Information Systems 71 (2017) 152–163

Contents lists available at ScienceDirect

Information Systems

journal homepage: www.elsevier.com/locate/is

Mobile recommendations based on interest prediction from consumer's installed apps–insights from a large-scale field study

Remo Manuel Frey[a,*], Runhua Xu[a], Christian Ammendola[b], Omar Moling[b], Giuseppe Giglio[c], Alexander Ilic[a]

[a] ETH Zurich, Weinbergstrasse 56/58, 8092 Zurich, Switzerland
[b] 42matters AG, Sumatrastrasse 3, 8006 Zurich, Switzerland
[c] Swisscom AG, Binzring 17, 8045 Zurich, Switzerland

ARTICLE INFO

Article history:
Received 14 July 2017
Revised 12 August 2017
Accepted 15 August 2017
Available online 17 August 2017

Keywords:
Mobile recommender system
User profiling
Personalization
M-commerce
Click-through rate

ABSTRACT

Recommender systems are essential in mobile commerce to benefit both companies and individuals by offering highly personalized products and services. One key pre-requirement of applying such systems is to gain decent knowledge about each individual consumer through user profiling. However, most existing profiling approaches on mobile suffer problems such as non-real-time, intrusive, cold-start, and non-scalable, which prevents them from being adopted in reality. To tackle the problems, this work developed real-time machine-learning models to predict user profiles of smartphone users from openly accessible data, i.e. app installation logs. Results from a study with 904 participants showed that the models are able to predict interests on average 48.81% better than a random guess in terms of precision and 13.80% better in terms of recall. Since the effectiveness of such predictive models is unknown in practice, the predictive models were evaluated in a large-scale field experiment with 73,244 participants. Results showed that by leveraging our models, personalized mobile recommendations can be enabled and the corresponding click-through-rate can be improved by up to 228.30%. Supplementary information, study data, and software can be found at https://www.autoidlabs.ch/mobile-analytics.

1. Introduction

According to Statista [1], the number of smartphone users was estimated at 1.6 billion in 2014 and will grow to 2.7 billion in 2019. Due to their ubiquitous and highly personalized nature, smartphones are nowadays an essential companion of our daily lives. On average, people spend 177 minutes of attention on smartphones and tablets each day, while television gets about 168 minutes of attention each day [2]. Consequently, smartphones are overtaking televisions and desktop computers to become the top channel for companies to conduct direct marketing. In 2015, 50 billion U.S. dollars were spent on mobile advertisements, and the forecast for 2017 expects an increase up to 95 billion U.S. dollars [3].

Through traditional media like televisions and newspaper, companies apply mass marketing to promote products and services. Although a large number of potential consumers can be reached, this type of marketing is less effective due to its high cost and low conversion rate. On the consumer side, problems such as receiving irrelevant promotions and facing too much advertisement occur consequently. To tackle the problem, recommender systems have been developed and widely adopted in online marketing activities. On the one hand, these systems help companies identify high potential consumers for different products thereby enabling targeted marketing at an individual level. On the other hand, they assist consumers during their shopping process and support them with purchase decisions.

Previous research has demonstrated the power of these recommender systems [4–9]. However, to deploy an effective online recommender system, a pre-requirement is to gain knowledge about each individual consumer, such as her demographics and previous purchases. Existing technologies typically combine Web data analytics with cookies to gain the knowledge based on a consumer's click events in her browser. Nevertheless, this approach is difficult to apply to smartphones. First, mobile users spend 90% of their time on using native applications. Second, cookies on mobile devices are still not throughout supported or often disabled by default. Consequently, existing cookie-based data analytical technologies are only limitedly applicable in the mobile context, which prevents companies from conducting effective personalized marketing.

* Corresponding author.
E-mail addresses: rfrey@ethz.ch (R.M. Frey), rxu@ethz.ch (R. Xu), christian@42matters.com (C. Ammendola), omar@42matters.com (O. Moling), giuseppe.giglio@swisscom.com (G. Giglio), ailic@ethz.ch (A. Ilic).

http://dx.doi.org/10.1016/j.is.2017.08.006
0306-4379/

Chaffey [2] found that the conversion rate on smartphones is less than half of that on desktop computers.

To gain knowledge about mobile consumers in a non-cookie setting, researchers and practitioners have started to develop new approaches for user profiling. As personal interests were shown to be useful in improving advertisements and recommendations [9–11], recent studies have started to predict personal interests based on analyzing an individual's social network content (e.g., posts, likes). However, such an approach cannot be applied by most companies because they do not have direct access to the social data on mobile. Thus, the focus of the present work is on the prediction of interests. We present an evaluation of a scalable machine-learning approach to predict personal interests and additionally gender from openly accessible information, i.e., mobile app installation events. Consequently, companies can integrate the models into their mobile apps, thus improving the performance of existing decision support systems and recommender systems.

Most research in the field of user profiling and recommender systems evaluates the developed models in terms of accuracy, precision, and recall. Although these metrics can reflect the goodness of a predictive model, lacking of research that shows the economic values of using the models in practice prevents companies from adopting them. To bridge the gap between new technologies and their business impact, a large-scale field experiment with 73,244 participants has been conducted together with our industry partner. The predictive approach was integrated into a mobile recommender system of our partner. Consequently, recommendations were personalized according to the predicted user profiles and corresponding business impact was measured quantitatively.

The contributions of this work are thus two-fold: First, an app-based approach for predicting interests and gender is evaluated in terms of precision and recall in a user study. To the best knowledge of the authors, the current approach has never been previously scientifically evaluated in terms of interests. Second, the predictive models were integrated into a mobile recommender system. For the first time, the effectiveness of such an app-based approach was measured in a real business case. This work thus provides insights for companies who intend to implement best practice quality from research in the real world.

The rest of the paper is structured as follows. Section 2 presents related work on recommender systems and user profiling approaches and Section 3 introduces the research design in detail. Section 4 and Section 5 describes the two empirical studies, respectively. Finally, the paper concludes with a discussion of the limitations and an outlook on future work.

2. Related work

2.1. Online recommender systems

Today's overload of products and services in online stores provokes a customer's need for decision support. Recommender systems can help to narrow down the considered product space, to compare similar products, and to come to the final purchase decision [12]. As a consequence, recommender systems are now an essential sales instrument in both brick–and–mortar stores and e-commerce [8]. There are two main techniques for recommender systems: content-based [13] and collaborative filtering [14]. Content-based Filtering tries to match preferences of individuals with properties of available shop items. For instance, a customer who likes fish is probably interested in a current salmon offer. In contrast, collaborative filtering tries to build a single recommendation using the knowledge about a bunch of customers. For instance, a company observed that people who like fish tend to buy rice as well. In this case, the company recommends rice to the fish buyer; even she has never bought rice in the past. Content-based Filtering and collaborative filtering have been extensively analyzed in previous research in the last two decades. Both techniques are typically combined in practice [15,16] in order to cover the advantages of both.

2.2. Importance of user profiling

However, the usage of the previously explained types of recommender systems requires information about users' preferences. Either users manually enter their preferences in an online form, or companies leverage data mining methods to gather automatically the information from previous user activities. The latter is less intrusive and less time-consuming for the user, because the data collecting process can be achieved in the background. Moreover, an ongoing automatic observation is able to track changes in user interests over time. Further, sophistic data mining methods are able to go beyond just matching demand and offer; they enable companies to better understand consumers in all their complexity and to tailor the offers to individual needs, behavior and habits. Weng and Liu [17] reported the potential for such personalized (or customized) product recommendations already in the early stages of e-commerce. "Ideally, profiles capture an aggregate view of the behavior of subsets of users based their interests or information needs" [18]. The objective of the personalization is thus to match each user's interests [19]. This approach is nowadays commonly referred as 'mass personalization'. Chen and Hsieh [4] identified six important design attributes for personalized promotions: interests, price, preference, promotion, brand, and type of mobile device. The present article considers two attributes: interests and gender. Both are popular in previous research about user profiling as shown later in this chapter. The benefits are measurable. In particular, one of the key findings of Zhang and Wedel [7] was that "customized promotions at all levels are more profitable than undifferentiated promotions in online stores." Thus, personalization has an essential business impacts and the generation of more detailed user profiles is a continuing trend. The downside contains several dimensions of privacy concerns, such as fraud and misuse of the personal data (internal and external), loss of control and collection in general [20]. Consumer weigh the potential benefits from personalized offers against the risks associated with a loss of privacy, today often referred as 'personalization privacy paradox' [21].

2.3. Prediction of interests and gender

2.3.1. Web-based approaches

When first recommender systems in e-commerce went online in the nineties, desktop computers were the preferred entry point for individual customers. A large number of research studies about user profiling on desktop computers were published, of which many tried to predict user's interests and demographics like gender and age. Research focused on analyzing users' visited websites used Nature Language Processing [22] for profiling. For instance, Kurtz and Mostafa [23] detected relevant topics and tracked interests in online news channels. Kim and Chan [24] built hierarchical clusters of keywords appearing in a user's visited websites, which was further used to build a topic model for predicting interests. Beside the textual content of browsed websites, Liu et al. [22] argued that the visual appearance of webpages helps to detect interests in a smaller granularity and hence with better accuracy than existing approaches. The measured accuracy of interest prediction was between 56.9% and 100%, depending on the analyzed webpage. The authors analyzed the layouts of popular websites and demonstrated with a simulation that vision-based personalization saved user's browsing time and improved user's browsing experience. Yang [25] captured users' web browsing behavior to generate unique user profiles. Jung [26] considered website bookmarking as the most important information to extract user preferences because each bookmarks reflects the intent of a user to revisit that website. The idea was to categorize bookmarks by using an ontology. The author used Bayesian networks to extract user interests. Similar approaches were evaluated for demographic estimation: Hu et al. [27] and Phuong and Phuong [28] predicted users' gender from their web browsing behaviors. Their proposed algorithms achieved a macro F1 score of 79.7% and 80.5%. Duong et al. [29] predicted gender based on catalog viewing data on e-commerce systems with a macro F1 score of 81.4%.

2.3.2. Social media based approaches

The dot-com bubble in 2001 marked a turning point for the Web. Afterwards, the Web was developed to 2.0 where users began to contribute their own content and leveraged network effects [30]. Virtual communities and platforms emerged in which users shared experiences, interests, ideas, and other personal information, today commonly referred to as social media content. A huge volume of social media content has been published and thus, easy to collect and to use for profiling. Chin and Wright [31] listed a number of potential content types for user profiling, such as posts (e.g. on Facebook, Twitter), blogs, forums, emails, and reviews. They concluded that each content type has different characteristics in terms of social conventions, change over time, number of spelling and grammatical errors. Therefore, each social media type “requires specialized processing” for user profiling. It is possible to combine different media types to improve the interest predictions [32]. Several researchers predicted personality traits based on social media content, such as Twitter [33], Facebook [34], and Emails [35]. Guo et al. [36] infer specific user interests from forum activities and Bhargava [37] et al. from Facebook posts with an accuracy between 71.7% and 90.2%. The prediction of gender is well-explored on social media as well. Tang et al. [38] achieved an accuracy of 96.3% with the analysis of the user names on Facebook. Deitrick et al. [39] predicted gender based on email content with accuracy between 88% and 95%. The evaluation of chats and tweets (short message on Twitter) reached an accuracy of 84.2% [40] and 99.3% [41]. However, the results could be over-optimistic because most of the above-mentioned studies had too small sample size, which easily leads to overfitting. Therefore, the present work will validate the machine-learning models with a large sample size. Also, the usage of social media content could provoke strong privacy concerns. Users are typically not aware of the possibilities and did not consciously agreed with the presented data proceeding methods.

2.3.3. Mobile-based approaches

With the increasing popularity of mobile phones in the past decade, mobile marketing brings a broad range of new opportunities for companies. Personalized interaction with end-consumers is possible in any place, at any time [5]. Researcher has investigated new approaches for user profiling on mobile devices. Compared to the classic Web-based solutions, the motivation is two-sided: On the one hand, people use apps instead of the web browser for online activities. On the other hand, the mobile device enables access to a wide range of new kind of data like GPS and voice records. As a consequence, researchers tried user profiling on acoustic measurements [42], phone call logs, and location-based information [43–45]. It is also conceivable that built-in cameras can be used for user profiling and several scientific works went into this direction [46,47]. In addition, Noulas et al. [48] and Wang et al. [49] estimated interests with mobility tracking technologies. The first article monitored that the mobility and social interactions in a conference environment and achieved an accuracy of 74% and 80%. The second article predicted points-of-interests.

In the past three years, several authors proposed to use the list of installed mobile apps on users' devices as a meaningful input stream for profiling. Such an approach does not read any data from the apps itself; instead, it just analyzes information about which apps are currently installed on the device. From the privacy perspective, this piece of information is less sensitive than email content or other data types mentioned before. In doing so, the authors estimated personality traits [50], demographics like gender, age, and salary [51,52], life events [53], and religion [54]. All studies were proof-of-concepts by semi-automated approaches or by the usage of machine learning techniques, such as Support Vector Machine or Random Forest. The resulting models were highly promising, in particular for e-commerce on mobile. Measured accuracy and precision were outstanding in some areas. Although many areas have been covered, there is still no research study on the prediction of interests. Seneviratne et al. [54] mentioned the idea about ‘interest prediction from a snapshot of apps installed on a smartphone’ in the discussion part of their article, but they left the investigation for future work. Thus, the first contribution of the present article is to close that research gap by measuring the performance of predictive models for interests, using a snapshot of currently installed mobile apps on users' smartphone.

2.4. Evaluation of recommender systems

There are several ways to evaluate recommender systems, but often in incompatible ways. One of the reason for the different evaluation methods are the varying targeted objectives [55]. Moreover, the authors observed strong differences in terms of the chosen items (which), timing (when), and presentation style (how). Only a few frameworks are able to cover the whole range of different approaches. However, accuracy, precision and recall are popular and well-established metrics for the evaluation. Cleverdon and Kean [56] proposed the metrics in the sixties for evaluating information retrieval systems. With the emerging e-commerce at the turn of millennium, accuracy, precision, and recall were introduced for measuring the effectiveness of online recommender systems, for instance in the study of Basu et al. [57] and in the analysis of Sarwar et al. [58]. On a regular basis, research studies have been published to show how recommender systems are to be evaluated. Two of these articles are from Herlocker et al. [59] and Shani and Gunawardana [60], in which the authors continued to suggest the usage of accuracy, precision and recall. Some researchers prefer the ‘F1 score’ [61] or ‘Area Under the ROC Curve’ [62] to take the trade-off between precision and recall into account [63,64].

In industry, the so-called Click-through rate (CTR) is a de-facto standard to assess the performance of a digital recommender system [65,66]. It measures the effectiveness of clickable recommendations and advertisements displayed on screens by counting the number of clicks per view. To the best of our knowledge, none of the mentioned app-based predictive models have been scientifically evaluated with regard to its effectiveness. We speculate that the unknown economic value prevents companies from adopting these models. A research study may resolve this unclear situation. Thus, the second objective of the present article is to close that research gap with an empirical study which investigates the CTR of the predictive models in a field experiment.

3. Research design

3.1. Research questions

Previous section reveals that information about personal interests is crucial in developing an effective recommender system, however, knowledge about an individual's interests is difficult to

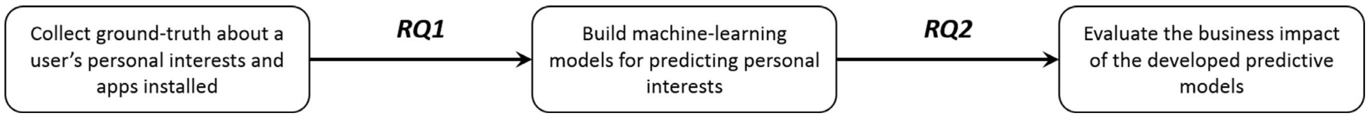


Fig. 1. Research framework.

obtain in the mobile context. As smartphone is a highly personalized and ubiquitous device and apps each consumer has installed could strongly mirror personal interests, the first research question of this work thus becomes:

RQ1: How accurate is the prediction of personal interests from a snapshot of apps installed on a smartphone?

Existing research typically evaluates a predictive model in terms of accuracy, or precision and recall. Although these metrics make sense in the field of computer science, they are not able to reflect the business impact a predictive model brings in a real business case. The absence of knowledge about the economic value of the predictive model prevents companies from adopting it in practice. To bridge the gap between technologies and business, this work also tries to answer:

RQ2: What is the effectiveness of the method for clickable recommendations?

To answer the research questions, two empirical studies have been conducted in this work. In Study 1, ground-truth about a smartphone user's personal interests as well as her installed mobile apps were collected. Afterwards, predictors were generated and machine-learning models were developed to predict each user's personal interests, thus addressing RQ1. In Study 2, the approach from Study 1 was integrated into an existing recommender system, which enables it to offer personalized recommendations based on the predicted knowledge about each user. In the end, a large-scale field experiment was conducted in a real business setting to quantify the business impact of our predictive models thereby answering RQ2. Fig. 1 shows the overall research framework and the link between the two research questions.

3.2. Predictors

One important factor for building an accurate predictive model is to select meaningful and representative predictors. Previous research that focused on mobile user profiling used the categories of the installed apps as predictors [67]. However, existing categorization in app stores is inaccurate because each app developer can freely choose which category the app belongs to when publishing it. Further, the number categories in apps stores is not very granular (in Google Play Store about 50), thus, apps with different descriptions/purpose can belong to the same category. Consequently, apps with similar description could belong to different categories, which significantly pollute the quality of predictors.

As the description of each mobile app in app stores is usually well written and provides more details about the app's features and functionality, this piece of information is leveraged to generate predictors. First, we crawled the Google Play Store to retrieve title and description of each mobile app. Based on that, a latent Dirichlet allocation (LDA) was applied to generate topic models for each app. LDA is a generative statistical model developed by Blei et al. [68], which has been widely used in natural language processing. It is a hierarchical Bayesian model where each document is described as a weighted mixture of finite topics and each topic is modeled as weighted key words. After applying LDA, each mobile app is explicitly represented as topic probabilities, which are numbers between 0 and 1. The number of topics was defined as one thousand in order to provide enough details as well as to avoid information overload. For each participant, her topic probabilities are accumulated according to the corresponding apps she has installed. For each participant, a vector of one-thousand probabilities serves as predictors.

3.3. Predicted variables

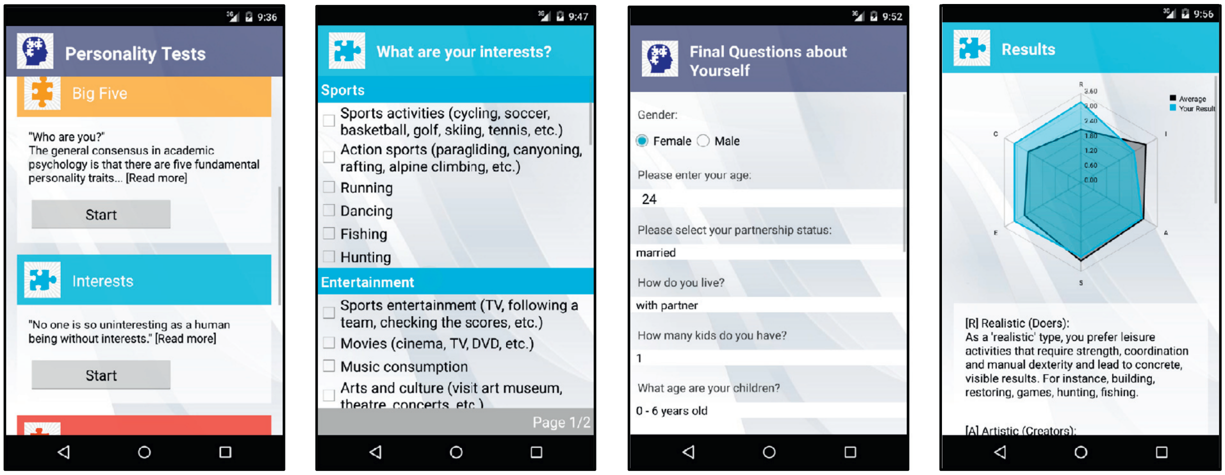
The two variables that were predicted are *interest* and *gender*. As already motivated in previous sections, interests are highly relevant for recommender systems and has never been predicted with the current app-based method in research. Together with our industry partner, we defined the six most relevant interests according to the products in the investigated online shop in Study 2. The six personal interests under study were: photography, running, health, technologies, music, and commuting. Commuting is a behavior rather than an interest, but it was also predicted because some of the recommended products had been frequently purchased by commuters. The second variable *gender* is a popular variable in user profiling activities and is therefore also investigated. Unlike *interest* where its values vary from study to study, the values for *gender* are always the same (female, male) and therefore helps other researchers to compare their own methods with ours.

3.4. Design of Study 1

3.4.1. Data collection

To collect ground-truth about user's personal interests and gender and her installed apps, we developed an Android app. The app is described as a test game where users provide answers to different personality and personal interests questions to know more about themselves as well as to compare their results with the average value of other people who have also played the game. Fig. 2 presents the screenshots of the mobile app.

When the app is opened for the first time, the user will be requested to accept the privacy policy. The policy contains a consent for the scientific use of the anonymized user data with the objective of "improving personalization on applications". We mention in the policy that only the list of installed apps on the user's device will be collected. If the user rejects the privacy condition, the app is forced to quit and no data will be collected and analyzed. If the user accepts the privacy terms, a background process is initiated, which retrieves app installation logs from the device: Android provides a public API called 'android.content.pm' for developers to retrieve a list of installed apps from each Android device. The list contains app name, package name, installed date, version of operation system, etc. We consider package name other than app name because it is a unique identifier to represent each individual app. The list is sent directly to our backend Web server and stored in a database. Meanwhile, the user is on the landing page, as shown in Fig. 2(a), where she can choose between several personality and personal interests tests. Each test is shortly explained and contains a questionnaire. A screenshot of the questionnaire about personal interests is shown in Fig. 2(b). The user can tick any number of interests. At the end of the first completed questionnaire, questions about demographics are displayed to the user, as shown in Fig. 2(c). Gender is used in our models, whereas



(a) (b) (c) (d)

Fig. 2. Screenshots of the mobile test game. From left to right: (a) landing page, (b) questions about interests, (c) demographic questions, and (d) feedback interests.

all other demographic data are only used for the descriptive statistics in Section 4.1. Finally, the app presents the results of the tests and compares the answers with the average of the other people who have already participated in the game. The results are mostly illustrated by a spider graph, as shown in Fig. 2(d). This is the only incentive for users to download the app and to participate in the study. They do not receive any additional monetary compensation.

Once all the questions are answered, the answers will be transmitted immediately to our Web server. It is impossible to redo the questionnaire on the same device more than once. In addition, if the user goes to the next page inside a questionnaire, it will not be possible to go back to change answers. By these restrictions, we prevent users from providing their own device to others who also want to do the test.

The app is listed on Google Play Store for free usage. We recruited users via Facebook promotion to download the app. For details on the participant recruitment strategy, please see Section 4.1.

After collecting data, participants who fail to answer all the questions that measure personal interests and gender are excluded in the analysis. To control for data quality, a simple mathematic problem ("$3+7=?$") was added in the questionnaire as attention check to screen out irrational answers. The sampled data thus serve as ground-truth to train and test our predictive model afterwards.

3.4.2. Regression analysis and validation

Because the relationship between interests/gender and input features developed from LDA could be non-liner, the Random Forest algorithm [63,64] was applied due to its ability to capture both linear and non-linear relationships and it usually performs better than other models. Random Forests is also a good choice in cases in which there are a large number features [69] or fast computation speed for real-time applications is needed [70]. In addition, Random Forest provides insights on what factors are more important in model generation and it almost cannot overfit [63,64], which makes models less sensitive to variance.

We divided our data samples randomly into two sets: 70% samples in a training set and 30% samples in a test set. Parameters of the predictive models like number of predictors to consider at each branch split in Random Forest is tuned through cross-validation [63,64] on the training samples. The best-performed model is then applied on the separate test data set to check the predictive precision and recall.

3.5. Design of Study 2

3.5.1. Integrating with a recommender system

We worked together with our industry partner to integrate the models with its existing recommender system. The partner is a large telecom provider in the investigated European country and it has its own brand app with more than 100,000 downloads. In the app, our partner runs an online shop and sells technical accessories to their customers. Fig. 3 shows five examples of its product offerings. The product recommendations were displayed on the top of the landing page. The existing system was able to provide recommendations either in a randomized order or based on product popularity. We extended the system with a third option for real-time personalized product recommendations based on each individual's interests and gender, predicted from the snapshot of installed apps as shown in Study 1. The recommender system was designed as simple as possible: The predicted interests/gender of a user were matched with the interests/gender associated with the products in the database. (That means the system is content-based and does not use collaborative filtering.) The association between interests/gender and products in the database was manually set in advance, according to partner's practical experience. For example, taking the articles from Fig. 3, the photo printer was tagged with 'Photography', the ski tracker and fitness tracker with 'Sport', the charger with 'Commuting', and the headphones with 'Music'. Some recommended products were further labeled according to gender preferences. For instance, headphones in white and rose gold were designed especially for women and thus, tagged as product for women.

The predictive models according to Study 1 had been integrated into our partner's mobile app and recommender system. Fig. 4 demonstrates the workflow of the extended system. First, the partner's mobile app retrieves a list of all installed apps from each smartphone through the Android API. As this piece of data is openly accessible, no additional permission is required from a user's perspective. Afterwards, the app sends the list to our machine-learning model, which in return predicts the personal interests and gender for each user. In the next step, the recommender system communicates with the product database and picks out suitable products based on the estimated user profile in real-time. Finally, a set of generated product recommendations is presented in the app. The user can click on one of the products to get more information and to trigger the process of purchasing the product in our partner's online store.

Fig. 3. Product offerings in industry partner's online shop. From left to right: photo printer, wearable for skiing, fitness tracker, charging device, headphones.

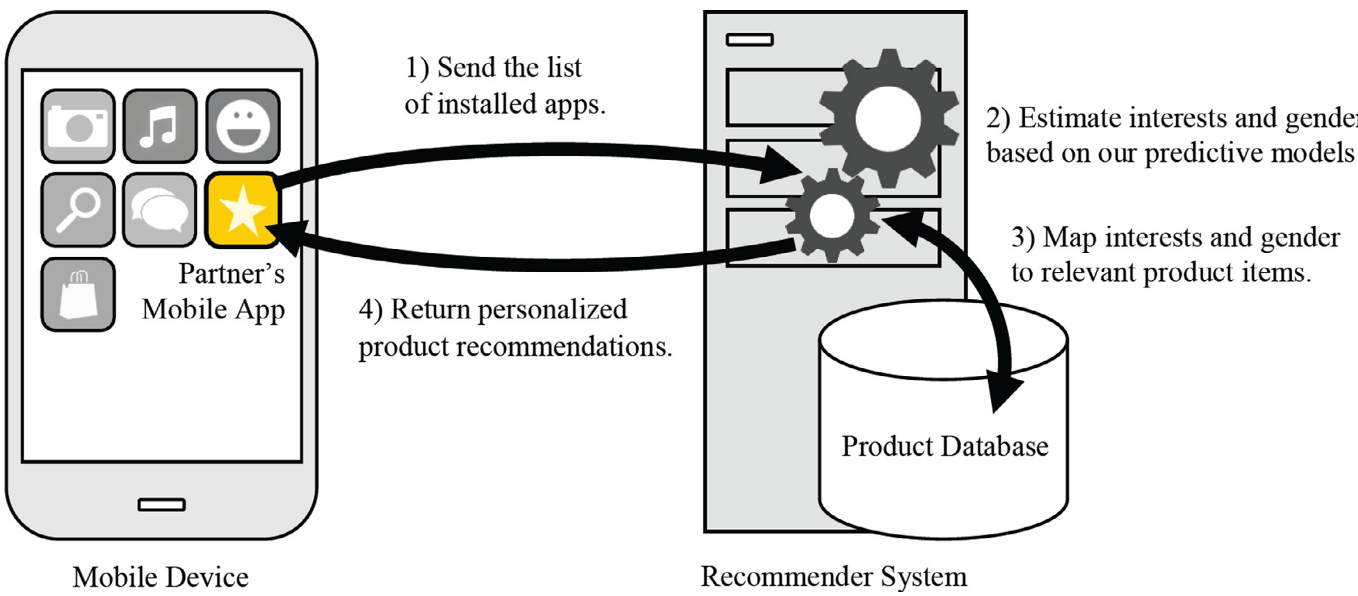


Fig. 4. Integrated predictive models and real-time recommender system.

3.5.2. Design of the field experiment

The field experiment uses the recommender system described above. As mentioned, the system offers three options for generating recommendations. To answer the second research question, the three options are evaluated and compared with each other. Users were assigned randomly in one of the three groups each time they landed on the recommendation view. They were unaware about that process, i.e. it was a blinded experiment. Here are the treatment groups:

- *Random:* Each user in this group got a list of non-personalized product recommendations. The displayed products were selected randomly by the system.
- *Top-selling:* Each user in this group got a list of non-personalized product recommendations. The displayed products were selected based on the top-selling product list. The list was dynamically updated according to the sales situation.
- *App-based:* Each user in this group got a list of personalized product recommendations based on her predicted personal interests and gender.

The definitions of the groups are comparable with the definitions in the study by Hegelich and Jannach [71] about the effectiveness of different recommender algorithms in the mobile Internet. The authors used two non-personalized item lists, namely top-selling and top-rating, for the comparison with more sophisticated personalized methods. As an additional control group, they used a quasi-randomly ordered item list, mainly based on the item release date or company contracts.

To guarantee the quality of the field experiment, only the recommended products were different according to each user's predicted profile, while the user interface design (e.g., number of recommended products, app layout, frequency of interaction, text size and font, etc.) remained the same. To prevent information overflow, the list of recommended products contained always exactly six products. The field experiment ran within a pre-defined time span, during which no changes are made to the running system. Similarly, the product database was kept the same for all the groups at any time.

3.5.3. Data analysis

The objective of the field experiment is to answer the second research question by comparing the effectiveness of three treatment groups. The metric is the Click-Through Rate (CTR) because it is widely used in research and practice to measure the success of advertisement on screens [65,66]. The distribution of CTR in each group is expected to be non-normal because the click actions on recommendations are typically rare. Therefore, a Kruskal–Wallis H test [72], which is ranked-based and non-parametric, is used to determine if there are statistically significant differences between the three groups. Since Kruskal–Wallis H only tests if at least two groups are different, Dunn's multiple comparison [73] is further applied to determine which of these groups differ from

Table 1
Characteristics of participants in study 1 ($N = 904$).

Respondents	Range	In percentage	Respondents	Range	In percentage
Gender	Female	75.7%	*Job Type*	Pupil/Student	33.6%
	Male	24.3%		Full-time job	33.8%
Age	10–19	37.3%		Part-time job	10.7%
	20–29	37.8%		Homemaker	7.5%
	30–39	15.8%		Job-seeking	6.9%
	40–49	4.1%		No job	4.6%
	50–59	2.9%		Disabled	2.2%
	>60	0.7%		Retired	0.6%
	No Answer	1.4%	*Highest Education*	No degree	7.5%
Net monthly salary ($)	<500	27.2%		Elementary school	3.1%
	500–1000	16.4%		Middle school	26.9%
	1000–1500	13.4%		High school	34.6%
	1500–2000	11.6%		Vocational school	15.6%
	2000–3000	6.3%		College/University	11.9%
	3000–4000	3.4%		Promotion/Habilitation	0.3%
	4000–5000	2.0%	*Language*	English	45.1%
	> 5000	1.7%		German	54.4%
	No Answer	18.0%		Others	0.4%

each other. The p-value is adjusted according to Bonferroni. Additionally, the effect size r is also calculated according to Rosenthal [74] to demonstrate how strong or weak the differences are. The classification of Cohen [75] is used to assess the effect size: $r = .10$ corresponds to small effect; $r = .30$ to medium effect; $r = .50$ to strong effect.

4. Results

4.1. Results of Study 1

4.1.1. Study participants

The described Android app was placed on Google Play Store for free download (https://play.google.com/store/apps/details?id=ch.autoidlabs.newpersonalitytest). The app was available in two countries on two continents and in two different languages with the aim to demonstrate the robustness of our method related to culture and language. A Facebook promotion was conducted in both countries for two weeks in August 2015. During this period, the app promotion page was shown to 211,090 people and 1465 of them installed the app. The conversion rate for installation was around 0.7%. Additional 910 people installed the app directly from Google Play Store without being reached by our Facebook promotion. In total, there were 2375 users who accepted the privacy policy, in which 920 people answered both the demographic and personal interest questions. However, eleven participants failed in the control question. Therefore, data from them was excluded in the analysis. In addition, five participants were removed in a later stage because their devices had a feature vector (see next subsection) of zero. The demographics of the 904 remaining participants are summarized in Table 1.

4.1.2. Performance of predicting personal interests and gender

Before modeling the six interests and gender, the collected research data was further cleaned. Due to the fact that some mobile apps were pre-installed by phone manufacturers or service providers on each Android device, such apps were excluded in building the topic models because they did not reflect users' interest or gender specific behavior. (Pre-installed apps are marked as such and therefore can easily be distinguished from the others.) In the end, a total of 27,991 distinct mobile apps had been installed by our 904 participants. The title and the description of each app in Google Play Store was used to feed the LDA algorithm with a pre-defined 1000 topics. After applying LDA, each app had a feature vector that consisted of 1000 components, where each component represented the probability for the app to belong to a specific topic. The sum of the probabilities of the 1000 topics of each app was 1. According to all the apps a participant had installed, the 1000 features were first summed up and then normalized to serve as predictors for the Random Forest algorithm.

The goal of the predictive models is to identify smartphone users who are interested in specific activities like photography and running. Because the data set was unbalanced in terms of the two classes (i.e., people interested and not interested in an activity), metrics like overall predictive accuracy were not used. Instead, the developed models were focused on precision and recall [60] for the positive class, the class that identifies people who have an interest. Precision is defined as the fraction of the retrieved instances that are relevant. It is a measure of the accuracy provided that a specific class has been predicted. On the other hand, recall is defined as the fraction of relevant instances that are retrieved and it is a measure of the ability of a model to select instances of a certain class from the whole data set.

The predictive results are displayed in Table 2. The baseline for performance comparison is a random model, which is defined as randomly allocating each user in the test set into one of the two classes (interested and not interested) according to the class probability distribution. The approach is similar to that of previous studies [54,67,76,77]. Take the first model that predicts photography for example, 34.48% of instances that were predicted as having interest by a random model were actually interested in photography, while the remaining 65.52% had been misclassified. On the other hand, our Random Forest model was able to increase the precision by 80.72% from 34.48% to 62.32%, while kept the recall at an acceptable level (38.74%). Given the large number of smartphone users, a high precision model is of more value in this context as long as the recall is not too low.

For music and commuting, where more than 80% of participants were interested in, our machine learning models were able to improve the predictive performance by only 5.84% and 7.81% respectively. However, for the other four interests, our models performed on average 69.81% better in terms of precision and 19.30% better in terms of recall.

In addition to personal interests, gender prediction is shown in Table 2 as well. The precision of predicting a male was 210.00% higher than random and the corresponding recall was 106.91% higher. Similarly, our models also predicted females 9.26% better than random in precision and 20.92% better in recall.

Table 2
Comparing performance of predictive models ($N_{training} = 637$, $N_{test} = 272$).

	Precision of random model	Precision of our model	Precision improvement	Recall of random model	Recall of our model	Recall improvement
Interest						
Photography	34.48%	62.32%	80.72%	39.60%	38.74%	−2.18%
Running	20.69%	34.09%	64.77%	21.82%	27.78%	27.31%
Health	34.69%	59.42%	71.27%	33.33%	39.05%	17.14%
Technology	29.03%	47.17%	62.47%	26.47%	35.71%	34.92%
Music	80.89%	87.20%	7.81%	81.61%	81.78%	0.20%
Commuting	81.11%	85.84%	5.84%	80.73%	85.09%	5.39%
Gender						
Female	79.43%	86.78%	9.26%	77.21%	93.36%	20.92%
Male	22.22%	68.89%	210.00%	24.56%	50.82%	106.91%

Table 3
Automatically computed top key words and topics related to interests and gender.

Interests	*Key words*			*Interests*	*Key words*		
Photography	Photo	Frame	Camera	Running	Training	Exercise	Goals
	Photograph	Exposure	Modes		Fitness	Workout	Gym
	Color	Blue	Paint		Message	Text	Email
	Shape	Thin	Eyebrow		Run	Race	Participant
	Black	White	Border		Food	Diet	Healthy
Health	Weight	Height	BMI	Music	Status	WhatsApp	Share
	Food	Diet	Healthy		Bus	Route	Time
	Training	Exercise	Goals		SMS	Messages	Send
	Citizen	Unwanted	Spam		Overview	Glance	cooperation
	Market	Trading	Investment		Skin	Glow	theme
Technology	Hunting	Speaker	Deer	Commuting	Active	Session	Shuffle
	Status	WhatsApp	Share		Maps	Lake	Areas
	Network	Wi-Fi	networks		Submit	Interact	Anonymous
	Truck	Loads	Dealer		Forward	Bookmark	Audiobook
	Car	Vehicle	Auto		Com	WWW	HTTP
Gender	*Key words*						
Female	Period	Cycle	Menstrual				
	Game	Play	Fun				
	Eyes	Makeup	Looks				
	Photo	Frame	Camera				
	Solve	Complex	Equation				

To illustrate what key words were most relevant in predicting personal interests and gender based on her installed apps, we present the top-five most powerful topics (rows) with their top-three associated key words (columns) for each model in Table 3. As shown in the table, some topics can be well structured. For instance, topics about photography terminologies, shape, and color were frequently used to distinguish people who were interested in photography from others. Similarly, topics about body mass index, exercises, and health eating were widely used in predicting interests in running and health, whereas topics about maps, reading, and Internet were related to commuting behavior. Similarly, topics like period tracking, game playing, makeup were most frequently used in gender prediction also makes total sense. Until here, RQ1 is addressed.

4.2. Results of Study 2

4.2.1. Study participants

The field experiment evaluated the user profiling models in a mobile recommender system. The experiment was conducted in a western European country. Table 4 displays the number of users, the number of clicks, the impression and the CTR of each group. All users in the app-based treatment group had agreed to provide their mobile app data in order to get real-time personalized product offers. As shown in the table, the app-based group has the highest average CTR with 3.48%, followed by the random group with 2.15% and top-selling group with 1.06%.

The average CTR for the top-selling group was counterintuitively lower than for the random group. However, this observation is in line with the results of the earlier mentioned study by Hegelich and Jannach [71] where the quasi-randomized control group outperformed the top-selling group in all cases. One possible explanation is that the six presented top-selling products in our online shop lacked under novelty and diversity [78].

Table 4
Comparing performance of three treatment groups.

	Random	Top-selling	App-based
Users	23,842	23,890	34,346
Impressions	207,024	205,848	429,498
Clicks	4459	2185	14,951
Average CTR	2.15%	1.06%	3.48%

Fig. 5 provides an overview on how the distribution of CTR looked like within the different groups. More than 80% of the users in all groups had a CTR of zero which means that they had never clicked on the displayed six recommendations. Users with a CTR larger zero are grouped in intervals between 0 and 10 percent, 10 and 20 percent, 20 and 100 percent. Overall, the CTRs in all the three groups were not normally distributed, therefore, a non-parametric test is appropriate in this case to compare the differences.

4.2.2. Comparison of CTR's

A Kruskal–Wallis H test was conducted to compare the measured CTRs of different groups. Results showed that there was a statistically significant difference in CTR between the groups, $\chi 2(2) = 1999$, $p < .001$, with a mean rank of 40,415 for the random

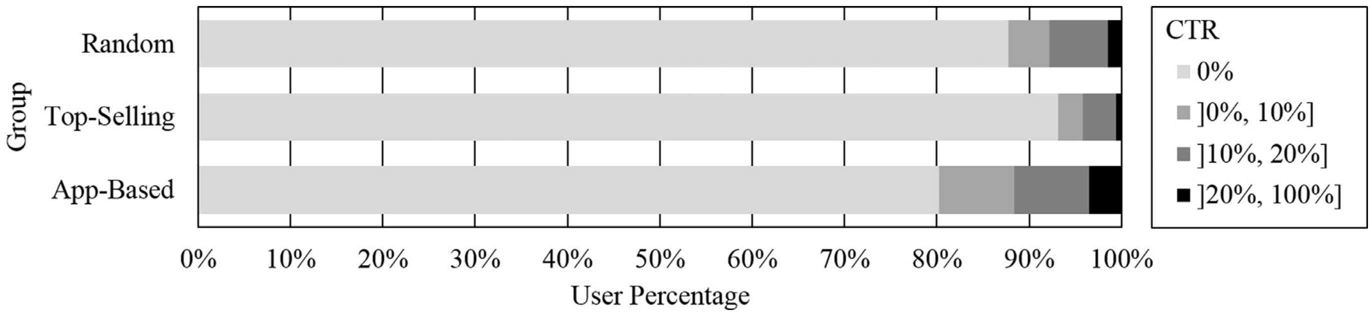


Fig. 5. Distribution of CTR of the three groups under study.

group, 38,188 for the top-selling group, and 43,456 for the app-based group.

Dunn's post hoc test compared the groups pairwise. The test rejected the three null hypotheses that each combination of two groups are equal, i.e. all groups significantly differed from each other. Dunn's test showed a significant increase from the random group to the app-based group (test statistic = 3041, $p < .001$), from top-selling to random (test statistic = 2227, $p < .001$), and from top-selling to the app-based (test statistic = 5268, $p < .001$). The measured effect sizes r was .09 between random and top-selling, .10 between group random and app-based, .18 between top-selling and app-based.

To sum up and answer the second research question, the CTR of the app-based group was significant better than of the other two groups. The improvements were on average plus 61.86% compared to the random group, and plus 228.30% compared to the top-selling group. Effect sizes of .10 and .18 mean that the used method to personalize recommendations had a small to medium effect on CTR.

5. Discussion, limitation and future work

5.1. Findings and contributions

A holistic comprehension of consumers' past shopping behavior and current demands can lead to a better understanding of their purchase decision making in the future. However, gaining knowledge about an individual's personal interests and gender is not straightforward but digital technologies provide new opportunities to achieve it. Researchers and practitioners used to predict personal interests and gender based on analyzing cookies from Web browsers. This approach is scalable and has been well adopted in different context. Nevertheless, cookie-based profiling approaches are not well applicable on mobile devices. The difficulty is that most time on mobile devices is spent on convenient native apps and not in Web browsers, but cookies do not work in native apps.

Consequently, this work used an open and scalable approach of predicting personal interests and gender based on a snapshot of an individual's installed mobile apps. A study involving 904 smartphone users was conducted to collect ground-truth about one's personal interests, gender and her installed apps. Afterwards, machine-learning models were developed and results showed that the developed models are able to predict on average 48.81% better than a random guess in terms of precision and 13.80% better in terms of recall. Due to the fact that predictors of our models are openly accessible, this machine-learning approach can be integrated into any mobile app to generate user profiles. Furthermore, it overcomes the cold start problem under which many other approaches suffer.

Besides product recommendations, the presented profiling approach promise hidden potential for other areas as well. A current example is the American presidential election 2016. Ted Cruz's presidential campaign heavily invested in a data analytics company that specializes in voter profiling [79]. The company analyzed Facebook data (posts, likes, demographics) for micro-targeting. After Cruz's campaign ended, the company started to work for Donald Trump's campaign [80] which might have led to his victory [81]. However, older populations typically rarely use Facebook, but represent a significant proportion of the electorate. Using our app-based approach instead could open up the opportunity to reach a much broader audience and thus, to generate an even more powerful impact in politics.

Existing research evaluates the performance of machine-learning models by examining their accuracy, or precision and recall. Although these metrics are objective and widely applied, they are not able to reflect the economic values of using them, which might prevent practitioners from adopting the approach in reality. To bridge the gap between applied machine-learning and business research, an additional field study was further conducted. We integrated the developed predictive models with our industry partner's recommender system and measured its impact with 73,244 participants in a real business setting. Empirical result showed that by adopting our models to predict personal interest, the CTR of a recommender system can be significantly improved by 61.86% compared to random recommendations and 228.30% compared to recommendations of top-selling products. To sum up, this work proposes a practicable approach that enables companies to implement best practice quality from research into the real world in a scalable and low-cost manner.

5.2. Managerial implications

This work brings implications to managers and practitioners as well. First, we provide managers with a real-time tool to gain knowledge about each consumer's personal interests and gender at a low cost. Different from existing approaches that leverage data only from desktop computer activities or only accessible to phone manufacturers or telecommunication service providers, our approach can be integrated into any mobile app without the installation of additional surveillance software. As a result, companies can integrate our models into their existing mobile apps thereby enabling real-time personalized product recommendations or user interactions.

Companies like Netflix and Amazon typically combine article features with user features to recommend the most likely products that a user would buy. User features used in this context are typically each user's browsing activities. In addition to information like what products a user has already looked at or purchased, our approach provides existing machine-learning algorithms with other valuable input features, i.e., personal interests. Consequently, the performance of recommender systems could be significantly improved, as shown in our field experiment.

In addition, the proliferation of mobile devices make it possible to obtain large amount of data that was not accessible in the past. Researchers in the field of applied machine-learning have shown opportunities of leveraging such new data sets to predict various aspects in our daily life. However, lacking studies showing the business values of these approaches prevents managers and practitioners from quantifying the corresponding benefits and further adopting them. Consequently, this work shows that a 48.81% better precision in predicting personal interests and gender is able to increase the CTR of in-app product recommendation by up to 228.30%.

Existing recommendation systems can vary significantly in terms of product content, timing, presentation style, etc. [55]. Some user interests are easier to predict (e.g. music with 87% precision), others are more difficult (e.g. running with 34% precision). In conclusion, different effect sizes are expected and it is not clear at the beginning whether the effort and expense would pay off. Thanks to the simplicity and low realization costs of our approach, companies can quickly test the effect with low financial risk under their own conditions.

Although proved to be useful, leveraging personal data for personalized recommendation might trigger privacy concerns. Previous research in the field of user profiling mostly requires the installation of additional surveillance software and has to trace user behavior for a certain period of time. Compared to existing methods, the proposed approach should trigger less privacy concerns because only a snapshot of app installation log is used. Also, no additional software needs to be installed for data collection because all predictors used in the machine-learning models are openly accessible. Results from our field experiment also proved the relative low privacy concerns: More than 40% of consumers freely agreed to share their app installation logs so as to receive in-app product recommendations. Nevertheless, we still suggest companies that leverage the approach to explicitly inform users of what data is collected for what purpose and what benefits consumers will get in return. Companies should also provide users with the rights to opt-out for sharing mobile app data at any time.

5.3. Limitations and future work

There are four limitations in this work, which leads to opportunities for future research. First, the samples in Study 1 are not representative in terms of age, gender, income, etc. This could be resulted from the fact that we promoted the mobile app through Facebook, which brings potential selection bias. For instance, women are more likely to heavily use Facebook [82]. Similarly, around 80% of the participants in Study 1 have commuting behavior. We noticed that most Facebook promotion feeds were sent out between five and seven o' clock in the evening. Thus, it is possible that commuting smartphone users are more likely to see and download the app on their way back home. Future research is called to build more stable models with more representative samples.

Second, a snapshot of app installation log was used to generate predictors together with a LDA topic model. In the current model, all installed apps were perceived to be equal. However, it is possible that some installed apps were used more frequently than others. Taking the frequency of app usage could bring more insights on each consumer's personal interests, but on the other hand might also trigger more privacy concerns. Further research is thus called to examine the trade-off between potential improvement on predictive precision and the raised privacy concerns.

Third, participants of the field experiment are not completely randomly assigned in one of the three groups. For data protection reasons, it had to be considered whether a customer had consciously decided against personalization. People in the random and top-selling groups are users who opted-out for personalized content. Thus, our results of the field experiment might suffer under a selection bias. Finally, the investigated online shop had a specific product range. Most of the products are technical accessories. Further experiments in other environments like in food or finance sector are suggested in future work to verify the efficiency and applicability of the approach in general.

6. Conclusion

Generating user profiles from installed mobile apps is a powerful method for recommender systems on mobile devices. This work developed machine-learning models that can significantly increase precision and recall in predicting a user's interests and gender. That app-based method is open to all app providers and can be easily integrated into existing apps. Furthermore, a large-scale field experiment demonstrated its applicability and efficiency in a real business case for the first time. The click-through rate significantly increased using the app-based method. As with all other personalization approaches, the privacy aspect must be considered. However, the presented method with mobile app installation log seems to be less intrusive than others that track online activities or read out social media content. No privacy concerns were observed during the experiment. To sum up, it is a practicable user profiling approach which allows companies to implement best practice quality from research in the real world.

Acknowledgements

We thank 42matters AG, an app market data analyst, and Swisscom AG, a telecommunications provider in Switzerland, for supporting the research project. The large-scale field experiment was conducted and analyzed together. Further, the companies provided support with expert knowledge and experience in the field of mobile technologies, m-commerce and decision-support systems.

Funding

This work was supported by the Swiss Commission for Technology and Innovation (CTI) [project no. 17041.1, project no. 19317.1].

References

[1] Statista, Number of smartphone users worldwide from 2014 to 2019, 2016. https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/.
[2] D. Chaffey, Statistics on consumer mobile usage and adoption to inform your mobile marketing strategy mobile site design and app development, http://www.smartinsights.com/mobiel-marketing/mobile-marketing-analytics/mobile-marketing-statistics, 2016.
[3] Statista, Mobile advertising spending worldwide from 2015 to 2017, 2016. https://www.statista.com/statistics/272859/mobile-advertising-spending-worldwide/.
[4] P.T. Chen, H.P. Hsieh, Personalized mobile advertising: its key attributes, trends, and social impact, Technolog. Forecast. Soc. Change 79 (2012) 543–557.
[5] P. Smutkupt, D. Krairit, V. Esichaikul, Mobile marketing : implications for marketing strategies, Int. J. Mobile Market. 5 (2010) 126–139.
[6] M. Wedel, R.T. Rust, C.P. Md, Up close and personalized : a marketing view of recommendation systems, System (2008) 3–4.
[7] J. Zhang, M. Wedel, The effectiveness of customized promotions in online and offline stores, J. Market. Res. 46 (2009) 190–206.
[8] S. Sivapalan, A. Sadeghian, H. Rahnama, A.M. Madni, Recommender systems in e-commerce, in: World Automation Congress Proceedings, 2001, 2014, pp. 179–184.
[9] R. Chen, A domain ontology in social networks for identifying user interest for personalized recommendations, J. Univ. Comput. Sci. 22 (2016) 319–339.
[10] J. Zhang, C.-Y. Chow, GeoSoCa : exploiting geographical, social and categorical correlations for point-of-interest recommendations categories and subject descriptors, in: 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, 2015, pp. 443–452.

[11] D. Li, Q. Lv, X. Xie, L. Shang, H. Xia, T. Lu, et al., Knowledge-based systems interest-based real-time content recommendation in online social communities, Knowledge-Based Syst. 28 (2012) 1–12.
[12] S. Usharani, A. Basu, System for product recommendation in e-commerce applications, Int. J. Eng. Res. Develop. 11 (2015) 52–56.
[13] M.J. Pazzani, A framework for collaborative, content-based and demographic filtering, Artif. Intell. Rev. 13 (1999) 393–408.
[14] D. Goldberg, D. Nichols, B.M. Oki, D. Terry, Using collaborative filtering to weave an information tapestry, Commun. ACM 35 (1992) 61–70.
[15] M. Claypool, T. Miranda, A. Gokhale, P. Murnikov, D. Netes, M Sartin, Combining content-based and collaborative filters in an online newspaper, in: Proceedings of Recommender systems workshop at ACM SIGIR., 1999, pp. 40–48.
[16] B. Smyth, P. Cotter, Personalized TV listings service for the digital TV age, Knowl.-Based Syst. 13 (2000) 53–59.
[17] S.-S. Weng, M.-J. Liu, Personalized Product Recommendation in E-Commerce, in: Proceedings of the 2004 IEEE International Conference on E-Technology, E-Commerce and E-Service (EEE'04), 2004.
[18] B. Mobasher, H. Dai, T. Luo, M. Nakagawa, Discovery and evaluation of aggregate usage profiles for Web personalization, Data MiningKnowl. Discovery 6 (2002) 61–82.
[19] P.S. Yu, Data mining and personalization technologies, database systems for advanced applications, in: 1999. Proceedings., 6th International Conference on, 1999, pp. 6–13.
[20] H.J. Smith, S.J. Milberg, S.J. Burke, Information privacy: measuring individuals' concerns about organizational practices, Manag. Inf. Syst. Quarterly 20 (1996) 167–196.
[21] E. Aguirre, A.L. Roggeveen, D. Grewal, M. Wetzels, The personalization–privacy paradox: implications for new media, J. Consum. Market. 33 (2016) 98–110.
[22] Y. Liu, W. Liu, C. Jiang, User interest detection on web pages for building personalized information agent, Adv. Web-Age Inf. Manag. 3129 (2004) 280–290.
[23] A. Kurtz, J. Mostafa, Topic detection and interest tracking in a dynamic online news source, in: Proceedings of the 3rd ACM/IEEE-CS Joint Conference on Digital Libraries, 2003, pp. 122–124.
[24] H.R. Kim, P.K. Chan, Learning implicit user interest hierarchy for context in personalization, in: Proceedings of the 8th International Conference on Intelligent User Interfaces, 2003.
[25] Y. Yang, Web user behavioral profiling for user identification, Decis. Support Syst. 49 (2010) 261–271.
[26] J.J. Jung, Collaborative web browsing based on semantic extraction of user interests with bookmarks, J. Univ. Comput. Sci. 11 (2005) 213–228.
[27] J. Hu, H.-J. Zeng, H. Li, C. Niu, Z. Chen, Demographic prediction based on user's browsing behavior, in: WWW '07: Proceedings of the 16th International Conference on World Wide Web, 2007, pp. 151–160.
[28] D.V. Phuong, T.M. Phuong, Gender prediction using browsing history, in: Knowledge and Systems Engineering, pringer International Publishing, 2014, pp. 271–283.
[29] D. Duong, H. Tan, S. Pham, Customer gender prediction based on e-commerce data. (2016) 91–95.
[30] T. O'Reilly, What is web 2.0: design patterns and business models for the next generation of software, Commun. Strategies 1 (2007) 17.
[31] D.N. Chin, W.R. Wright, Social media sources for personality profiling, in: Proceedings of the 2nd Workshop Emotions and Personality in Personalized Services, Aalborg, Denmark, 2014, pp. 613–620.
[32] J. Han, H. Lee, Characterizing user interest using heterogeneous media, in: World Wide Web Conference, 2014, pp. 289–290.
[33] A. Minamikawa, H. Fujita, J. Hakura, M. Kurematsu, Personality estimation application for social media, in: SoMeT, IOS Press, 2012, pp. 327–335.
[34] R. Wald, T. Khoshgoftaar, C. Sumner, Machine prediction of personality from facebook profiles, in: Proceedings of the 2012 IEEE 13th International Conference on Information Reuse and Integration, IRI 2012, 2012, pp. 109–115.
[35] J. Shen, O. Brdiczka, J. Liu, UnderstandingEmail Writers, Personality prediction from email messages, in: The 21st International Conference on User Modeling, Adaptation and Personalization, Rome, Italy, 2013, pp. 318–330.
[36] L. Guo, Q. Chen, W. Han, Y. Tian, Y. Cui, W. Wang, et al., Collaborative topic prediction model for user interest recommendation in online social networks, Int. J. Digital Content Technol. Appl. 6 (2012) 62–73.
[37] P. Bhargava, O. Brdiczka, M. Roberts, Unsupervised modeling of users' interests from their Facebook profiles and activities, in: International Conference on Intelligent User Interfaces, Proceedings IUI. 2015-Janua, 2015, pp. 191–201.
[38] C. Tang, K. Ross, N. Saxena, R. Chen, What's in a name: a study of names, gender inference, and gender behavior in facebook, Database Syst. Adv. Appl. (2011) 344–356.
[39] W. Deitrick, Z. Miller, B. Valyou, B. Dickinson, T. Munson, W. Hu, Author gender prediction in an email stream using neural networks, 2012 (2012) 169–175.
[40] T. Kucukyilmaz, B. Cambazoglu, Chat mining for gender prediction, Adv. Inf. (2006) 274–283.
[41] Z. Miller, B. Dickinson, W. Hu, Gender prediction on twitter using stream algorithms with n-gram character features, Int. J. Intell. Sci. 02 (2012) 143–148.
[42] F. Pianesi, N. Mana, A. Cappelletti, B. Lepri, M. Zancanaro, Multimodal recognition of personality traits in social interactions, in: Proceedings of the 10th International Conference on Multimodal Interfaces, 2008, pp. 53–60.
[43] G. Chittaranjan, J. Blom, D. Gatica-Perez, Mining large-scale smartphone data for personality studies, Personal. Ubiquitous Comput. (2013) 433–450.
[44] I. Trestian, A. Nucci, S. Ranjan, A. Kuzmanovic, Measuring Serendipity, Connecting people, locations and interests in a mobile 3 G network, in: Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference - IMC '09, Chicago, 2009, pp. 267–279.
[45] Y.A. De Montjoye, J. Quoidbach, F. Robic, A. Pentland, Predicting personality using novel mobile phone-based metrics, in: Social Computing, Behavioral-Cultural Modeling and Prediction, Springer, Berlin Heidelberg, 2013, pp. 48–55.
[46] J. Hu, L. Peng, L. Zheng, XFace: A Face Recognition System for Android Mobile Phones, in: Proceedings - 3rd IEEE International Conference on Cyber-Physical Systems, Networks, and Applications, CPSNA, 2015, 2015, pp. 13–18.
[47] H. Han, C. Otto, X. Liu, A.K. Jain, Demographic estimation from face images: human vs. machine performance, IEEE Trans. Pattern Analys. Mach. Intell. (2014) 1148–1161.
[48] A. Noulas, M. Musolesi, M. Pontil, C. Mascolo, Inferring interests from mobility and social interactions, NIPS Workshop (2009) 1–8.
[49] R. Wang, C.Y. Chow, Y. Lyu, V.C.S. Lee, S. Nutanong, Y. Li, et al., Exploring cell tower data dumps for supervised learning-based point-of-interest prediction (industrial paper), GeoInformatica 20 (2016) 327–349.
[50] R. Xu, R.M. Frey, D. Vuckovac, A. Ilic, Towards Understanding the impact of personality traits on mobile app adoption – a scalable Approach, in: 23rd European Conference on Information Systems, Münster, Germany, 2015, pp. 1–12.
[51] S. Seneviratne, P. Mohapatra, A. Seneviratne, A. Mahanti, Your installed apps reveal your gender and more!, ACM SIGMOBILE Mobile Comput. Commun. Rev. 18 (2014) 55–61.
[52] R. Xu, R.M. Frey, A. Ilic, Individual differences and mobile service adoption : an empirical analysis, IEEE BigData Service 2016, Oxford, UK, 2016.
[53] R.M. Frey, R. Xu, A. Ilic, Reality-mining with smartphones: detecting and predicting life events based on app installation behavior, International Conference on Information Systems, ICIS 2015, Fort Worth, United States, 2015.
[54] S. Seneviratne, A. Seneviratne, P. Mohapatra, A. Mahanti, Predicting user traits from a snapshot of apps installed on a smartphone, ACM SIGMOBILE Mobile Comput. Commun. Rev. 18 (2014) 1–8.
[55] F. Hernández del Olmo, E. Gaudioso, Evaluation of recommender systems: a new approach, Expert Syst. Appl. 35 (2008) 790–804.
[56] C. Cleverdon, M. Keen, Factors determining the performance of indexing systems, (1966).
[57] C. Basu, H. Hirsh, W. Cohen, Recommendation as classification : using social and content-based information in recommendation, in: Proceedings of the Fifteenth National Conference on Artificial Intelligence, 1998, pp. 714–720.
[58] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, Analysis of recommendation algorithms for e-commerce, Organization 5 (2000) 158–167.
[59] J.L. Herlocker, J.A. Konstan, L.G. Terveen, J.T. Riedl, Evaluating collaborative filtering recommender systems, ACM Trans. Inf. Syst. (TOIS) 22 (2004) 5–53.
[60] G. Shani, A. Gunawardana, Evaluating recommendation systems, in: Recommender Systems Handbook, 2011, pp. 257–298.
[61] C.J. Van Rijsbergen, Information Retrieval, Butterworth-Heineman, Newton, MA, USA, 1979.
[62] D. Bamber, The area above the ordinal dominance graph and the area below the receiver operating characteristic graph, J. Math. Psychol. 12 (1975) 387–415.
[63] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second etuded, Springer, New York, 2009.
[64] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning: With Applications in R, Springer, 2014.
[65] F. Garcin, B. Faltings, O. Donatsch, A. Alazzawi, C. Bruttin, A. Huber, Offline and online evaluation of news recommender systems at swissinfo.ch, in: Proceedings of the 8th ACM Conference on Recommender Systems - RecSys '14, 2014, pp. 169–176.
[66] C. Hand, H. Robinson, A. Wysocka, Internet advertising effectiveness: the effect of design on click-through rates for banners, Int. J. Advert.: Rev. Market. Commun. 26 (2007) 527–541.
[67] R. Xu, R.M. Frey, E. Fleisch, A. Ilic, Understanding the impact of personality traits on mobile app adoption – insights from a large-scale field study, Comput. Human Behav. 62 (2016) 244–256.
[68] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent Dirichlet Allocation, J. Mach. Learn. Res. 3 (2003) 993–1022.
[69] R. Caruana, N. Karampatziakis, A. Yessenalina, An empirical evaluation of supervised learning in high dimensions, in: Proceedings of the 25th International Conference on Machine Learning - ICML '08, 2008, pp. 96–103.
[70] L. Breiman, in: Random forests, Machine Learning, 45, 2001, pp. 5–32.
[71] K. Hegelich, D. Jannach, Effectiveness of different recommender algorithms in the mobile internet: a case study, in: Proceedings of the 7th International Conference on Intelligent Techniques for Web Personalization & Recommender Systems, 528, 2009, pp. 41–50.
[72] W.H. Kruskal, W.A. Wallis, Use of ranks in one-criterion variance analysis, Source J. Am. Stat. Assoc. 47 (1952) 583–621.
[73] O.J. Dunn, in: Multiple Comparisons Using Rank Sums, 6, Technometrics, 1964, pp. 241–252.
[74] R. Rosenthal, Parametric measures of effect size, in: H. Cooper, L.V. Hedges (Eds.), The Handbook of Research Synthesis, Russel Sage Foundation, 1994, pp. 231–244.
[75] J. Cohen, A Power Primer, Psycholog. Bullet. 112 (1992) 155–159.
[76] R.M. Frey, R. Xu, A. Ilic, Reality-mining with smartphones: detecting and predicting life events based on app installation behavior, in: Proceedings of the 36th International Conference on Information Systems (ICIS), Fort Worth, United States, 2015.

[77] W.R. Wright, D.N. Chin, Personality profiling from text: introducing part-of-speech n -grams, in: the 22nd conference on user modeling, in: Adaptation and Personalization, Aalborg, Denmark, 2014, pp. 243–253.
[78] S. Vargas, P. Castells, Rank and relevance in novelty and diversity metrics for recommender systems, in: Proceedings of the Fifth ACM Conference on Recommender Systems, 2011.
[79] F.S. Sellers, Cruz Campaign Paid $750,000 to “Psychographic Profiling” Company, Washington Post, 2015.
[80] A. Altman, Silent Partners, Time Magazine, 2016.
[81] S. Bertoni, Exclusive interview: how jared kushner won trump the white house, Forbes (2016).
[82] S.H. Thompson, E. Lougheed, Frazzled by facebook? An exploratory study of gender differences in social network communication among undergraduate men and women, Coll. Stud. J. 46 (2012) 88–99.