and the rapid development of AI techniques have made it possible to *automatically infer gender for industry*. Applications (*Abbr.* apps) installed on a smartphone provide the opportunity to profile user gender in a lightweight fashion. First, users determine what apps to install based on their personal needs, interests, and tastes, which is likely shaped by their gender–the behavioral, cultural, or psychological traits typically associated with one sex [2]. The list of installed apps is a potentially

S. Zhao, Y. Xu, Z. Jiang, Z. Luo, and S. Li are with the College of Computer Science and Technology, Zhejiang University, Hangzhou, Zhejiang, 310027 China. E-mail: {szhao, yzxu, zjujzw, luozhiling, shijianli}@zju.edu.cn.

X. Ma is with Hong Kong University of Science and Technology, Hongkong, China. Email: mxj@cse.ust.hk

L.T. Yang is with St. Francis Xavier University, Antigonish, NS, Canada. Email: ltyang@ieee.org

A. Dey is with University of Washington, Seattle, WA, USA. E-mail: anind@uw.edu

G. Pan is with the State Key Lab of CAD&CG, Zhejiang University, Hangzhou, China, and College of Computer Science and Technology, Zhejiang University, Hangzhou, China

Dr. Gang Pan is the corresponding author. E-mail: gpan@zju.edu.cn

Manuscript received May 28, 2019; revised July 17, 2019.

[1] https://www.i-scoop.eu/internet-of-things-guide/5g-iot/

[2] https://www.statista.com/statistics/271258/facebooks-advertising-revenue-worldwide/

good indicator of gender, since females and males may seek different functions and values in smartphone apps.

Second, the list of apps installed on a smartphone is accessible to other applications. Although the permission for an app to access personal data, such as contacts and location, must be requested from a user at the time of app installation in Android, the list of apps installed on a user's smartphone can be obtained without his/her permission through any app installed in Android [5]. Some advertisement tracking libraries have reportedly embedded this feature to collect lists of installed apps [5]. Third, the list of installed apps is lightweight, only a single snapshot of the apps installed on a smartphone. It does not need continuous tracking of users' activities and maintaining historical records which are expensive and time-consuming. With the help of advanced techniques, we have the potential to quickly determine one user's gender at the moment when accessing his/her current list of apps installed on the smartphone. It can be used in conjunction with user behavior tracking technologies to address the cold-start issues in tracking-based systems, especially at the beginning of data collection for an individual. Despite the advantages, there have been few studies so far to systematically profile user gender from a single snapshot of apps installed on a smartphone.

In this paper, we present an empirical study of gender profiling from a single snapshot of apps installed on a smartphone. We try to answer three research questions that guide the remainder of the this paper:

- **RQ1:** What differences between females and males can be explored from the list of apps installed?
- **RQ2:** Can user gender be reliably inferred from a snapshot of apps installed? Which snapshot feature(s) are the most predictive? What is the best combination of features for building the gender prediction model?
- **RQ3:** What are the limitations of a gender prediction model based solely on a snapshot of apps installed on a smartphone?

In order to answer each of our research questions, we conducted experiments based on a large-scale dataset of installed app lists collected from 7500 females and 7500 males. Our key findings are as follows.

(1) Females and males have distinctive differences in terms of top popular apps, the key functions they seek, and trend in icon designs at both an individual app and app category level.

(2) Features extracted from a snapshot of apps installed on a smartphone can be used to infer the gender of its user. We investigated the predictive ability of individual features and found that the discriminative apps selected by IG (Information Gain) are the most powerful for distinguishing gender. The combination of apps selected by IG and topics learned from app description by LDA (Latent Dirichlet Allocation) leads to the best gender classification results, outperforming the existing work by around 5% in accuracy (76.62% *vs.* 71.5%), and about 10% in AUC (84.23% *vs.* 74.0%). This performance is nearly on par with the state-of-the-art gender profiling results based on months of app usage data.

(3) Further investigation of misclassified users suggest profitable lines of future work. We conducted an error analysis about the users we had difficulty inferring to understand our model's limitations, and provided possible reasons for these misclassifications.

II. Related work

Studies on gender differences have a long history. Over the past few decades, gender differences have been studied using traditional methods, i.e., questionnaires. Although questionnaires are cost-efficient, lack of conscientious responses can be a serious issue, since respondents usually answer questions subjectively. As with the increasing development of devices and technologies, more and more personal history data of users is recorded in cyber space, such as data about social networks, web searching, and mobile phone usage. Big data about users in cyber space provides a great opportunity for studying gender objectively and extensively. The captured history data describes users' behaviors objectively. Additionally, this data can be continuously recorded in detail for a long duration. Many studies have revealed the gender differences through analyzing users' behaviors in cyber space, and even inferred gender. In this section, we review the related work in the two areas, namely the studies about gender differences of behaviors in cyber space and inferring gender from behaviors.

A. Gender differences of behaviors in cyber space

Differences in behaviors in cyber space based on gender have been learned through the analysis of various datasets, such as social network usage, web searching and mobile phone usage. A variety of datasets have been used to analyze gender differences in *social network usage* [6]–[10] and *web searching behaviors* [11]–[13]. For example, Mazman *et al.* [6] learned the difference in usage purpose of social networks between females and males by investigating 870 Facebook users. It was found that females use Facebook for "maintaining existing relationships", "academic usage" and "following agenda" more than males do, while males only use Facebook for "making new relationships" more than females. Zhou reported that males were engaged in more search activities than females, as seen in the large number of searches, search queries, and times males updated their queries [11].

There are fewer studies analyzing gender differences in *mobile phone usage*. Andone *et al.* showed females used smartphones for longer periods of time than males, with a daily mean of 166.78 minutes vs. 154.26 minutes [14]. They simply calculated the daily mean phone usage time and usage duration of app categories for females and males. As with the rising development of mobile app market, the apps on smartphones provide a great opportunity for analyzing gender differences. Compared with logs about social network usage and web searching, apps on smartphones can reflect more comprehensive and personalized user information. Smartphone apps can be considered as the entry point to access many life services. The apps installed on a smartphone reveal rich clues regarding one user's identity and information, since the smartphone is usually associated with the same user.

Unfortunately, there are very few studies extensively analyzing gender differences from installed apps on smartphones.

B. Gender inferring from behaviors

Since there are gender differences of behaviors in cyber space, many studies have sought to infer gender from the behaviors, such as social network usage and web searching [15]–[17]. For example, Rao *et al.* [15] inferred 500 users' gender from Twitter language, with an accuracy of 72%. Bi *et al.* inferred the gender of 3.3 million users with AUC of 80%, by analyzing their Bing query logs [16].

There have been also studies inferring gender from mobile phone usage behaviors [18]–[23]. [18] and [24] used GPS, call log, media, Bluetooth, calendar, acceleration, and application use frequency to predict about 200 phone users' gender. Malmi *et al.* analyzed the apps used by 3,760 Android users, and inferred gender using logistic regression with an accuracy of 82.3% [22]. Wang *et al.* [23] inferred 25 users' gender from app usage behaviors with an accuracy of 86.5%. However, all of these analyses required data from users' phones collected over time, which is expensive and time consuming. There have been very few studies using installed app lists to infer user gender.

Seneviratne *et al.* analyzed about 200 users' installed app lists and discovered the gender differences in installed app lists [19]. The dataset was collected through a specific Android app from a relatively small population. They extracted different features from the number of installed apps, categories, and app description, then trained SVM and Naive Bayes classifiers to infer gender and obtained an accuracy of around 70%. Compared with Seneviratne's work [19], our work has the following differences. First, we explored a much bigger dataset with 15,000 users, and our dataset is natural, not artificially collected as [19]. Second, we extracted new features and new feature combinations, and inspected each feature at both intra-category and inter-category levels. Third, we compared more state-of-the-art algorithms for gender prediction, and achieved the state-of-the-art performance. Finally, we performed an error analysis on misclassified users.

Based on the large-scale dataset of installed app lists from 15,000 Android users, we present an empirical study of gender profiling from a single snapshot of apps installed on a smartphone. Firstly, we explore the gender differences in terms of popularity, functions, and icons of apps installed at both individual app and app category level. Then, we extract features for user representation from a snapshot of apps installed on smartphones, and investigate the predictive ability of individual features and combinations of different features for gender. Finally, we perform an error analysis on misclassified users, and discuss the implications and limitations of our study.

III. Dataset

The dataset we used to profile user gender consists of app lists installed on Android smartphones, provided by a mobile Internet company in China. The users' installed app lists were collected through an embedded advertising SDK (Software Development Kit) in apps on Android smartphones. In addition to the advertisements displayed to users, a brief questionnaire about demographic attributes such as gender was present. Users voluntarily answered the questionnaire, and the user who reported the demographic attribute was offered a compensation with a 20 Yuan e-coupon. The dataset contains 16,003 unique smartphone users with self-reported gender information, and 29,788 unique apps, in which there are 7,500 females and 8503 males from multiple provinces in China. To keep the balance, we randomly selected 7,500 males and there are 15,000 unique users in total used in this work. Each record consist of a:

- User ID: the unique identity of the sampled smartphone. Each user ID is anonymized for security and privacy reasons before the data was collected.
- Gender: the gender of the user.
- Installed app list: the apps installed on one smartphone. Each list consists of app package names which are used to identify apps.

In addition to the installed app lists, we crawled the meta data of the apps in the dataset from appstore websites, including *icons*, *description* that introduces an app, and *tags* that describe core functionality and characteristics of each app. In order to get a high level understanding of the apps installed by users, we categorized all the apps into 29 *categories* [25].

User privacy. The original data was collected for the purpose of improving user experience, with a strict policy of data collection, transmission and storage. The gender information was collected with explicit user agreements. In this work, we take careful steps to protect user privacy. First, the dataset was completely anonymized before provided to the authors, especially the user ID. Second, throughout the paper, we report only average statistics without revealing any identifiable information of individuals. Third, all the researchers are regulated by the strict non-disclosure agreement and the dataset is located in a secure off-line server.

Basic analysis. We calculated the distribution of the number of apps installed by users. Most users install between 20-60 apps, accounting for about 60% of our population, with just very few users install more than 100 apps. Each user in our data set installed 34.87 apps on average (median: 30, std: 22.35). The top 5 most frequently installed apps are WeChat, QQ (an IM client), Taobao (a shopping app like Amazon), SinaWeibo (a social network app like Twitter), and Sogou Input method (a Chinese language input method). We found that the top 1,288 frequent apps (out of 29,788) account for 80% of all the installations.

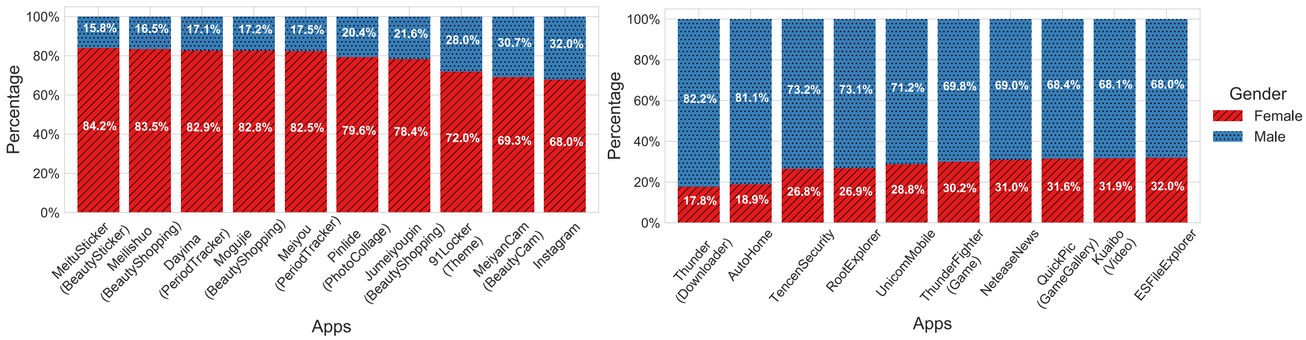
The 5 most popular categories are SON_and_IM (social online network and instant messaging), System tool, Media_and_video, Lifestyle (consisting of apps that can make people's lives easier, *e.g.* flashlight), and Shopping.

IV. The Gender differences

We explored the gender differences in terms of popularity, functions and icons of apps installed, at both the individual app and app category level.

A. The differences in the number and popularity

1) ***Male users install more apps than female users in general:*** We firstly compared the average number of apps



(a) (b)

Fig. 1: The top 10 popular apps for (a) females, and (b) males.

and categories installed by females and males. Each female installed 32.32 apps on average (median: 29, std: 19.39), while males installed 37.41 (median: 32, std:24.69). The difference is significant under a two-detailed z-test [26] with a $z - score = 14.05 (p < 0.01)$. It can be concluded that males installs about 5 more apps than females on average. Similarly, male users install apps that represent more app categories than female users: males install apps in 13.63 app categories on average, compared to females with 12.80. This difference is also significant under a two-detailed z-test [26] with $z - score = 7.5386 (p < 0.01)$.

*2) **Beauty-related apps attract more female users, while users of system tool and car apps are mostly male.**:* We then investigated the top popular apps for females and males, respectively. A top popular app for females refers to an app is popular among females but not so popular among males. It means that there is a distinct difference in the number of females and males. We focused on the frequent apps that are installed by more than 500 users. To be specific, for females we computed the difference of each app in the percentage of its female and male users. Then, we ranked the apps in a descending order according to the difference. We summarized the top 10 popular apps for females, shown in Fig. 1 (a). Similarly, we computed those for males, shown in Fig. 1 (b).

As shown in Fig. 1 (a), beauty-related apps are more attractive to female users. The top 10 popular apps for females are the ones related to photo beautifying and sharing (MeituSticker,Pinlide, MeiyanCam, and Instagram), beauty shopping (Meilishuo, Mogujie, and Jumeiyoupin), period tracker (Dayima and Meiyou), and theme (91 Locker). The apps of MeituSticker,Pinlide, MeiyanCam, and Instagram provide services for photography, taking pictures with smart beautifying functions, and sharing photos. The apps of Meilishuo, Mogujie, and Jumeiyoupin are for fashion e-commerce with clothing and cosmetics targeting females. The apps of Dayima and Meiyou are designed for a woman to track her menstrual cycle. For these apps, the female users account for more than 68%, especially for the apps of MeituSticker (beauty sticker) and Meilishuo (beauty shopping), the female users accounting for 84.2% and 83.5%, respectively. But, only around 16% of male users install MeituSticker and Meilishuo. The great difference shows that females have the beauty-related apps much more than males.

As we can see from Fig. 1 (b), tools and car-related apps are the most attractive to males. The top 10 popular apps for males are related to downloading (Thunder), cars (AutoHome), system tools (Tencent Security, Root Explorer, and UnicomMobile), games (ThunderFighter), news (NeteaseNews), videos (Kuaibo), and file managers (QuickPic-PhotoGallery and ESFileExplorer). For the apps of Thunder (downloader) and AutoHome (Car), males account for 82.2% and 81.1%, respectively, while female users account for only 17.8% and 18.9%. Thunder is a downloading tool, and Autohome provides services about cars, technologies, and purchasing.

*3) **Females install more photo and shopping types of apps, while males have more tool, game, and video apps on the phones.** :* We also calculated the popular categories for females and males, by comparing the average number of apps in each category installed by female and male users. The categories were ranked in a descending order based on the difference. For 23 out of 29 categories, males install more apps than females on average. This is to be expected, since males install more apps than females in general. The category with the greatest difference between females and males is system tool, for which each male installs about 4.8 apps on average while each female installs about 3.8. Males also install more apps on average than females in the categories of game_other (role playing games, action games, simulation games, and adventure games), media and video, news and reading, business, navigation, car and travel. But in the category of photography_and_beauty, females install about twice as many apps as males on average, 2.1 apps per female and 1.1 apps per male. Females also have more shopping and parent_and_child apps than males. Females install as many apps as males, on average, in the categories of social network and IM, education, and weather.

B. The differences in app functions

Considering females and males may seek different functions and values in smartphone apps because of different needs and interests, we explored the gender differences in app functions at both an individual app level and app category level.

*1) **Females are interested in apps that improve their look and feel, while males like apps that improve their access to the world.**:* In order to discover the gender differences in app functions at the individual app level, we extracted the app functions through discovering the semantic topics from the description text of apps. Each topic indicates one kind of function, such as playing music and taking pictures. In order to learn latent semantic topics, each user was regarded as a document consisting of the words appearing in the description of all his/her installed apps, and all the users constituted a corpus. Each user (document) is considered to have a set of topics that can be learned from his/her words. Here, we applied LDA to learn the semantic topics from the app description to extract the app function. More specifically, we extracted words using Jieba, a tool for Chinese text segmentation, to select nouns, verbs and adjectives, and used words to represent each user as a vector input to LDA. LDA assigned a user multiple topics with a probability distribution of the topics, indicating the probability that the topic belongs to the user. Each topic consists of a probability distribution of the words, indicating the probability that the word belongs to the topic. Fig. 2 illustrates how to get topics from app description.

We obtained 300 topics and each has a probability distribution of words from LDA (the choice of 300 topics is explained in Section V-C3). In order to understand the gender differences in app functions, we computed the significant topics for females and males, respectively. The significance of the ith topic s_i was computed by Equation 1, i.e., the absolute value of the difference between the average probability each female belongs to the ith topic and the average probability each male belongs to the ith topic. If the average probability each female belongs to the ith topic is bigger than that of each male, the ith topic is one significant topic for females, otherwise it is one significant topic for males. We ranked the topics in a descending order according to their significance to females and males, respectively. The first topic in the ranking is the one that females (or males) are the most interested in.

$$s_i = \left|\frac{\sum_{j=1}^{7500} F_{i,j}}{N_F} - \frac{\sum_{k=1}^{7500} M_{i,k}}{N_M}\right| \tag{1}$$

where $F_{i,j}$ and $M_{i,k}$ refers to the probability that the ith topic belongs to the jth female and kth male, respectively,

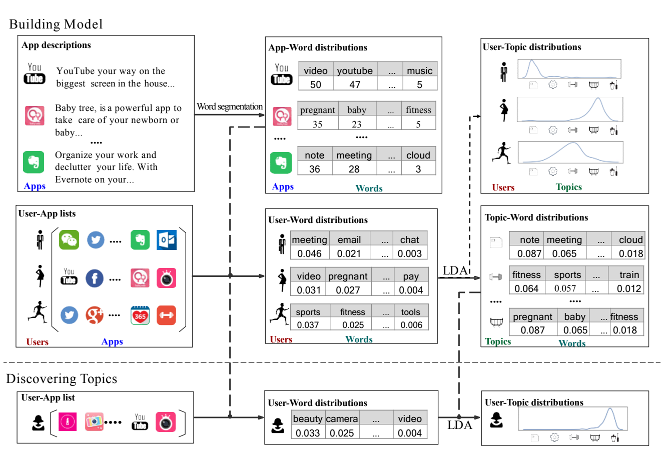


Fig. 2: The illustration of how to get topics from app description.

(a) The top 5 topics for female users.

(b) The top 5 topics for male users.

Fig. 3: The word clouds of the top 5 topics for females and males, respectively.

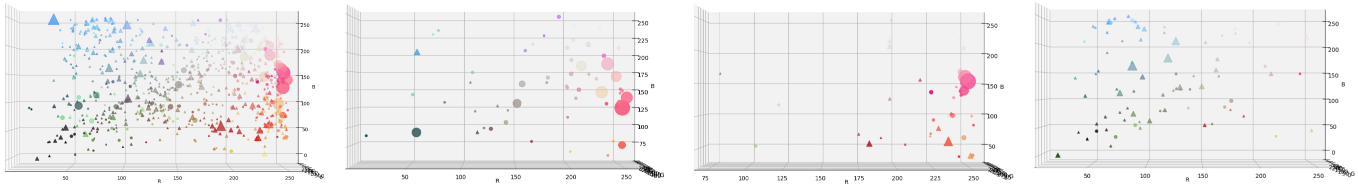
and $N_F = 7500$ and $N_M = 7500$ are the total number of females and males.

To give a sense of what function each topic expresses, we selected the top 30 words with the highest probability for each topic and generated a word cloud. We listed the word clouds of the top 5 topics for females and males, respectively. For each of the listed topics, we also computed the average probability for each female p_F, and for each male p_M, shown in Fig. 3. In each word cloud, the size of one word indicates the probability that it belongs to the topic, with higher probability corresponding to bigger size.

Figure 3 (a) shows the word clouds of the 5 most significant topics for female users, and the average probability of each topic for females and males. It can be seen female users are more interested in apps that improve their look and feel. The top 5 functions that females are interested in are about camera, shopping, photography, collocation, and slimming. When we looked into the most attractive function, camera-related, the average probability for females is twice as much as for male (0.01629 *vs.* 0.00765). As was expected, females have higher probability than males for shopping (0.01126 *vs.* 0.00263), photography (0,01454 *vs.* 0.00614), collocation (0.01071 *vs.* 0.00542), and slimming (0.00653 *vs.* 0.00159).

As we can see from Fig. 3 (b), compared with females, males are more interested in apps with functions about cars, live broadcasting, accessing WiFi, maps, and travel that can improve their access to the world. Males have much higher probability to these 5 functions than females: car (0.06597 vs. 0.01923), live broadcasting (0.04529 *vs.* 0.00105), WiFi (0.01143 *vs.* 0.00851), map (0.01134 *vs.* 0.00855), and travel (0.01035 *vs.* 0.00767), especially for functions related to cars for which the probability for males is around three times bigger than that for females. The findings for the top 5 functions correspond to our earlier finding that the users of tool and car related apps are mostly male.

*2) **Within the same app category, females and males have different preferences to apps.**:* Apps in the same category can have different core functions; for example, PictureEditor and Meitu in the category of Photography_and_beauty. PictureEditor, like photoshop, is for professionally editing pictures, while Meitu is for beautifying and customizing photos, such as smoothing skins and enlarging eyes. We looked into the gender differences in app functions within the same category,



(a) The top 1000 apps (b) Photography_and_beauty (c) Shopping (d) System tool

Fig. 4: The apps mapped into RGB space (‘△’: male, ‘o’: female).

(a) Photography_and_Beauty (b) Health_and_fitness (c) Media_and_video (d) SON_and_IM (e) Shopping

Fig. 5: The word clouds of the five popular categories for females and males, respectively (F: Female, M: Male, SON: Social Online Network).

and found females and males have different preferences for apps in a particular category. Here, we investigated five popular categories: Photography_and_beauty, Health_and_fitness, Media_and_video, Shopping, and SON_and_IM. The word clouds for females and males were generated by selecting top 30 significant tags of the apps in each category, respectively, shown in Fig. 5. The significance of each tag is its TF-IDF (Term Frequency-Inverse Document Frequency).

As shown in Fig. 5 (a), the comparison shows that females and males have distinct differences in needs and preferences. Within the category of Photography_and_beauty, males install more apps that support professionally editing pictures, picture management, and making funny photos, while females show more preference to apps that support beautifying photos, taking selfies or recording videos with smart beautifying functions. For the apps in Health_and_fitness in Fig. 5 (b), females install ones which are used for slimming, losing weight, diet, menstruation, and parenting, whereas males install more apps for workout, fitness, sports, and medical health. As shown in Fig. 5 (c), the most significant tag in Media_and_video for females is ‘editor’, and that for males is ‘player’. Females install more apps for editing videos, but males have more apps for playing videos. In addition, females pay more attention to short movies, drama, and shows, while males are interested in content about nijigen (anime, comic and games), HD movie, and hot girls. Within the category of SON_and_IM in Fig. 5 (d), males install apps about dating, love, and searching soulmate. Different from males, females have more apps that can be used for photography and beautifying movies. For shopping apps in Fig. 5 (e), females are interested in the goods on sale, luxury goods, and cosmetics, while males pay more attention to electric appliances, sporting goods, and drinks.

C. The differences in app icons

We also investigated the gender differences in app icon design. Here, we focused on the top 1000 discriminative apps selected by the method of IG, to analyze the *icon color preferences* of female and male users. IG measured each app’s relevance to gender, which measures the expected reduction in entropy by learning the state of a random variable [27]. The higher the information gain, the greater the relevance of the app to gender. We ranked the apps according to their information gain, and the first app is considered to be the most discriminative one for gender.

*1) **Females have more apps in pink, while males install more apps in different shades of blue.**:* At the individual app level, we mapped the top 1000 discriminative apps selected via IG into the RGB space, as shown in Fig. 4(a). Specifically, we averaged the R, G, B values of all the pixels in each icon, respectively, and then mapped the icon into the RGB space based on the averaged R, G, B values. For each selected app, the app is labeled with ‘△’ if there are more male users than females, otherwise, it is labeled with ‘o’. The size of each app in the figure corresponds to its information gain.

As we can see from Fig. 4(a), there are more ‘△’ than ‘o’. Among the top 1000 discriminative apps, there are 739 apps in ‘△’ which are installed by more males and 261 in ‘o’. It is reasonable because males install 5 more apps than females in average in our dataset. Compared to males, females have more apps in pink and males install more apps in different shades of blue. There are more pink ‘o’ than pink ‘△’, indicating pink apps are installed by more females than males. There is a cluster of apps labeled with pink ‘o’s in the right where the values of both R and B are bigger, and the information gain of these apps is very big, such as Mogujie-beautyshopping and Meituxiuxiu-photography with the biggest information gain of 0.029 and 0.027. In the right bottom, there is a cluster of apps labeled with red ‘△’, implying red apps are installed by more males. Many ‘△’ are located in the left top and left bottom, where the R (IG) values are small overall. In particular, the big ‘△’ in the left top corresponds to the app of AutoHome, with an information gain of 0.021. Taken together, it means that males install more apps with light blue and dark blue icons.

*2) **Within the same app category, females install more pink ones, versus blue for males.**:* We looked into the icon color of the apps in the same category. There were some categories where the app icons were not available, such as the category of Health_and_fitness. We analyzed the categories where the icons of most apps were available. The apps in one category were mapped into the RGB space, in a similar way as Fig. 4(a). We found that females have more pink apps, versus blue for males, in the categories of Photography_and_beauty, Shopping, and System_tool in Fig. 4.

As we can see from Fig. 4(b), for the category of Photography_and_beauty, there is a small cluster in which the apps with pink icon are labeled with ‘o’. There are many more pink apps labeled with ‘o’ than labeled with ‘△’. Males install more apps in blue (locate on the left of the graph) where the R value is relatively small. Similarly, in the category of Shopping shown in Fig. 4(c), females install more pink apps. A few apps labeled with ‘△’ are in red, such as the apps of Jingdong-electronics and Tmall. For the apps in System tool shown in Fig. 4(d), males install many more apps in different shades of blue than females, and there are almost no pink or red apps installed by either gender (perhaps because fewer exist in those colors).

V. The gender predictive ability

In this section, we first extracted features from the snapshot of installed apps based on the gender differences mentioned above, and then investigated the ability of individual features and combinations of different features to predict gender.

A. Features extracted from snapshot of apps installed

*1) **App-based user representation**:* We intuitively exploited the installed apps to represent users for gender prediction. In detail, we took each app as a dimension and represent each user as an app-based vector. If an app is installed, the corresponding value of its dimension is set to 1, and if not, the value is 0. Formally, a user u is represented by $u = (a_1, a_2, a_3, \cdots, a_k, \cdots, a_m)$, where a_k is for the kth app, and it has two values, 1 and 0, for indicating whether the app is installed. Considering the sparsity and redundancy, we used IG to select the discriminative apps to compactly represent each user, so as to improve computation efficiency.

*2) **Category-based user representation**:* Based on the differences in 29 app categories between females and males, we converted the differences to features for user representation. More specifically, each user is represented as a vector of 29 dimensions, $u = (c_1, c_2, \cdots, c_{29})$, where c_k is for the kth category, indicating the number of the apps in the category installed by the user.

*3) **Topic-based user representation**:* Considering the differences in app functions conveyed by topics that were learned from app descriptions, we applied the n topics to represent each user as a vector. One user can be represented by $u = (d_1, d_2, \cdots, d_k, \cdots, d_n)$, where d_k is the kth topic, and the value is the probability the topic belonging to the user.

*4) **Tag-based user representation**:* Considering there are differences in word clouds generated by the tags of apps in the same category, we built a user representation vector from tags. There were 19,056 tags for about 3,700 apps, and 1965 unique tags, *i.e.*, 1965 words in Chinese. Each user was firstly represented as a tag-based vector using the tags, and the value of each dimension is the TF-IDF of the tag. Then, we applied the pre-trained Chinese word vectors [3], where each word is represented as a vector with 300 dimensions [28]. A user representation vector with 300 dimensions was obtained by multiplying the tag based vector of each user with the pre-trained word vectors.

*5) **Icon-based user representation**:* In order to extract features from an app icon color, we set 64 color bins, which was obtained by evenly dividing each color value (R, G, and B) into four pieces. Each bin represents a color range. The dominant color of each app icon was computed [29], and the app was placed in a color bin if its dominant icon color was in the corresponding color range of the bin. Here, we focused on the icons of the top 1000 apps selected via IG. For each user, his/her app icon falls into one of the 64 color bins. We represent each user as a vector with 64 dimensions, $u = (o_1, o_2, \cdots, o_{64})$, where o_i is the number of the apps in the corresponding color bin.

B. Implementation and performance metrics

We trained different classifiers to infer gender, including SVM, LR (Logistic Regression) , GBDT (Gradient Boosting Decision Tree), and DNN (Deep Neural Network). In particular, in the DNN model, features were input into a wide layer, followed by three hidden layers of fully connected Rectified Linear Units (ReLU). There were 64, 16, and 4 neurons on the first, second, and third hidden layer, respectively. In the training procedure, a cross-entropy loss was minimized with gradient decent on the output of the sampled softmax, and the optimizer we used was RMSprop. Learning rate was set as 0.001. We applied the early stopping policy and $patience$ was set to 2 for training the network. The activation function we selected was sigmoid. The number of epochs was set as 10, and the batch size was set as 128.

We employed a five-fold cross-validation policy. The dataset was randomly divided into five folds as evenly as possible. In each round, four folds were used for training classifiers and the rest for validation. Thus, any user for testing will never simultaneously appear in the training set and testing set. We repeated the procedure five times and report the averages of the tests below.

We used three criteria to measure the performance of the classification: *ACC* (accuracy), *AUC* (area under the curve), and *F1 score*.

C. Results

*1) **Gender prediction performance.**:* We investigated the gender predictive ability of individual features and combinations of different features using different classifiers. Table I summarizes the performance of the classifiers over the samples. As we can see from Table I, the best performance for predicting gender is using the DNN model with the combination of apps and topics, with an ACC of 76.62%, AUC of 84.23%, and F1 score of 0.76, while the icon-based feature performs the worst (compared to all the feature sets), *e.g.*, with an ACC of 69.27%, AUC of 72.86%, and F1 score of 0.72 for an LR model. The performance of our best model outperformed the existing work [19] by about 5% in accuracy (76.62% *vs.* 71.5% (the original result reported in [19]), and 10% in AUC (84.31% *vs.* 74.0%). The model of DNN with 3 hidden layers performs the best for 8 out of the 16 feature sets. The neural network is helpful for feature fusion for inferring gender over most features.

[3]https://github.com/Embedding/Chinese-Word-Vectors

TABLE I: Gender prediction results.

Feature	Model	ACC	AUC	F1	Feature	Model	ACC	AUC	F1
App (1000)	SVM	0.7427	0.8202	0.7364	App+Icon (1064)	SVM	0.7419	0.8184	0.7383
	LR	0.7450	0.8213	0.7427		LR	0.7457	0.8212	0.7437
	GBDT	0.7470	0.8176	0.7341		GBDT	0.7421	0.8118	0.7323
	DNN	**0.7655**	**0.8424**	**0.7654**		DNN	0.7659	0.8414	0.7726
Topic (300)	SVM	0.7355	0.7969	0.7456	Topic+ Category (329)	SVM	0.7353	0.8039	0.7420
	LR	0.7306	0.7909	0.7392		LR	0.7357	0.7984	0.7404
	GBDT	0.7399	0.8034	0.7394		GBDT	0.7429	0.8072	0.7393
	DNN	0.7446	0.8122	0.7445		DNN	0.7411	0.8027	0.7338
Category (29)	SVM	0.6919	0.7580	0.6910	Topic+Tag (600)	SVM	0.7378	0.7963	0.7467
	LR	0.6907	0.7545	0.6828		LR	0.7361	0.7984	0.7404
	GBDT	0.6977	0.7594	0.6894		GBDT	0.7315	0.7945	0.7226
	DNN	0.6987	0.7432	0.6977		DNN	0.7187	0.7890	0.6994
Tag (300)	SVM	0.7240	0.7908	0.7231	Topic+Icon (364)	SVM	0.7378	0.7963	0.7467
	LR	0.7267	0.7965	0.7269		LR	0.7347	0.7965	0.7396
	GBDT	0.6881	0.7523	0.6812		GBDT	0.7443	0.8068	0.7362
	DNN	0.6987	0.7432	0.6977		DNN	0.7422	0.7992	0.7399
Icon (64)	SVM	0.6885	0.7412	0.6984	Category+ Tag (329)	SVM	0.7297	0.7984	0.7289
	LR	0.6927	0.7286	0.6912		LR	0.7318	0.8009	0.7305
	GBDT	0.6813	0.7470	0.6748		GBDT	0.7121	0.7730	0.7036
	DNN	0.6865	0.7357	0.6812		DNN	0.6901	0.7595	0.6923
App+Topic (1300)	SVM	0.7437	0.8194	0.7421	Category+ Icon (93)	SVM	0.7147	0.7762	0.7196
	LR	0.7453	0.8215	0.7438		LR	0.7096	0.7680	0.7098
	GBDT	0.7505	0.8181	0.7453		GBDT	0.7183	0.7820	0.7107
	DNN	**0.7662**	**0.8423**	**0.7661**		DNN	0.7081	0.7717	0.7211
App+ Category (1029)	SVM	0.7419	0.8174	0.7393	Tag+Icon (364)	SVM	0.7292	0.7968	0.7318
	LR	0.7446	0.8215	0.7419		LR	0.7290	0.7972	0.7304
	GBDT	0.7441	0.8126	0.7362		GBDT	0.7086	0.7734	0.7004
	DNN	0.7642	0.8424	0.7702		DNN	0.6964	0.7645	0.7056
App+Tag (1300)	SVM	0.7444	0.8181	0.7415	All (1693)	SVM	0.7445	0.8182	0.7437
	LR	0.7466	0.8217	0.7442		LR	0.7465	0.8217	0.7450
	GBDT	0.7323	0.7959	0.7186		GBDT	0.7423	0.8081	0.7335
	DNN	0.7630	0.8357	0.7696		DNN	**0.7655**	**0.8431**	**0.7653**

When we solely used one type of feature, the ACC for apps, topics, tags, categories, and icons is 76.55%, 74.46%, 72.67%, 69.87%, and 69.27%, respectively. It indicates that discriminative apps selected by IG and topics learned from app description by LDA are the two most useful features for distinguishing gender, while the category-based and icon-based features are the worst. For the app-based user representation method, we selected the top 1000 significant apps that are the most powerful by applying the IG method. The app-based features are with binary values and whether each of them was installed or not by one user helps to identify the gender of the user. For the topics extracted from app description, the topic-based features construct a fine-grained representation of users, where each topic reflects a specific function and the probability one user belongs to the topic indicates the user preferences to the topic. Compared with topic-based features, category-based features are a relatively coarse grained representation method. There are only 29 dimensions in total, which is very small compared to 300 and 1000. The value of each dimension is the number of installed apps in the corresponding category, and there are few differences in the values among users. Thus, the category-based features are not so powerful to infer gender. Similarly, icon-based representation method is a relatively coarse grained way, resulting in the difficulty in powerfully identifying the gender of each user.

Combining different features do not provide a significant performance improvement. In particular, combining all five features achieves an ACC of 76.55%. However, we can achieve the same performance when only using discriminative apps. In addition, the combination of the two feature types of apps and topics (ACC: 76.62%), and the combination of apps and icons (ACC: 76.59%) outperformed the combination of all five feature types (by a small margin). In addition, the combination of some features even degraded the performance. For example, solely using topics resulted in an ACC of 74.46%, but, the combination of topics and tags resulted in an ACC of 73.78%.

TABLE II: The 10 most discriminative apps for gender.

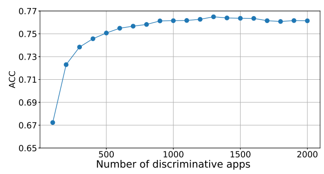
No.	Name	Apps	IG	# Male	# Female
1		Mogujie-BeautyShopping	0.01978	202	**973**
2		Meitu-Photography	0.01904	2002	**3401**
3		Meilishuo-BeautyShopping	0.01495	142	**721**
4		AutoHome	0.01423	**783**	183
5		Meiyou-PeriodTracker	0.01319	144	**677**
6		Meiyan-BeautyCam	0.01259	655	**1477**
7		360-BeautyCam	0.01162	900	**1768**
8		Meipai-BeautyMV	0.01102	821	**1639**
9		Kuaibo-Video	0.01075	**1421**	667
10		AdobeFlashPlayer	0.01040	**2184**	1277

The most powerful features are the app- and topic-based ones, while the category- and icon-based ones are the best.

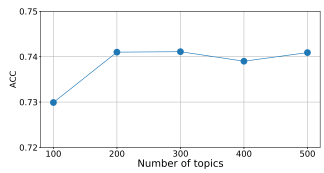
*2) **Discriminative apps selected by IG:*** In this experiment, we demonstrated the discriminative apps selected by IG which are the most powerful for distinguishing gender.The top 10 most discriminative apps were listed, shown in Table II. The most discriminative app is Mogujie, an app for online fashion shopping targeting women, which 973 females and only 202 males install. The second most discriminative app is Meitu, an app for photography, and the third most discriminative app is Meilishuo, an app for beauty shopping similar to Mogujie. The two apps attract many more females than males. The fourth most discriminative app is AutoHome, an app related to automotive content, that has many more males installing it than females (783 *vs.* 183). Among the top 10 apps, 7 apps have more female installers and 3 have more males. The apps attractive to female users are related to beauty shopping, photography, period tracker, and beautifying videos, and the apps attractive to males are about cars and videos.

*3) **Performance with varying apps and topics for user representation :*** As shown in Table I, the binary app-based vectors describing whether one app is installed or not have the most powerful predictive ability for gender, with an accuracy of 76.55%. The topic-based user representation is the second most powerful for inferring gender, with an accuracy of 74.46%. We investigated the impact on performance of varying the number of apps and topics for user representation. We used different numbers of discriminative apps to represent each user. We experimented with 100, 200, 300, ..., 1900, 2000 discriminative apps with an interval of 100 for user representation. The performance of ACC is shown in Fig. 6(a). As we can see, when the number of discriminative apps increases from 100 to 500, the performance dramatically increases. The performance eventually becomes steady with nearly 1000 apps, with an accuracy of 76.55%. Compared to the top 100 discriminative apps, the top 1000 discriminative apps produced a much better performance for inferring gender.

We also investigated the effect of varying the number of topics for inferring gender. We experimented with 100, 200, 300, 400, and 500 topics. The performance of ACC is shown in Fig. 6(b). As we can see, the performance grows quickly when the number of topics varies from 100 to 200, and changes slightly after 200. When the number of topics is 300, the performance is the best (74.46%), while the performance is the worst when it is 100. That is to say, using only 100

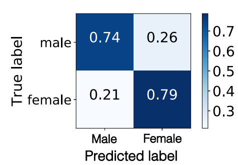


(a)



(b)

Fig. 6: Performance with varying number of (a) discriminative apps and (b) topics.



(a)

	Apps	Proportion of the users in each predicted group installing the app			
		Male→ Female	Male→ Male	Female→ Male	Female→ Female
1	Mogujie-BeautyShopping	8.99%	0.50%	1.00%	16.21%
2	Meitu-Photography	37.67%	22.88%	19.42%	52.35%
3	Meilishuo-BeautyShopping	6.62%	0.25%	0.50%	12.08%
4	AutoHome-Car	1.60%	13.51%	7.21%	1.15%
5	Meiyou-PeriodTracker	5.63%	0.63%	1.25%	11.13%
6	Meiyan-BeautyCam	16.28%	6.11%	5.95%	23.41%
7	360-BeautyCam	18.04%	9.90%	9.77%	27.30%
8	Meipai-BeautyMV	16.69%	8.95%	6.83%	25.91%
9	Kuaibo-Video	5.94%	23.47%	16.10%	6.94%
10	Adobe FlashPlayer	12.45%	34.91%	21.67%	15.77%

(b)

Fig. 7: (a) Confusion matrix. (b) The proportion of the users in each predicted group installing the significant apps selected by IG. Apps in red (female apps) are installed by more females than males in the dataset, and the apps in blue (male apps) are installed by more males than females.

interests learned by LDA is not enough for optmizing gender inferring. Using 300 topics for user representation is advisable for inferring gender.

VI. Error analysis

Our DNN model performs well for predicting gender, but not perfectly. To understand its limitations, we analyzed the falsely predicted samples output by the DNN model with users being represented by the selected 1000 apps, as shown in Fig. 7. The 1000 apps are the most powerful ones for distinguishing gender selected by the method of IG (details can be seen in Section IV-C). Fig. 7(a) shows the confusion matrix of the results, where 26% of males are misclassified as females and 21% of females are misclassified as males. In order to further explore, we analyzed the apps installed by the true samples and false samples. We divided the user samples into four user groups: male→female, male→male, female→male, and female→female, where → means the user samples on the left that were classified as the user samples on the right, for example, male→female refers to the set of male users that were classified as females.

For the error analysis, we listed the top 10 discriminative apps selected by IG, and calculated the proportion of the users in each group installing the app, as shown in Fig. 7(b). Apps whose names are in red (female apps) are installed by more females than males in the dataset, and the apps in blue (male apps) are installed by more males than females. As we can see, the users that were classified as females install more female apps than the users inferred as males, and install fewer male apps:

- The proportion of the users in the group of male→female installing the female apps of Mogujie-BeautyShopping, Meitu-Photography, Meilishuo-BeautyShopping, Meiyou-PeriodTracker, Meiyan-BeautyCam, 360-BeautyCam, and meipai-BeautyMV, is much bigger (8.99%) than that of the users in the group of male→male (0.50%). In contrast, the proportion of the users in the group of male→female installing the male apps of AutoHome-Car, Kuaibo-Video, and AdobeFlashPlayer is much smaller than that of the users in the group of male→male.
- For the users in the group of female→male (*i.e.*, females classified as males), the proportion of them with female apps, such as Mogujie-BeautyShopping and Meitu-Photography is much smaller than that of the users in the group of female→female (*i.e.*, females correctly classified as females). In contrast, the proportion of the users in the group of female→male installing male apps is much larger than that of the users in the group of female→female. For example, 7.21% of the users in female→male install the app of AutoHome-Car, but only 1.15% of the users in female→female install the app.
- For the female apps, the proportion of the users in both groups of male→female and female→female is much larger than that of the users in the groups of male→male and female→male. But for the male apps, we see the opposite result, as expected.

We looked into the list of apps installed by a particular randomly chosen male (ID: 6787) who was misclassified as a female. He installed 31 apps in total, and there were 5 apps which have very significant features related to females: two apps for tracking menstruation (Dayima and Meiyou), Meitu for photography, two apps of Meilishuo and Jumeiyoupin for beauty shopping. In particular, the apps of Dayima and Meiyou are designed for females to track their period. Meitu is installed by more females in the dataset, and has a big information gain for gender. Meilishuo and Jumeiyoupin are social commerce fashion apps targeting young ladies. If we speculate about the reason for these app installations, it could be that User 6787 installed these apps by his girlfriend, wife or another significant female in his life. Or maybe these apps were installed but rarely or never used. It is difficult for our model to correctly identify the gender of such users. Perhaps having other knowledge could help, such as installation time, usage over time and uninstallation operations, which can reflect one's priority or preferences to apps. A follow-up interview with users we had the most difficulty identifying is required to verify the possible reasons for app installation.

VII. Discussion

Through our analysis, we have discovered gender differences in terms of popularity, functions, and icons of apps installed. We have demonstrated that we can accurately predict gender from a single snapshot of apps installed on a smartphone. In particular, we achieved an accuracy of approximately 76%, AUC of 84%, and F1 score of 0.76, outperforming the existing work by about 5% in accuracy (76.62% *vs.* 71.5%), and 10% in AUC (84.31% *vs.* 74.0%).

We compared our results with two studies conducted on app usage records of two relatively large-scale populations, and our performance is nearly on par with their results. The two studies used one-month and four-month long datasets, whereas we used only a single snapshot of smartphone apps installed. Malmi *et al.* studied the lists of apps used at once within one month, and inferred 3760 Android users' gender using logistic regression with an accuracy of 82% [22]. Qin *et al.* analyzed the app usage logs over four months, and inferred about 32,000 users' gender with a F1 score of 0.80 and accuracy of 81% [21]. Compared with the two studies, our model had only about a 5% drop in accuracy, and similar F1 score as Qin's study. Compared to a single snapshot of apps installed on a smartphone, monitoring app usage on a smartphone can be a resource-intensive task, which is more expensive and time-consuming. Moreover, monitoring and recording historical app usage records is intrusive, while the lists of apps installed can be obtained without requiring users' permission. The comparison shows that the single snapshot of smartphone apps installed provides comparable performance to app usage records that require tracking app usage over a period of time.

A. Practical implications

*1) **Gender profiling to address cold-start problem**:* The approach we described in this paper builds a gender profile from the snapshot of apps installed on a user's smartphone, and its performance compares favorably to user tracking techniques that can be expensive, time-consuming, and intrusive. A user's gender can instead be quickly determined by a single access of his/her list of installed apps. Gender profiling with our method could potentially be used in conjunction with user tracking technologies to address the cold-start problem in tracking-based systems. The cold-start problem can be characterized by an inference system that has difficulty making inferences due to a lack of a sufficient amount of data. By using gender to base decisions on, especially at the beginning of data collection, systems can provide a potentially better user experience until a sufficient amount of data has been collected over a period of time for training more sophisticated models.

*2) **Improving user experience of smartphone usage**:* We foresee many opportunities of gender profiling for improving the user experience of smartphone usage. Mobile phone designers, mobile carriers, and application developers can improve services and devices based on the gender differences in installed app lists. In particular, mobile phone designers can design smartphones that are targeted towards improving the user experience of females or males by providing features that females (or males) may value more than males (or females). For example,females may value an improved camera sensor since they have greater interest in taking selfie. Mobile carriers could allow for the customization of what apps are made available for females and males based on the most discriminative apps and the top interests. For example, mobile carriers can pre-install apps related to automotive cars for male users. Application developers can recommend apps to females and males according to their interests, and provide personalized applications. After inferring gender from the installed app lists, app developers can recommend specific apps that males or females are the most interested in (*e.g.*, by using a list similar to the top discriminative apps as we discuss earlier).

B. Theoretical Implications

We believe our findings provide a support for existing theories. For example, beauty-related apps attract more female users than males. Females pay more attention to their appearance in the sociocultural climate surrounding them where how they look like are capable of representing them [30]. It is the association between women's worth and their physical appearance that explains their preferences for apps that can improve their looks and feelings. Users of car apps are mostly male, reflecting that men's relationships with cars are premised on passion and pleasure [31]. The finding that females have more apps in pink and males installing more apps in different shades of blue, is consistent with Cohen's conclusion that females prefer pink more than males do and males preferring blue more than females do [32]. It could also reflect, however, that app developers tend to use female-oriented colors (*e.g.*, pink) for the icons of apps developed primarily for females (and *vice-versa* for males).

C. Limitations

Although we showed the interest differences in installed app lists between females and males, we must acknowledge the limitations of the dataset used in this study. First, most users in our dataset were from China, and the conclusions we obtained were mainly for a Chinese population. Although females in the world have similarities, as do males, whether our conclusions can be generalized to user populations from other countries is still unknown. It requires further investigation into the gender differences in installed app lists for other user populations. Second, we investigated gender differences from a snapshot of installed app lists, without information on how frequently an app is used. Some apps installed may rarely be used, whereas others may be frequently used. It would be interesting to explore gender differences through app usage behaviors as part of future work.

Another important limitation is that the labels provided for users to select from to indicate their gender was limited to female and male. In some cultures, it is becoming more common to provide non-binary options for individuals to indicate their gender. Such options would increase the complexity and challenge of gender profiling.

Finally, while there are clearly some benefits of being able to identify gender from smartphone data, there are also some risks. Having a system that can identify whether a phone user is male or female could be used to target individuals. As well, while our approach works because of what may appear to be stereotypes (or adherence to normative structural gender [33]) about males and females (*e.g.*, males like cars and females like pink), it works because the stereotypes to some degree hold true for the individuals in our dataset. As these stereotypes become less representative of actual gendered differences, there is still the risk that app developers may 'enforce' these stereotypes through the designs of their apps, both in the icons and text used to describe them, and in how they are marketed. It will require further study to better understand these risks.

VIII. Conclusion and future work

In this work, we have presented an empirical study of gender profiling from a snapshot of installed smartphone apps. We conducted our study based on a large-scale dataset of installed app lists from 15,000 Android users. We demonstrated the gender differences in terms of popularity, functions, and icons of apps installed. Then, we extracted features from a snapshot of apps installed on smartphones to infer the gender of each user. Sequentially, we investigated the gender predictive ability of individual features and combinations of different features. The combination of apps selected by Information Gain and the topics learned from app descriptions using LDA provides the best gender classification results, outperforming the existing work by about 5% in accuracy (76.62% *vs.* 71.5%), and about 10% in AUC (84.23% *vs.* 74.0%). This performance from a one-time snapshot is nearly on par with the state-of-the-art gender profiling results based on months of collected app usage data. Finally, we performed an error analysis on our misclassified users and discussed the implications of our findings and the limitations of this work.

To address the limitations in correctly inferring gender from a subset of our population, we plan to conduct follow-up interviews to understand the choice for app installations. In addition, our method can potentially be readily extended to a series of other user attributes, such as age, income level, and occupation. In our future work, we will explore the differences among other user attributes from a snapshot of apps installed on a smartphone. Further, we will explore the risks associated with identifying these types of attributes from smartphone data.

IX. Acknowledgment

This work was supported by National Key R&D Program of Chin (2018YFC1504006), NSF of China (No. 61772460, 61802340, and 61802342), China Postdoctoral Science Foundation under Grant No. 2017M620246 and 2018T110591. Dr. Gang Pan is the corresponding author.

References

[1] F. Ahmed and M. H. Kabir, "Facial feature representation with directional ternary pattern (dtp): Application to gender classification," in *IRI'12*. IEEE, 2012, pp. 159–164.

[2] J. A. Howard and J. A. Hollander, *Gendered situations, gendered selves: A gender lens on social psychology.* Rowman & Littlefield, 1997, vol. 2.

[3] H.-C. Kim, D. Kim, Z. Ghahramani, and S. Y. Bang, "Appearance-based gender classification with gaussian processes," *Pattern Recognition Letters*, vol. 27, no. 6, pp. 618–626, 2006.

[4] J. Wang, Y. Wang, D. Zhang, J. Goncalves, D. Ferreira, A. Visuri, and S. Ma, "Learning-assisted optimization in mobile crowd sensing: A survey," *TII*, vol. 15, no. 1, pp. 15–22, 2019.

[5] M. C. Grace, W. Zhou, X. Jiang, and A.-R. Sadeghi, "Unsafe exposure analysis of mobile in-app advertisements," in *WISEC'12*. ACM, 2012, pp. 101–112.

[6] S. G. Mazman and Y. K. Usluel, "Gender differences in using social networks," *TOJET*, vol. 10, no. 2, 2011.

[7] N. L. Muscanell and R. E. Guadagno, "Make new friends or keep the old: Gender and personality differences in social networking use," *Computers in Human Behavior*, vol. 28, no. 1, pp. 107–112, 2012.

[8] V. Kisekka, S. Bagchi-Sen, and H. R. Rao, "Extent of private information disclosure on online social networks: An exploration of facebook mobile phone users," *Computers in human behavior*, vol. 29, no. 6, pp. 2722–2729, 2013.

[9] X. Lu, W. Ai, X. Liu, Q. Li, N. Wang, G. Huang, and Q. Mei, "Learning from the ubiquitous language: an empirical analysis of emoji usage of smartphone users," in *UbiComp'16*. ACM, 2016, pp. 770–780.

[10] M. Al-Qurishi, M. S. Hossain, M. Alrubaian, S. M. M. Rahman, and A. Alamri, "Leveraging analysis of user behavior to identify malicious activities in large-scale social networks," *TII*, vol. 14, no. 2, pp. 799–813, 2018.

[11] M. Zhou, "Gender difference in web search perceptions and behavior: Does it vary by task performance?" *Computers & Education*, vol. 78, pp. 174–184, 2014.

[12] M. Roy and M. T. Chi, "Gender differences in patterns of searching the web," *Journal of Educational Computing Research*, vol. 29, no. 3, pp. 335–348, 2003.

[13] F. Lamberti, G. Paravati, V. Gatteschi, and A. Cannavo, "Supporting web analytics by aggregating user interaction data from heterogeneous devices using viewport-dom-based heat maps," *TII*, vol. 13, no. 4, pp. 1989–1999, 2017.

[14] I. Andone, K. Błaszkiewicz, M. Eibes, B. Trendafilov, C. Montag, and A. Markowetz, "How age and gender affect smartphone usage," in *UbiComp'16: Posters*, 2016, pp. 9–12.

[15] D. Rao, D. Yarowsky, A. Shreevats, and M. Gupta, "Classifying latent user attributes in twitter," in *SMUC'10*, 2010, pp. 37–44.

[16] B. Bi, M. Shokouhi, M. Kosinski, and T. Graepel, "Inferring the demographics of search users: Social data meets search queries," in *WWW'13*, 2013, pp. 131–140.

[17] Y. Dong, Y. Yang, J. Tang, Y. Yang, and N. V. Chawla, "Inferring user demographics and social strategies in mobile social networks," in *SIGKDD'14*. ACM, 2014, pp. 15–24.

[18] S. Brdar, D. Culibrk, and V. Crnojevic, "Demographic attributes prediction on the real-world mobile data," in *2012 Mobile Data Challenge by Nokia Workshop*, 2012.

[19] S. Seneviratne, A. Seneviratne, P. Mohapatra, and A. Mahanti, "Your installed apps reveal your gender and more!" *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 18, no. 3, pp. 55–61, 2015.

[20] S. Zhao, S. Li, J. Ramos, Z. Luo, Z. Jiang, A. K. Dey, and G. Pan, "User profiling from their use of smartphone applications: A survey," *Pervasive and Mobile Computing*, p. 101052, 2019.

[21] Z. Qin, Y. Wang, H. Cheng, Y. Zhou, Z. Sheng, and V. Leung, "Demographic information prediction: A portrait of smartphone application users," *TETC*, 2016.

[22] E. Malmi and I. Weber, "You are what apps you use: Demographic prediction based on user's apps," *arXiv preprint arXiv:1603.00059*, 2016.

[23] Y. Wang, Y. Tang, J. Ma, and Z. Qin, "Gender prediction based on data streams of smartphone applications," in *BIGCOM'15*. Springer, 2015, pp. 115–125.

[24] K. Mo, B. Tan, E. Zhong, and Q. Yang, "Report of task 3: Your phone understands you," in *2012 Mobile Data Challenge by Nokia Workshop*, 2012.

[25] S. Zhao, J. Ramos, J. Tao, Z. Jiang, S. Li, Z. Wu, G. Pan, and A. K. Dey, "Discovering different kinds of smartphone users through their application usage behaviors," in *UbiComp'16*. ACM, 2016, pp. 498–509.

[26] R. A. Fisher, "Statistical methods for research workers," in *Breakthroughs in Statistics*. Springer, 1992, pp. 66–70.

[27] T. M. Mitchell *et al.*, "Machine learning." McGraw-Hill Boston, MA:, 1997.

[28] S. Li, Z. Zhao, R. Hu, W. Li, T. Liu, and X. Du, "Analogical reasoning on chinese morphological and semantic relations," in *ACL'18 (Volume 2: Short Papers)*. Association for Computational Linguistics, 2018, pp. 138–143.

[29] A. Miniukovich and A. De Angeli, "Pick me!: Getting noticed on google play," in *CHI'16*. ACM, 2016, pp. 4622–4633.

[30] N. A. Heflick, J. L. Goldenberg, D. P. Cooper, and E. Puvia, "From women to objects: Appearance focus, target gender, and perceptions of warmth, morality and competence," *Journal of Experimental Social Psychology*, vol. 47, no. 3, pp. 572–581, 2011.

[31] C. Landström, "A gendered economy of pleasure: Representions of cars and humans in motoring magazines," *Science & Technology Studies*, 2006.

[32] P. N. Cohen, "Children's gender and parents' color preferences," *Archives of Sexual Behavior*, vol. 42, no. 3, pp. 393–397, 2013.

[33] S. G. Harding, *The science question in feminism*. Cornell University Press, 1986.

Sha Zhao is currently a Postdoctoral Research Fellow of the College of Computer Science and Technology, Zhejiang University. She received the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2017. Zhao visited the Human-Computer Interaction Institute at Carnegie Mellon University as a visiting PhD student from September 2015 to September 2016. Her research interests include pervasive computing, mobile sensing, data mining, and machine learning. She won the Best Paper Award of ACM UbiComp'16.

Yizhi Xu is currently a Master candidate of the College of Computer Science and Technology, Zhejiang University, Hangzhou, China. He received his B. Sc. degree in Electrical Engineering from Zhejiang University, Hangzhou, China, in 2017. His research interests include machine learning and data mining.

Xiaojuan Ma is an assistant professor of Human-Computer Interaction (HCI) at the Department of Computer Science and Engineering (CSE), Hong Kong University of Science and Technology (HKUST). She received the Ph.D. degree in Computer Science at Princeton University. She was a post-doctoral researcher at the Human-Computer Interaction Institute (HCII) of Carnegie Mellon University (CMU), and before that a research fellow in the National University of Singapore (NUS) in the Information Systems department. Before joining HKUST, she was a researcher of Human-Computer Interaction at Noah's Ark Lab, Huawei Tech. Investment Co., Ltd. in Hong Kong. Her background is in Human-Computer Interaction. She is particularly interested in data-driven human-engaged AI and Human-Robot Interaction in the domain of education, health, and design.

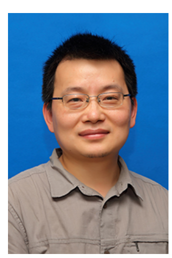
Ziwen Jiang received his B. Sc. degree in Software Engineering from Zhejiang University, Hangzhou, China, in 2015 and Master degree in Computer Science from Zhejiang University, Hangzhou, China, in 2018. He is a staff of Alibaba, Hangzhou, China. His research interests include machine learning and data mining.

Zhiling Luo was an Assistant Professor in Computer Science at Zhejiang University, China. He received his B.S. and Ph.D. degree in Computer Science from Zhejiang University in 2012 and 2017, respectively. He was the visiting scholar of Georgia Institute of Technology, US, in 2016. His research interests include service computing, machine learning and data mining.

Shijian Li is currently with the College of Computer Science and Technology, Zhejiang University. He received the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2006.In 2010, he was a Visiting Scholar with the Institute Telecom SudParis, Evry, France. He was published over 40 papers. His research interests include sensor networks, ubiquitous computing, and social computing. Dr. Li serves as an Editor of the International Journal of Distributed Sensor Networks and as Reviewer or PC Member of more than ten conferences.

Laurence T. Yang received the B.E. degree in Computer Science and Technology and B.Sc degree from Applied Physics both from Tsinghua University, China and the PhD degree in Computer Science from University of Victoria, Canada. He is a professor and W.F. James Research Chair in Department of Computer Science, St. Francis Xavier University, Canada. His research interests include parallel and distributed computing, embedded and ubiquitous/pervasive computing, and big data. His research has been supported by the National Sciences and Engineering Research Council, and the Canada Foundation for Innovation.

Anind Dey is a Professor and Dean of the Information School and Adjunct Professor in the Department of Human-Centered Design and Engineering, University of Washington. Anind is renowned for his early work in context-aware computing. His research is at the intersection of human-computer interaction, machine learning, and ubiquitous computing. Anind has focused on passively collecting large amounts of data about how people interact with their phones and the objects around them, to use for producing detection and classification models for human behaviors of interest. He applies a human-centered and problem-based approach through a collaboration with an amazing collection of domain experts in areas of substance abuse (alcohol, marijuana, opioids), mental health, driving and transportation needs, smart spaces, sustainability, and education. Anind was inducted into the ACM SIGCHI Academy for his significant contributions to the field of human-computer interaction in 2015.



Gang Pan received the B.Eng. and Ph.D. degrees from Zhejiang University, China, in 1998 and 2004, respectively. He is currently a professor of the Department of Computer Science, and deputy director of State Key Lab of CAD&CG, Zhejiang University, China. From 2007 to 2008, he was a visiting scholar at the University of California, Los Angeles. His current interests include artificial intelligence, pervasive computing, brain-inspired computing, and brain-machine interfaces. He received three best paper awards (e.g. ACM UbiComp'16) and three nominations from premier international conferences. He is the recipient of IEEE TCSC Award for Excellence (Middle Career Researcher), CCF-IEEE CS Young Computer Scientist Award, and the State Scientific and Technological Progress Award. He serves as an Associate Editor of IEEE Trans. Neural Networks and Learning Systems, IEEE Systems Journal, Pervasive and Mobile Computing, and ACM Proceedings of Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT).