Incentive Mechanisms for Data Privacy Preservation and Pricing

Doctoral Consortium

Mengxiao Zhang
The University of Auckland
Auckland, New Zealand
mengxiao.zhang@auckland.ac.nz

ABSTRACT

The advent of data marketplaces and the increasing awareness of data privacy call for data privacy pricing mechanisms which can set prices of data properly while assuring privacy preservation. Moreover, the trade-off between the privacy of data and the accuracy of results need to be considered. Additionally, data marketplaces have different structures, including buy-sided market, sell-sided market, two-sided market, and two-sided platform. I will explore data privacy pricing mechanisms under these four structures.

KEYWORDS

Data marketplaces, data pricing, mechanism design, privacy

ACM Reference Format:
Mengxiao Zhang. 2020. Incentive Mechanisms for Data Privacy Preservation and Pricing. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020,* IFAAMAS, 3 pages.

1 INTRODUCTION

Data is being produced at unprecedented rates. Modern computing and communications infrastructure allow to capture, transport, and store data in ways not seen before. Such technological progress gives rise to the emergence of data marketplaces, which can be seen as platforms that facilitate data trade by bringing together data providers, data consumers, analysts, and application developers.

As data become easy to access, there is an increasing public concern about protecting its privacy aspects. The recent privacy breach of millions of Facebook users is highlighting the importance of privacy preservation and is raising the awareness of the value of personal data. Consequently, it is expected that data providers start to demand compensation as a result of their privacy loss when their data are utilised. Furthermore, data market owners need to be conscious of the privacy cost and privacy protection when collecting and selling data.

Data pricing problems have attracted many scholars' attention (e.g. [3, 4, 7, 8]) . In the setting of conventional mechanism design, individual rationality and incentive compatibility (or truthfulness) are identified as desirable properties of a mechanism. When privacy is taken into consideration, more properties need to be explored, such as privacy, and accuracy.

Moving beyond data, as mentioned above, the institution that facilitates the exchange of data, the data marketplace, is set up following a range of different data marketplaces structures. For instance, Datacoup works as a data broker, purchasing data from individuals and selling data to data consumers and making a profit from data trade. Alternatively, Datum plays a role of platform provider, gathering two sides of data transactions but not getting involved in data transactions. These two contexts need to be considered separately since the goals are different. The goal of data brokers is to minimise the costs of data collection and to maximise the profit generated from data consumers. The aim of platform providers is to profit from attracting more participants on the sides of the platform. I will explore pricing mechanisms in these two contexts.

This proposal proceeds as follows: Section 2 reviews relevant literature, which leads to Section 3, research questions, followed by research design. At last, the conclusion is provided.

2 RELATED WORK

2.1 Differentially Private Mechanisms

To measure the extent to which privacy is preserved by a mechanism, Dwork et al. [2] propose the notion of ε-differential privacy, where ε is a non-negative number known as *privacy parameter*. Building upon this idea, two differentially private mechanisms are proposed, namely, the Laplace mechanism [2] and the exponential mechanism [7].

However, when an ε-differentially private mechanism is applied to a dataset representing a group of individuals I with heterogeneous requirements ε_i, to satisfy the privacy requirement from every one, ε has to be set as the smallest value among all requirements, i.e., $\varepsilon = \min_{i \in I} \varepsilon_i$. It means that the mechanism adds more noise than is necessary for some data providers. Considering this, Jorgensen, Yu, and Cormode [5] and Alaggan et al. [1] propose a new privacy concept called personalised differential privacy. Four existing mechanisms can achieve personalised differential privacy, including the stretching model [1], Sampling mechanism, $\mathcal{PE}$ mechanism [5] and the partitioning-based mechanism [6].

2.2 Mechanisms for privacy pricing and protection

The transaction on data is either a purchase or a sale. For a sale, it happens between a broker with data for sale and data consumers with heterogeneous valuations on private data. When setting prices of data, the broker wants to maximise each consumer's payment so as to gain the maximum revenue, while consumers are reluctant to do so since the lower price would lead to a higher surplus. Therefore, a mechanism designed to encourage consumers to reveal their true valuation is needed. McSherry and Talwar [7] and Riederer et al. [8] propose forward auction to price data. On the contrary, some researchers considered the side of data providers. There is

Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, G. Sukthankar (eds.), May 9–13, 2020, Auckland, New Zealand.

a monopolistic broker willing to purchase data and many data providers owing data for sale. The goal of the broker is to minimise the total payment for each provider. In other words, s/he wants to reduce the price to the minimum amount that the providers can tolerate. Ghosh and Roth [4] and Fleischer and Lyu [3] design reverse auctions to achieve this goal.

3 RESEARCH QUESTIONS

The emergence of data marketplaces and considerable literature manifest that the problem of private data pricing is of practical and theoretical significance. Also, the increasing privacy awareness leads to the privacy protection requirements from public. This study aims to design incentive mechanisms that set prices for private data while preserving privacy.

The consulted literature identifies desirable properties that a data pricing scheme should satisfy. The most relevant of these properties to the present proposal follow.

1. Differential privacy and personalised differential privacy. In the process of magnetisation of private data, the public concern about privacy preservation and the requirement for quantification of privacy arise. When an ε-differentially private mechanism is applied, ε has to be the most stringent requirement among all data providers, which means the mechanism adds more noise than is necessary for some data providers. Considering this, $\vec{\varepsilon}$-personalised differential privacy is proposed, where $\vec{\varepsilon} = \{\varepsilon_1, \ldots, \varepsilon_n\}$ is the collection of specified privacy requirement ε_i of each data provider.
2. Accuracy. On the other hand, privacy is not what data consumers care about. Instead, they care about whether the estimates derived from datasets are accurate. Given that, a broker needs to ensure the mechanism can satisfy their accuracy requirements. There is a natural tension between privacy and distortion. Better privacy is preserved, more noise is added, and less accuracy can be achieved.
3. Payment minimisation or revenue maximisation. In most cases, to guarantee more accurate results, more accurate data and more individuals are needed, which will definitely increase the overall payment. There is a trade-off between the accuracy and the payment. A data broker needs mechanisms which can minimise overall payment. On the other side, a broker aims to generate the maximum revenue when selling data.
4. Truthfulness. The valuation of private data is known by data consumers or data providers themselves, but hidden from anyone else, which is known as information asymmetry. However, the information about true valuation is helpful to increase market efficiency. Hence, mechanisms towards the revelation of private information are needed.
5. Individual rationality. Both data providers and data consumers should have incentives to participate in the mechanism. In other words, the surplus of involving the data transactions should not be negative, and the mechanism which satisfies this requirement is individually rational.

Data marketplaces adopt different models. To be specific, the market structure, according to [9], is known as:

- Buy-sided market, where a data broker collects data from multiple data providers and compensates them accordingly.
- Sell-sided market. Sell-side market deals with the data transactions between a data broker and data consumers.
- Two-sided market. In such a model, all data providers and data consumers trade data through a data broker. The profit of the data broker is the difference between the revenue generated from data selling and the costs incurred by data collection.
- Two-sided platform, where data consumers and data providers can make transactions directly if they are members of the data marketplace. Datum, CitizenMe and DataWallet are examples of data marketplaces with such a market structure.

I will consider designing the mechanisms that satisfies the above-mentioned properties in four different data market structures, respectively. And I propose four research questions:

- For a buy-sided broker, given accuracy requirements from data consumers, is there a mechanism is truthful, individually rational, $\vec{\varepsilon}$-personalised differentially private, and minimise the overall payment and how to design it?
- For a sell-sided broker, given privacy protection requirements from data providers, is there a mechanism is truthful, individually rational, accurate and maximises the revenue and how to design it?
- For a two-sided market, is there a mechanism can achieve truthfulness, individual rationality, $\vec{\varepsilon}$-personalised differential privacy, accuracy and a balance and how to design it?
- For a two-sided platform, is there a mechanism can maximise the profit while guaranteeing privacy preservation and accuracy and how to design it?

4 RESEARCH DESIGN

I will take mechanism design approach and follow the research process as shown below. Firstly, unsolved problems are identified from practice and literature. In addition, the properties that the mechanism should satisfy are identified and these properties have been justified by extant research. After that, the problem can be formulated as a constrained optimisation problem. The process of searching for the optimal mechanism is equivalent to finding the optimal solution of the problem.

The next step consists of developing a tool for validation of the model defined by the designed mechanism. In validating the model, an agent-based model (ABM) is built and the corresponding data marketplace is simulated. The data generated by agent-based model and simulation will be collected and used to evaluate the performance of the designed mechanism.

5 CONCLUSION

The advent of data marketplaces and the increasing awareness of data privacy call for data privacy pricing mechanisms which can set prices of data properly while assuring privacy preservation. This study aims to explore this problem in different structures of data marketplaces.

REFERENCES

[1] Mohammad Alaggan, Sébastien Gambs, and Anne-Marie Kermarrec. 2015. Heterogeneous differential privacy. *arXiv preprint arXiv:1504.06998* (2015).

[2] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*. Springer, 265–284.

[3] Lisa K Fleischer and Yu-Han Lyu. 2012. Approximately optimal auctions for selling privacy when costs are correlated with data. In *Proceedings of the 13th ACM Conference on Electronic Commerce*. ACM, 568–585.

[4] Arpita Ghosh and Aaron Roth. 2015. Selling privacy at auction. *Games and Economic Behavior* 91 (2015), 334–346.

[5] Zach Jorgensen, Ting Yu, and Graham Cormode. 2015. Conservative or liberal? Personalized differential privacy. In *2015 IEEE 31St international conference on data engineering*. IEEE, 1023–1034.

[6] Haoran Li, Li Xiong, Zhanglong Ji, and Xiaoqian Jiang. 2017. Partitioning-based mechanisms under personalized differential privacy. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 615–627.

[7] Frank McSherry and Kunal Talwar. 2007. Mechanism Design via Differential Privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*. IEEE, 94–103.

[8] Christopher Riederer, Vijay Erramilli, Augustin Chaintreau, Balachander Krishnamurthy, and Pablo Rodriguez. 2011. For sale: your data: by: you. In *Proceedings of the 10th ACM WORKSHOP on Hot Topics in Networks*. 1–6.

[9] Florian Stahl, Fabian Schomm, Gottfried Vossen, and Lara Vomfell. 2016. A classification framework for data marketplaces. *Vietnam Journal of Computer Science* 3, 3 (2016), 137–143.