Towards Ubiquitous Privacy Decision Support: Machine Prediction of Privacy Decisions in IoT

Hosub Lee & Alfred Kobsa

Smartphone apps and websites increasingly ask users to make privacy decisions about personal information disclosure, e.g., to grant or deny an app permission to access their location or their phone book. However, previous research indicates that people are often unable to make these decisions due to limits in their available time, motivation, and abilities to fully understand the tradeoff between utility and privacy in the disclosure of personal information. This problem will continue to grow in ubiquitous computing environments like the Internet of Things (IoT), as an array of sensors around the user unobtrusively collects his/her personal information [1]. Clearly, rich personal information helps IoT systems better understand users, thus providing better-tailored services to them. At the same time, however, this leads to considerable increase in privacy concerns that may lead users to stop using the service [4, 11, 9]. Therefore, providing IoT services with minimized privacy risks is crucial for both protecting users' privacy and keeping IoT ecosystems sustainable. One possible way to achieve this objective is to assist users with making better privacy decisions, by predicting decisions based on their historical decision-making behavior and recommending privacy settings accordingly (i.e., privacy decision support).

In this chapter, we proposed a novel machine learning mechanism for predicting privacy decisions of users in IoT environments. The aim of this mechanism is to correctly predict users' decisions whether or not to allow the given personal information monitoring based on both the user's current context and personal attitudes on privacy. We tested the proposed mechanism on a privacy-related behavioral dataset collected from 172 users who were presented with descriptions of personal information tracking scenarios relating to their physical location on a university campus [6]. These scenarios are defined by five different contextual parameters such as place (*where*), type of collected information (*what*), entity (*who*), purpose (*reason*), and frequency (*persistence*) of the monitoring. The dataset also contains each user's privacy decision behavior on each scenario in terms of five reaction parameters such as willingness to be notified (*_notification*), willingness to allow (*_permission*), and subjective evaluations of comfort, risk, and appropriateness (*_comfort*, *_risk*, and *_appropriateness*) of the monitoring. We treated the five contextual parameters as basic features which collaboratively represent the current context in which information monitoring is performed by IoT devices. Additionally, we assigned each user to a specific privacy perception segment based on a small portion of their data (i.e., prior privacy decisions). We then used this privacy segment information as an additional feature. We used the reaction parameter *_permission* as target value (or label), since it best reflects users' substantive privacy decisions in IoT environments (namely, allow or reject the monitoring). The dataset contains 6,618 data points, and each data point (row) represents a user's privacy decisions in a specific IoT scenario.

First, we utilized a state-of-the-art machine learning model, called linear model and deep neural networks (LMDNN, [2]), to make privacy decisions on behalf of users. LMDNN, which is also known as Wide & Deep Learning, jointly trains wide linear models and deep neural networks. By doing so, it can take the benefits of memorization (linear models) and generalization (neural networks) at the same time.

LMDNN showed a remarkable performance on binary classification problems with sparse input features [2, 10]. Because the dataset collected in [6] is also composed of categorical data with many possible feature values (e.g., 24 values for the contextual parameter *what*), we decided to use LMDNN to build predictive models for privacy decision support in IoT. We also selected machine learning models that have been widely used in the literature (e.g., decision trees) and compared them with LMDNN in terms of predictive performance on the dataset. Experimental results indicated that LMDNN outperforms all the conventional models.

Next, we explored the most suitable combination of features for building LMDNN models with a reasonable predictive performance. We chose the five contextual parameters as basic features because these parameters are known to be related to users' privacy decisions in IoT [3, 5, 6]. In addition, we considered each user's privacy segment information as an additional feature. This is because previous research indicates that privacy segment information is helpful for machine learning models to better predict users' privacy decisions [7, 8]. We applied an unsupervised data clustering algorithm, K-modes clustering, on a subset of users' privacy decisions, in order to segment the users by their perceptions of privacy (i.e., privacy segmentation). Therefore, we tested the following feature combinations: (1) contextual parameters only (basic features), (2) contextual parameters with interactions between them, and (3) contextual parameters with interactions between them and privacy segment information. Experimental results showed that the feature combination (3) gives LMDNN models with the highest predictive performance. It also means that both interactions between contextual factors (e.g., *who* by *what*) and privacy segment information are useful for LMDNN models to predict privacy decisions of the users.

Lastly, we investigated different approaches for training machine learning models. There exist two traditional approaches in the literature: individual and one-size-fits-all modeling. Individual modeling is a process of building a user-specific predictive model based on each user's data only. This approach is known to be effective for modeling each user's unique characteristics (e.g., habits and personality). A model with a reasonable predictive performance, however, typically requires a considerable amount of training data from each individual user. In contrast, one-size-fits-all modeling utilizes multiple users' data as a single training data and constructs a universal model for all of them (including new users). This approach enables predictive models to make general predictions, therefore it can be useful for new users who did not provide data to the system yet, but want to get recommendations immediately. However, prediction results may not be personalized to each individual user. We present another approach called one-size-fits-segment modeling, a variant of one-size-fits-all, taking privacy segment information into account in building predictive models. The basic idea is to serve each user with a machine learning model trained by data collected from others who share the same notion of privacy with this user. We divided the dataset based on the results of privacy segmentation, then trained an LMDNN model for each segment of the users. By using both contextual and privacy segment information as input features, we compared the predictive performance of individual, one-size-fits-all, and one-size-fits-segment modeling. Experimental results confirmed that the proposed approach performs the best in the dataset. Final LMDNN models trained via one-size-fits-segment modeling showed an average area under curve (AUC) of 0.6782 across all users. We noticed that one-size-fits-segment modeling performs much better than individual modeling for about 80% of the users.

To sum up, our proposed prediction mechanism not only showed a reasonable performance for most users, but also can cause less burden and privacy risks to users since it mainly utilizes non-personal contextual information which can often be automatically collected from the IoT environment, and only prompts each user for a small amount of privacy decisions (answers to five reaction parameters about a single scenario) to determine his/her privacy segment. We also presented some practical implications for designing and developing machine learning-based privacy decision support systems for IoT.

References

[1] Luigi Atzori, Antonio Iera, and Giacomo Morabito. The Internet of Things: A survey. *Computer Networks*, 54(15):2787–2805, 2010.

[2] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, et al. Wide & deep learning for recommender systems. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*, pages 7–10. ACM, 2016.

[3] Richard Chow, Serge Egelman, Raghudeep Kannavara, Hosub Lee, Suyash Misra, and Edward Wang. HCI in business: A collaboration with academia in IoT privacy. In *International Conference on HCI in Business*, pages 679–687. Springer, 2015.

[4] Delphine Christin, Andreas Reinhardt, Salil S Kanhere, and Matthias Hollick. A survey on privacy in mobile participatory sensing applications. *Journal of Systems and Software*, 84(11):1928–1946, 2011.

[5] Hosub Lee and Alfred Kobsa. Understanding user privacy in Internet of Things environments. In *Internet of Things (WF-IoT), 2016 IEEE 3rd World Forum on*, pages 407–412. IEEE, 2016.

[6] Hosub Lee and Alfred Kobsa. Privacy preference modeling and prediction in a simulated campuswide IoT environment. In *Pervasive Computing and Communications (PerCom), 2017 IEEE International Conference on*, pages 276–285. IEEE, 2017.

[7] Jialiu Lin, Bin Liu, Norman Sadeh, and Jason I Hong. Modeling users' mobile app privacy preferences: Restoring usability in a sea of permission settings. In *Proceedings of the 10th Symposium on Usable Privacy and Security (SOUPS 2014)*, pages 199–212, 2014.

[8] Bin Liu, Mads Schaarup Andersen, Florian Schaub, Hazim Almuhimedi, SA Zhang, Norman Sadeh, Alessandro Acquisti, and Yuvraj Agarwal. Follow my recommendations: A personalized privacy assistant for mobile app permissions. In *Proceedings of the 12th Symposium on Usable Privacy and Security (SOUPS 2016)*, pages 27–41, 2016.

[9] Charith Perera, Rajiv Ranjan, Lizhe Wang, Samee U Khan, and Albert Y Zomaya. Big data privacy in the Internet of Things era. *IT Professional*, 17(3):32–39, 2015.

[10] Shaoyun Shi, Min Zhang, Hongyu Lu, Yiqun Liu, and Shaopin Ma. Wide & deep learning in job recommendation: An empirical study. In *Asia Information Retrieval Symposium*, pages 112–124. Springer, 2017.

[11] Sabrina Sicari, Alessandra Rizzardi, Luigi Alfredo Grieco, and Alberto Coen-Porisini. Security, privacy and trust in Internet of Things: The road ahead. *Computer Networks*, 76:146–164, 2015.