A Hybrid Strategy for Privacy-Preserving Recommendations for Mobile Shopping

Toon De Pessemier
iMinds-WiCa-Ghent University
G. Crommenlaan 8 box 201
B-9050 Ghent, Belgium
toon.depessemier@ugent.be

Kris Vanhecke
iMinds-WiCa-Ghent University
G. Crommenlaan 8 box 201
B-9050 Ghent, Belgium
kris.vanhecke@ugent.be

Luc Martens
iMinds-WiCa-Ghent University
G. Crommenlaan 8 box 201
B-9050 Ghent, Belgium
luc1.martens@ugent.be

ABSTRACT

To calculate recommendations, recommender systems collect and store huge amounts of users' personal data such as preferences, interaction behavior, or demographic information. If these data are used for other purposes or get into the wrong hands, the privacy of the users can be compromised. Thus, service providers are confronted with the challenge of offering accurate recommendations without the risk of dissemination of sensitive information. This paper presents a hybrid strategy combining collaborative filtering and content-based techniques for mobile shopping with the primary aim of preserving the customer's privacy. Detailed information about the customer, such as the shopping history, is securely stored on the customer's smartphone and locally processed by a content-based recommender. Data of individual shopping sessions, which are sent to the store backend for product association and comparison with similar customers, are unlinkable and anonymous. No uniquely identifying information of the customer is revealed, making it impossible to associate successive shopping sessions at the store backend. Optionally, the customer can disclose demographic data and a rudimentary explicit profile for further personalization.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Information Filtering; K.4.1 [**Computers and Society**]: Public Policy Issues—*Privacy*

Keywords

Recommender System, Shopping Assistant, Privacy, Mobile

1. INTRODUCTION

Data gathering and analysis, i.e. one of the fundamentals of traditional recommender systems, is a serious concern for many, increasingly privacy-aware users. A data collector may disclose personal information to untrusted parties. This could occur either on purpose, i.e., by selling the personal information to a third party, or involuntarily through a security breach. As a result, users are becoming apprehensive about using applications or services that collect personal data. For shopping applications for example, many customers are already concerned with the data collection practices related to loyalty programs [8, 11]. This can be exacerbated when the loyalty program is an application running on the customer's own smartphone. These devices contain a large amount of personal data such as the customer's phone number, e-mail address, and social networking account details.

Despite these privacy concerns, Mobile Shopping Assistants (MSAs) are becoming increasingly popular due to the benefits they offer to both customers and retailers. An MSA can enhance the shopping experience by incorporating features such as loyalty programs, discount vouchers, easy checkout, and various personalized services. The applications are easy and inexpensive to roll out because they can run on the customer's own smartphone. The retailer does not have to invest in specialized hardware and many customers are already familiar with smartphones and the concept of mobile apps. To address these privacy concerns, Put et al. [9] have created inShopnito, a transparent, privacy-preserving MSA that still offers all the features that customers and retailers have come to expect, including a rudimentary recommender system. In this paper, we have extended the MSA with an advanced, hybrid recommendation strategy. Section 3 describes this contribution in detail. As the security and privacy-enhancing technologies used for (anonymous) authentication and transactions have already been described [9], Section 2 of this paper provides only a brief overview of the functionality and the implications of privacy-preserving measures on recommendations.

2. PRIVACY-PRESERVING MOBILE SHOPPING

Preserving the customer's privacy during the usage of inShopnito is of primary importance, which has significant implications for the recommender. At registration time, the customer is issued an Idemix [5] anonymous credential containing attributes with personal information such as name, zip code, or gender. When the customer enters the store, the inShopnito MSA uses the credential to initiate a new shopping session on the store backend system. The customer chooses which attributes (name, zip code, gender, explicit profile) to disclose during this authentication phase. These different levels of privacy provide the customer the necessary

CBRecSys 2014, October 6, 2014, Silicon Valley, CA, USA.

flexibility in the trade-off between privacy and personalization. The backend system knows that the customer has a valid credential, and it knows the content of the attributes that the customer opted to disclose. However, it does not know which particular customer it is dealing with, because no uniquely identifying information is disclosed during authentication. This also means that a customer's successive shopping sessions can not be tied together.

The customer can now proceed to scan products using the camera of her smartphone and add them to the inShopnito shopping cart. During checkout, inShopnito can be used to redeem loyalty points and vouchers to get a discount. To provide the customer with a complete overview of her shopping history, the MSA stores this information securely on the smartphone where it can be used for recommendation purposes. In Section 3, we expand on the recommender components of the inShopnito MSA and backend, and propose a practical solution that preserves the privacy of the customer while still offering advanced personalized services.

Recommender systems initially face the cold start problem, because nothing is known about the user [7]. Usually, a user's actions can be tracked over time. As more information about the user becomes available, the quality of the recommendations increases. With inShopnito, each shopping session is associated with a different, anonymous user identifier. Thus, a server-side recommender system will always have to address the cold start problem, whether it is the customer's first store visit, or her hundredth. Client-side recommenders pose their own set of challenges [3].

Related research [4] into privacy-preserving, personalized ad delivery has proposed a coarse-grained filtering of ads based on the personal information that customers choose to disclose. Subsequently, a further filtering can be performed at client side based on the purchase details stored on the customer's smartphone. Compared to existing solutions, the hybrid recommender strategy of inShopnito goes further than a filtering of information, by analyzing individual shopping carts and comparing them with the purchases of similar customers at the backend.

3. HYBRID RECOMMENDATIONS

The hybrid strategy combines five recommendation approaches. For each approach, preserving the customer's privacy is of crucial importance. Figure 1 provides a schematic overview of these approaches and the data they use.

3.1 Explicit Profile Recommendations

Through an explicit profile on her smartphone, the customer can specify her preferred product categories, as shown in Figure 2(a). These categories, grouping individual products that are typically located in the same section of the store, allow customers to quickly express their interests and filter out irrelevant product groups. Product categories such as pet supplies, garden tools, car/motorcycle supplies, toys, or baby products, are not relevant for every customer. This explicit profile is created automatically based on the customer's purchases, but can be altered at her own discretion. Although the explicit profile contains only category preferences and no details regarding individual products, disclosing the explicit profile is an optional feature for privacy reasons. In addition, customers can opt to disclose some demographic data such as age, municipality, and gender, in order to further filter the product categories such as shaving products, make-up, personal care products, etc.

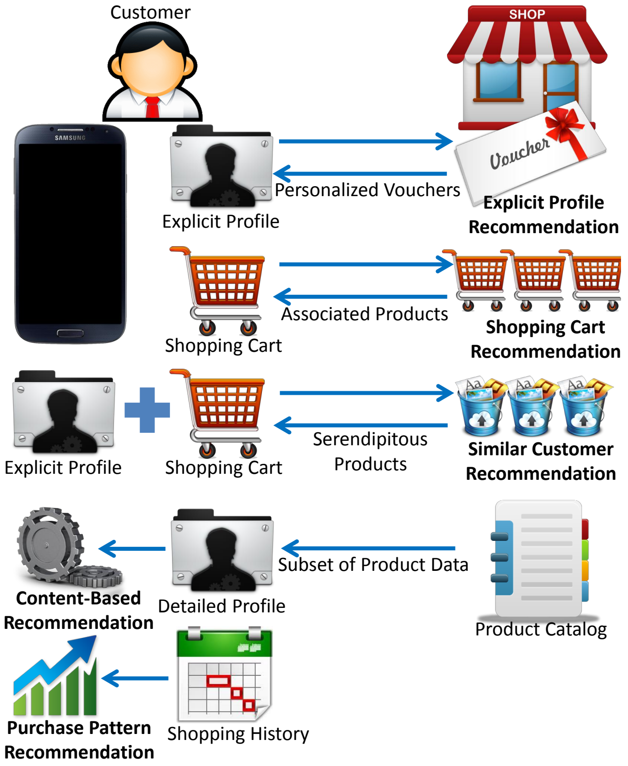

Figure 1: Schematic overview of the different recommendation techniques.

If the customer opts to disclose (parts of) her explicit profile and demographic data, this information is sent to the store backend and used to personalize the *coupons and vouchers* she receives. These targeted vouchers have benefits for the retailers (the vouchers are more effective) as well as for the customers (more relevant vouchers are offered). For privacy reasons, these explicit profiles are only used during the current sessions and removed from the store backend after issuing the vouchers.

3.2 Shopping Cart Recommendations

During every shopping session, the content of the shopping cart is sent to the store backend for analysis. For privacy reasons, no uniquely identifying reference to the customer is stored at server side. Only the content of the individual shopping carts, together with the date of the purchase, are stored. The date provides useful information regarding trends in purchasing behavior, or seasonal products. A detailed timestamp with the exact moment of the purchase (hours/minutes) would have no extra value for the recommender and is omitted because this might induce a privacy risk. If a customer's time of purchase is known (e.g., by observation) and the exact timestamp would be stored, linking the content of the shopping cart to the customer's identity would be possible.

Analysis of the content of the shopping carts of different customers provides insight into the shopping habits of the customers and reveals which products are often bought together. For instance, customers will buy both pasta and bolognese sauce if they intend to prepare spaghetti or lasagna. Product association rules are used to discover which products belong together. Interesting recommendations are products that are not yet added to the shopping cart, but are often bought in combination with the products that are in the shopping cart. So, if bolognese sauce is in the shopping cart, pasta is a good recommendation. More generally, the best recommendation is the product, Y, with the highest probability to be bought, given the current content of the shopping cart, X. Here, X can be a single product or a set of products that the customer wants to buy.

However, highly popular products will always be bought in combination with a large variety of other products, even though no direct link (e.g., a recipe) exists between them. The probability that the customer will buy these popular products is always high, regardless of the content of the shopping cart. In order to take into account the general popularity of products, this probability, $P(X,Y|X)$, is normalized by dividing it by the probability of buying Y, if the content of the shopping cart is different from X. The products with the highest normalized probability are recommended to the customer, as illustrated in Figure 2(b).

$$\operatorname*{Max}_{X \subset Cart} \frac{P(X,Y|X)}{P(!X,Y|!X)} = \operatorname*{Max}_{X \subset Cart} \frac{\frac{P(X,Y)}{P(X)}}{\frac{P(!X,Y)}{P(!X)}} \tag{1}$$

In addition to these automatically derived combinations of products using product association rules, domain knowledge helps to recommend the best matching products. For the shopping cart recommendations, the domain knowledge consists of a set of recipes. If the customer's shopping cart already contains several products that match the ingredients of a certain recipe, the missing ingredients are recommended and the recipe is suggested to try out. Since these recommendations do not require a user profile with an extensive purchase history, they can help to overcome the *cold start problem.*

3.3 Similar Customer Recommendations

Storing the customers' individual consumption behavior on a central server induces a privacy risk and is therefore undesirable. With inShopnito, each shopping session has a different, anonymous user identifier, and successive shopping sessions cannot be linked (Section 2). Because collaborative filtering is based on calculating the similarity between the historical consumption behavior of individual users (or products), a traditional collaborative filtering approach is not possible in this situation.

As an alternative, customers are compared based on their explicit profile, which is voluntarily disclosed and contains only data about product categories but not of individual product purchases. Calculating the similarity between customers based on their explicit profile might be less accurate than based on their complete consumption behavior; but this approach induces no privacy risk. Based on this explicit shopping profile, customers are partitioned into groups of similar customers, just as the neighborhoods of similar users in the traditional collaborative filtering approach. Each group is represented by a bucket that contains all products that have been bought by the customers of that group.

After every visit to the store, the content of the customer's shopping cart is added to the bucket of the customer's group. If two customers have a similar explicit profile, their shopping carts will end up in the same bucket. Since the bucket contains only product information and no link to the identity of the customer, the purchasing history of an individual customer cannot be deduced if the bucket groups purchases of many customers. Analysis of the products in the bucket of the customer allows to generate recommendations based on what people who like similar products have bought in the past. The products that are most popular with other customers of the group are recommended, with the exception of products that are already in the shopping cart of the customer. The popularity of products within a group is normalized with respect to the general popularity of a product. These recommendations, based on the purchases of similar customers, aim to offer more *serendipitous recommendations* to the customers, just as collaborative filtering algorithms do.

This recommendation technique can also be combined with the approach that compares shopping carts (Section 3.2). Individual shopping carts (without a reference to the customer's identity) can be stored per group of similar customers, as defined by the explicit profile. Subsequently, product association rules can be applied on the groups of similar customers, instead of on the complete population of customers. This partitioning of customers according to their preferences can help to refine the product association rules.

3.4 Content-based Recommendations

For privacy reasons, detailed historical information about purchases cannot leave the secured environment of the customer's smartphone. As a result, these detailed purchase data can only be exploited if the recommendation algorithm runs on the customer's smartphone. In this customer-centric personalization approach [1], each user has its own mobile recommendation engine. Storing this detailed user profile securely on the smartphone also has advantages. For instance, the user profile can be shared amongst different shops without the privacy risk that one retailer abuses these purchase data for commercial profits.

Since only purchase data of the target user (i.e. the user for who recommendations are calculated) are available on the smartphone, a content-based recommendation algorithm is the most worthwhile solution to process this detailed profile.

Content-based recommendation algorithms determine the products that best match the user's profile, based on a description of the product characteristics [6]. Since the detailed profile cannot leave the customer's smartphone, the product descriptions have to be transferred to the smartphone for comparison with the detailed profile. However, the complete product catalog of the store and the corresponding descriptions can be quite extensive for processing on a smartphone. Therefore, only products and descriptions of categories that are relevant for the customer are sent to the smartphone to reduce the data traffic. Recommendations for pet supplies may be irrelevant for customers who have never bought any pet supplies in the past. They may not have pets, or buy their supplies through other channels. The explicit profile is used to determine which categories are relevant and have to be considered. The resulting subset of the product catalog has to be downloaded only once, the first time that the customer visits the store. From then on, updates of the catalog are sufficient to keep track of new products, changed descriptions, and products that are not

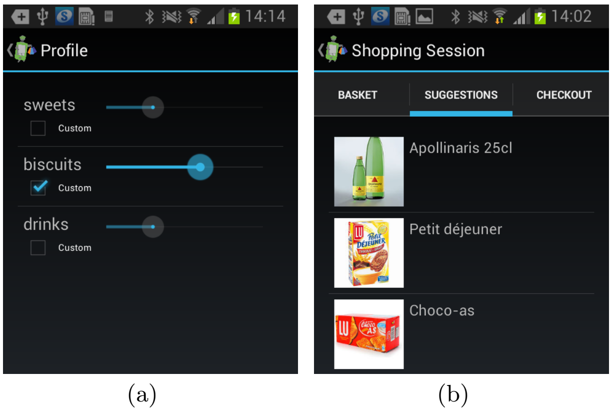

Figure 2: Screenshots of the mobile application: (a) the explicit user profile (b) the personalized suggestions.

available anymore.

Different types of content-based recommendation algorithms can be used, but with the limitation that the computational requirements must fit within the available resources of the smartphone. Our approach uses the InterestLMS algorithm of the Duine recommender framework [10]. Content-based algorithms often suffer from over-specialization [7], since they recommend only products similar to those already bought by the customers. In certain application domains, items should not be recommended if they are too similar to something the user has already seen, such as a different news article describing the same event. For shops however, various situations exist in which customers are interested in similar products: cheaper or discounted products of a different brand, new or similar food products to replenish their house stock, or alternatives for products that are out of stock.

3.5 Purchase Pattern Recommendations

The last type of recommendations focuses on the repetitive purchase behavior of customers [2]. Specific products, such as toothpaste or coffee, are used on a regular basis, and as a result, need to be replenished regularly. Patterns in the purchase behavior can be detected, and used to predict the next purchase of a certain product. E.g., if one tube of toothpaste is bought every month, predicting the next purchase of toothpaste is obvious.

Based on the shopping history (i.e. the time and amount of the last purchase), the recommender estimates if the customer needs to buy a certain product. If this is the case, and the customer has not yet added the product to the shopping cart, it will be recommended. So, the aim of the purchase pattern recommendations is to *remind customers* to buy products that they might forget but probably need because of their *repetitive consumption behavior*.

4. CONCLUSIONS

The growing importance of privacy in online services emphasizes the need for privacy-preserving recommender systems, not the least in the domain of shopping. Traditional collaborative filtering algorithms, which rely on a central storage and comparison of detailed user profiles, may induce a privacy risk. But limiting the disclosed customer data introduces a trade-off between the accuracy of the recommendations and the privacy of the customer. Therefore, we present a privacy-preserving, hybrid strategy that combines client-side and server-side recommendation techniques. At server-side, the recommender is based on information that customers opt to disclose, and performs an analysis of the shopping cart using product association rules, and a comparison with the shopping carts of similar customers. At client-side, detailed customer information is used for content-based recommendations and suggestions based on purchase patterns.

5. ACKNOWLEDGMENTS

This research was funded by the IWT-SBO Project MobCom: A Mobile Companion (https://www.mobcom.org)

6. REFERENCES

[1] G. Adomavicius, Z. Huang, and A. Tuzhilin. Personalization and recommender systems. *Tutorials in Operations Research, Informs*, pages 55–107, 2008.

[2] H. Baumgartner. Repetitive purchase behavior. In A. Diamantopoulos, W. Fritz, and L. Hildebrandt, editors, *Quantitative Marketing and Marketing Management*, pages 269–286. Gabler Verlag, 2012.

[3] L. N. Cassel and U. Wolz. Client side personalization. In *DELOS Workshop: Personalisation and Recommender Systems in Digital Libraries*, pages 8–12, 2001.

[4] M. Hardt and S. Nath. Privacy-aware personalization for mobile advertising. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, CCS '12, pages 662–673, New York, NY, USA, 2012. ACM.

[5] IBM Research Security Team. Specification of the Identity Mixer Cryptographic Library v. 2.3.4. Technical report, 2012.

[6] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich. *Recommender Systems: An Introduction*. Cambridge University Press, New York, NY, USA, 1st edition, 2010.

[7] P. Lops, M. Gemmis, and G. Semeraro. Content-based recommender systems: State of the art and trends. In F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, editors, *Recommender Systems Handbook*, pages 73–105. Springer US, 2011.

[8] V. Pez. Negative effects of loyalty programs: An empirical investigation on the french mobile phone sector. Technical report, Université Paris-Dauphine, 2007.

[9] A. Put, I. Dacosta, M. Milutinovic, B. De Decker, S. Seys, F. Boukayoua, V. Naessens, K. Vanhecke, T. De Pessemier, and L. Martens. inshopnito: An advanced yet privacy-friendly mobile shopping application. In *Proceedings of the IEEE 10th World Congress on Services (SERVICES 2014)*. IEEE, 2014.

[10] Telematica Instituut / Novay. Duine Framework, 2009. Online available at `http://duineframework.org/`.

[11] S. Worthington and J. Fear. The hidden side of loyalty card programs. *The Austalian centre for retail studies*, 2009.