TOWARDS EXPLAINABLE MACHINE LEARNING FOR PRIVACY-AWARE LOCATION TRACKING

Benjamin Baron, University College London
Mirco Musolesi, University College London and The Alan Turing Institute

Keywords: Interpretable Machine Learning, Pervasive Systems, Privacy, Digital Traces, Location Information.

With the emergence of connected devices (*e.g.*, smartphones and smartmeters), pervasive systems generate growing amounts of digital traces as users undergo their everyday activities. These traces are crucial to service providers to understand their customers, to increase the degree of personalization, and enhance the quality of their services. For instance, personal digital traces stemming from public transit smartcards help transportation providers understand the commuting patterns of users; the usage statistics of home appliances can be used to improve energy efficiency; bank transaction logs can be used to spot unusual activity in accounts.

However, sharing these digital traces generated by pervasive systems with service providers might raise concerns with regards to user privacy, as the processing and analysis of these traces can surface latent information about user behaviors. Using machine learning techniques, third parties such as advertisers can identify a single individual from inadequately aggregated datasets shared by service providers either publicly or privately. The main focus of the existing work has been on the performance and interpretability of the techniques to infer personal information and identify users from their digital traces. For instance, Kosinski *et al.* were able to infer personal attributes from your likes on Facebook [1], de Montjoye *et al.* were able to uniquely identify individuals from their credit card spending [2], Lisovich *et al.* were able to infer personal information about in-home activities from smartmeter traces [3], and El Mahrsi *et al.* were able to extract socio-economic information about passengers from their public transportation smartcard usage [4].

However, little interest has been shown to explain how these techniques can infringe the user privacy, given the nature of the inference itself [5]. In this context, we have conducted a user research study, which aims at exploring the privacy expectations of individuals when it comes to tracking their location and inferring personal information about them. In particular, we have developed a mobile application called TrackingAdvisor that infers the participants' significant places and the information about them from the places they visit [6]. We further provide users with interpretations of the personal information inference so that they can be aware of the privacy risks related to location tracking and better understand how to protect their privacy.

References

[1] M. Kosinski, D. Stillwell, and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," *Proceedings of the National Academy of Sciences*, vol. 110, no. 15, pp. 5802–5805, 2013.

[2] Y.-A. De Montjoye, L. Radaelli, V. K. Singh, *et al.*, "Unique in the shopping mall: On the reidentifiability of credit card metadata," *Science*, vol. 347, no. 6221, pp. 536–539, 2015.

[3] M. A. Lisovich, D. K. Mulligan, and S. B. Wicker, "Inferring personal information from demand-response systems," *IEEE Security & Privacy*, vol. 8, no. 1, 2010.

[4] M. K. El Mahrsi, E. Come, J. Baro, and L. Oukhellou, "Understanding Passenger Patterns in Public Transit Through Smart Card and Socioeconomic Data: A case study in Rennes, France," in *ACM UrbComp*, (New York, NY, USA), p. 9, Aug. 2014.

[5] B. Baron and M. Musolesi, "Interpretable machine learning for privacy-preserving iot and pervasive systems," *arXiv preprint arXiv:1710.08464*, 2017.

[6] TrackingAdvisor. Online: `https://itunes.apple.com/gb/app/trackingadvisor/id1353081290`.