Pervasive and Mobile Computing 25 (2016) 125–142

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

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

A machine-learning based approach to privacy-aware information-sharing in mobile social networks[☆]

Igor Bilogrevic[a,*,1], Kévin Huguenin[c,1], Berker Agir[b], Murtuza Jadliwala[d], Maria Gazaki[b], Jean-Pierre Hubaux[b]

[a] Google, 8002 Zurich, Switzerland

[b] School of Computer and Communication Systems, EPFL, 1015 Lausanne, Switzerland

[c] LAAS-CNRS, 31400 Toulouse, France

[d] Department of Electrical Engineering and Computer Science, Wichita State University, Wichita, KS 67260, USA

ARTICLE INFO

Article history:
Received 30 June 2014
Received in revised form 18 December 2014
Accepted 23 January 2015
Available online 31 January 2015

Keywords:
Information-sharing
Decision-making
Machine learning
User study
Privacy

ABSTRACT

Contextual information about users is increasingly shared on mobile social networks. Examples of such information include users' locations, events, activities, and the co-presence of others in proximity. When disclosing personal information, users take into account several factors to balance privacy, utility and convenience — they want to share the "right" amount and type of information at each time, thus revealing a selective sharing behavior depending on the context, with a minimum amount of user interaction. In this article, we present SPISM, a novel information-sharing system that decides (semi-)automatically, based on personal and contextual features, whether to share information with others and at what granularity, whenever it is requested. SPISM makes use of (active) machine-learning techniques, including cost-sensitive multi-class classifiers based on support vector machines. SPISM provides both ease of use and privacy features: It adapts to each user's behavior and predicts the level of detail for each sharing decision. Based on a personalized survey about information sharing, which involves 70 participants, our results provide insight into the most influential features behind a sharing decision, the reasons users share different types of information and their confidence in such decisions. We show that SPISM outperforms other kinds of policies; it achieves a median proportion of correct sharing decisions of 72% (after only 40 manual decisions). We also show that SPISM can be optimized to gracefully balance utility and privacy, but at the cost of a slight decrease in accuracy. Finally, we assess the potential of a one-size-fits-all version of SPISM.

1. Introduction

Mobile devices are becoming the gatekeepers of people's digital selves. More and more personal and private information is stored, shared and managed on-the-go. Having access to people's personal data and physical contexts (through an

[☆] This article is a revised and extended version of a paper that appears in the Proceedings of the 15th ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2013) Bilogrevic et al. (2013).

* Corresponding author.

E-mail addresses: igor.bilogrevic@gmail.com (I. Bilogrevic), kevin.huguenin@laas.fr (K. Huguenin), berker.agir@epfl.ch (B. Agir), murtuza.jadliwala@wichita.edu (M. Jadliwala), maria.gazaki@epfl.ch (M. Gazaki), jean-pierre.hubaux@epfl.ch (J.-P. Hubaux).

[1] Parts of this work were carried out while the author was with EPFL.

http://dx.doi.org/10.1016/j.pmcj.2015.01.006

increasing number of embedded sensors), mobile devices represent a simple means to quickly share information with others via mobile social networks, such as Facebook, WhatsApp or Google, without the need for manually typing their current locations; location and photos are just two examples of data that can be easily shared. In addition to the user-triggered sharing decisions, applications such as Foursquare and the now-closed Gowalla enable users to configure their smartphones to share their location and co-presence automatically, in a push-based fashion. With a small set of default information-sharing policies, users have the possibility to adjust the settings in order to match their sharing behaviors with their privacy concerns.

Usually, there are several behavioral and contextual factors that influence users when they share their personal information, as extensively shown in [1–5]. By analyzing people's sharing behaviors in different contexts, it is shown in these works that it is possible to determine the features that most influence users' sharing decisions, such as the identity of the person that is requesting the information and the current location [2]. For instance, tools such as the location-sharing systems Locaccino [6] and PeopleFinder [4] have been used to gain significant insight into the benefits of providing users with the ability to set personal sharing policies. Two recurrent findings in studies related to information sharing are that (i) users are not particularly good at effectively articulating their information-sharing policies (compared to their actual behavior) [4] and (ii) that sharing policies evolve over time [6,4].

In order to overcome these two issues, machine-learning techniques have been applied to improve to some extent the decision-making process [7,8,4,9]. The advantage of such systems is that they can decide, in a semi-automatic fashion, whether or not to share information. Most existing schemes, however, enable users to share only a specific kind of information (e.g., location). Moreover, they only make binary decisions on whether to share the requested information. This last issue in particular, is often mentioned as a crucial catalyst for overcoming concerns related to privacy [10,11] and to a more open, sharing behavior.

In our work, we perform a comprehensive study of information-sharing in mobile social networks, by tackling, all at once, the issues related to context, user-burden and privacy trade-offs. We introduce SPISM (for Smart Privacy-aware Information Sharing Mechanism), a novel *pull-based* information-sharing system (i.e., users explicitly request information from their friends) implemented on Android; it decides in a semi-automatic fashion, whether or not to share information and the level of detail of the information to be shared with other users or services, based on personal and contextual features and past behavior. SPISM makes use of a cost-sensitive multi-class classifier, typically based on naive Bayes (NB) or non-linear Support Vector Machines (SVM) with Gaussian or polynomial kernels, fed with different contextual features including the time of day, the current location of the user and the identity of the requester. The cost-sensitive aspects enable SPISM to find an appropriate balance between over-sharing (i.e., unduly sharing the requested information) and under-sharing (i.e., unduly retaining the requested information). The multi-class aspects of the classifier enable SPISM to decide the correct level of detail of the shared information (e.g., whether to share the name of the street or the city the user is in). The decision-making core is supported by an active learning method that enables SPISM to either decide automatically – whenever the confidence in the decision is high enough (based on the entropy of the distribution, computed over the possible decisions of the classifier) – or to rely on the user's input otherwise. As such, SPISM continuously learns from the user and, over time, it requires less and less user-input. Note that, in this paper, we do not develop new machine learning techniques. Instead, we leverage on appropriate existing machine learning techniques to provide useful features for protecting the users' privacy while reducing their burden. As such, the main contribution of this paper is the design and implementation of a feature-rich decision-making mechanism and the evaluation of its performance, in addition to the analysis of users' sharing behaviors, based on a novel and rich dataset. SPISM can work with any existing (mobile) social networks and could even be used transparently by users, as it can operate at the operating-system level, filtering all requests for personal information from mobile apps and websites and replying according to the user's behavior.

The contributions of this work are as follows. First, we develop a novel information-sharing system (SPISM) for (semi-) automatic decision-making in mobile social networks: It enables users to share different types of information (location, activity and co-presence of other people) with other users or services in a privacy-aware fashion. Second, we conduct a personalized online study which involves 70 participants where, in addition to collecting data about their sharing behaviors, we provide insights into two other crucial factors in studies related to information sharing [12]: The *reason* behind a decision to share and the confidence that the user has in her decision. Third, we evaluate SPISM with respect to the amount of training data (provided by the user) and its performance, and we compare it against two policy-based mechanisms. Our results show that SPISM significantly outperforms individual privacy policies specified by users, and it achieves a median proportion of correct binary sharing decisions (i.e., whether to share the requested information) of 72% when trained on only 40 manual decisions. Moreover, SPISM is able to infer the level of details at which the information should be shared (e.g., street-level accuracy vs. city-level accuracy for location information) with a median proportion of correct decision of 60%. We also demonstrate the advantages of active learning techniques in our setting. Fourth, in this substantially extended article (as compared to [13]), we enrich SPISM with a new functionality that significantly enhances the privacy of the users while incurring a negligible decrease of performance and no side-effects on the usability. We also assess the potential of a one-size-fits-all model for decision making.

The rest of the paper is organized as follows. We survey the related work in Section 2, and we present the SPISM information-sharing platform in Section 3. We describe the survey methodology and discuss the participants in Section 4. We present the performance results in Section 5, and we conclude this article in Section 6.