Constrained-Based Differential Privacy

Ferdinando Fioretto
Syracuse University, NY, USA

Abstract

Data sets and statistics about groups of individuals are increasingly collected and released, feeding many optimization and learning algorithms. In many cases, the released data contain sensitive information whose privacy is strictly regulated. For example, in the U.S., the census data is regulated under Title 13, which requires that no individual be identified from any data released by the Census Bureau. In Europe, data release is regulated according to the General Data Protection Regulation, which addresses the control and transfer of personal data.

Differential privacy [1] has emerged as the de-facto standard to protect data privacy. In a nutshell, differentially private algorithms protect an individual's data by injecting random noise into the output of a computation that involves such data. While this process ensures privacy, it also impacts the quality of data analysis, and, when private data sets are used as inputs to complex machine learning or optimization tasks, they may produce results that are fundamentally different from those obtained on the original data and even rise unintended bias and fairness concerns.

In this talk, I will first focus on the challenge of releasing privacy-preserving data sets for complex data analysis tasks. I will introduce the notion of *Constrained-based Differential Privacy* (C-DP), which allows casting the data release problem to an optimization problem whose goal is to preserve the salient features of the original data. I will review several applications of C-DP in the context of very large hierarchical census data [3], data streams [2], energy systems [4], and in the design of federated data-sharing protocols. Next, I will discuss how errors induced by differential privacy algorithms may propagate within a decision problem causing biases and fairness issues [5, 6]. This is particularly important as privacy-preserving data is often used for critical decision processes, including the allocation of funds and benefits to states and jurisdictions, which ideally should be fair and unbiased. Finally, I will conclude with a roadmap to future work and some open questions.

2012 ACM Subject Classification Security and privacy → Privacy-preserving protocols; Computing methodologies → Artificial intelligence; Computing methodologies → Optimization algorithms

Keywords and phrases Optimization, Differential Privacy, Fairness

Digital Object Identifier 10.4230/LIPIcs.CP.2021.2

Category Invited Talk

Funding This research is partially supported by NSF grant 2133169. Its views and conclusions are those of the authors only.

References

1. Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284, 2006.
2. Ferdinando Fioretto and Pascal Van Hentenryck. Optstream: Releasing time series privately. *Journal of Artificial Intelligence Research*, 65:423–456, 2019.
3. Ferdinando Fioretto, Pascal Van Hentenryck, and Keyu Zhu. Differential privacy of hierarchical census data: An optimization approach. *Artificial Intelligence*, pages 639–655, 2021.
4. Terrence W.K. Mak, Ferdinando Fioretto, Lyndon Shi, and Pascal Van Hentenryck. Privacy-preserving power system obfuscation: A bilevel optimization approach. *IEEE Transactions on Power Systems*, 35(2):1627–1637, March 2020. `doi:10.1109/TPWRS.2019.2945069`.
5. Cuong Tran, Ferdinando Fioretto, Pascal Van Hentenryck, and Zhiyan Yao. Decision making with differential privacy under the fairness lens. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, page (to appear), 2021.
6. Keyu Zhu, Pascal Van Hentenryck, and Ferdinando Fioretto. Bias and variance of post-processing in differential privacy. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 11177–11184, 2021.

27th International Conference on Principles and Practice of Constraint Programming (CP 2021).
Editor: Laurent D. Michel; Article No. 2; pp. 2:1–2:1
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany