AN EFFICIENT PRICING SCHEME FOR DATA MARKETS IN REAL TIME ENVIRONMENT

[#1]**RIMSHA TAHREEM,** *M.Tech Student,*
[#2]***Dr. E.SRIKANTH REDDY***, *Associate Professor,*
[#3]***Dr. N.CHANDRAMOULI***, *Associate Professor &HOD,*
Department of Computer Science and Engineering
VAAGESWARI COLLEGE OF ENGINEERING, KARIMANGAR, TELANAGANA.

ABSTRACT: *The society's insatiable appetites for personal data are driving the emergence of data markets, allowing data consumers to launch customized queries over the datasets collected by a data broker from data owners. In this paper, we study how the data broker can maximize its cumulative revenue by posting reasonable prices for sequential queries. We thus propose a contextual dynamic pricing mechanism with the reserve price constraint, which features the properties of ellipsoid for efficient online optimization and can support linear and non-linear market value models with uncertainty. In particular, under low uncertainty, the proposed pricing mechanism attains a worst-case cumulative regret logarithmic in the number of queries. We further extend our approach to support other similar application scenarios, including hospitality service and online advertising, and extensively evaluate all three use casesover MovieLens 20M dataset, Airbnb listings in U.S. major cities, and Avazu mobile ad click dataset, respectively. The analysis and evaluation results reveal that: (1) our pricing mechanism incurs low practical regret, while the latency and memory overhead incurred is low enough for online applications; and (2) the existence of reserve price can mitigate the cold-start problem in a posted price mechanism, thereby reducing the cumulative regret.*

1. INTRODUCTION

Nowadays, tremendous volumes of diverse data are collected to seamlessly monitor human behaviors, such as product ratings, electrical usages, social media data, web cookies, health records, and driving trajectories.

However, for the sake of security, privacy, or business competition, most of data owners are reluctant to share their data, resulting in a large number of data islands. Because of data isolation, potential data consumers (e.g., commercial companies, financial institutions, medical practitioners, and researchers) cannot benefit from private data. To facilitate personal data circulation, more and more data brokers have emerged to build bridges between the data owners and the data consumers. Typical data brokers in industry include Factual [2], DataSift [3],
Datacoup [4], CitizenMe [5], and CoverUS [6]. On the one hand, a data broker needs to adequately compensate the data owners for the breach of their privacy caused by using their data to answer any data consumer's query, thereby incentivizing active data sharing. On the

other hand, the data broker should properly charge the online data consumers for their sequential queries over the collected datasets, because both underpricing and overpricing may result in loss of revenue for the data broker. The data circulation ecosystem is conventionally called "data market" in the literature [7].

In this paper, we study how to trade personal data for revenue maximization from the data broker's standpoint in online data markets.We summarize three major design challenges as follows. The first and the thorniest challenge is that the objective function for optimization is quite complicated.

The principal goal of a data broker in data markets is to maximize its cumulative revenue, which is defined as the difference between the prices of queries charged from the data consumers and the privacy compensations allocated to the data owners. Let's examine one round of data trading. Given a query, the privacy leakages together with the total privacy compensation, regarded as the reserve price of the query, are virtually fixed. Thus, for revenue maximization, an ideal way for the data broker is to post a price, taking the larger value of the query's reserve price and marketvalue.

However, the reality is that the data broker does not know the exact market value andcan only estimate it from the context of the current query and the historical transaction records. Of course, a loose estimation will lead to different levels of regret: (1) if the reserve price is higher than the market value, implying that the posted price must be higher than the market value, the query definitely cannot be sold, no matter whether the data broker knows the market value or not. Thus, the regret is zero; and (2) if the reserve price is no more than the market value, a slight underestimation of the market value incurs a low regret, whereas a slight overestimation causes the query not to be sold, generating a high regret.

Therefore, the initial goal of revenue maximization can be equivalently converted to minimizing the cumulative regret, particularly, the difference between the data broker's cumulative revenues with and without the knowledge of the market values. Considering even the single-round regret function is piecewise and highly asymmetric,it is nontrivial to perform optimization for multiple rounds.

Another challenge lies in how tomodel the market values of the customized queries from the data consumers. For regret minimization in pricing online queries, the pivotal step for the data broker is to gain a good knowledge of their market values. However, markets for personal data significantly differ from conventional markets in that each data consumer asa buyer rather than the data broker as a seller can determine the product, namely, a query. In general, each query involves a concrete data analysis method and a tolerable level of noise added to the true [8], [9].

Hence, the queries from different data consumers are highly differentiated and are uncontrollable by the data broker. This striking property further implies that most of the dynamic pricing mechanisms, which target identical products or a manageable number of distinct products, cannot apply here.

In addition, existing work on data market design either considered a single query [10] or investigated the determinacy relation among multiple queries [9],but ignored whether the data consumers accept or reject the marked prices. Thus, these work omitted modeling the market values ofqueries and is parallel to this work.

The ultimate challenge comes from the novel online pricing with reserve price setting. For the estimation of a query's market value, the data broker can exploit only the current and historical queries. Thus, the pricing of sequential queries can be viewed as an online learning process.

Besides the usual tension between exploitation and exploration, our pricing problem has three atypical aspects: (1) the feedback after trading one query is very limited. The data broker can observe only whether the posted price for the query is higherthan its market value or not, but cannot obtain the exact market value, which makes standard online learning algorithms inapplicable; (2) thereserve price essentially imposes a lower bound on the posted price beyond the market value estimation,while the ordering between the reserveprice and the market value is unknown.

In addition, the impact of such a lower bound on the whole learning process has not been studied as of yet; and (3) the online mode requires our design of the posted price mechanism to be quite efficient. In other words, the data broker needs to choose each posted price and further update its knowledge about the market value model with low latency.

2. SYSTEM ANALYSIS

EXISTING SYSTEM

First regards general (insensitive) data trading. The researchers from the database community (e.g., Koutris et al. [11]–[14], Lin and Kifer [15]) mainly focused on arbitrage freeness in pricing queries over the relational databases. The existence of arbitrage means that the data consumer can buy a query with a lower price than the marked price through combining a bundle of other cheaper queries. Thus, the data broker needs to rule out arbitrage opportunities to preserve its revenue. Stahl et al. surveyed several empirical pricing strategies in practical data markets [43]. Their later work [44]–[46] introduced data quality as a criterion of pricing

ISSN: 2278-4632
Vol-13, Issue-08, August 2023

and allowed the data consumers to suggest their own prices.

Chawla et al. [47] considered the static revenue maximization problem with the prior knowledge of the data consumers' queries and valuations, while leaving the online setting as an open problem. They mainly adopted two static pricing strategies, called uniform bundle pricing and item pricing. Agarwal et al. [48] proposed a combinatorial auction mechanism to trade data for machinelearning tasks.

Specific to personal data trading, the researchers routinely adopted the cost-plus pricing strategy, where the data broker first compensates each data owner for its privacy leakage and then scales up the total privacy compensation to determine the price of query for the data consumer. Different researchers investigated distinct types of queries from the data consumers.

Ghosh and Roth [10] considered single counting query. The follow-up work by Li et al. [9] further extended to multiple noisy linear queries.We considered the queries of noisy aggregate statistics over private correlated data [16], [17]. Hynes et al. [49] investigated model training requests. Chen et al. [50] studied how to price a trained model with different levels of noise perturbation, by an analogy to the queries over personal data. They also considered how to statically optimize the data broker's revenue under the assumption that the error demands and corresponding valuations of the data consumers are known.

Disadvantages

In the existing work, the system does not provide revenue maximization methods for online pricing. This system is less performance due to lack of Ellipsoid-Based Pricing Mechanism.

ISSN: 2278-4632
Vol-13, Issue-08, August 2023

PROPOSED SYSTEM

The ultimate challenge comes from the novelonline pricing with reserve price setting. For the estimation of a query's market value, the data broker can exploit only the current and historical queries. Thus, the pricing of sequential queries can be viewed as an onlinelearning process. Besides the usual tension between exploitation and exploration, our pricing problem has three atypical aspects:

The feedback after trading one query is very limited. The data broker can observe onlywhether the posted price for the query is higher than its market value or not, but cannot obtain the exact market value, which makes standard online learning algorithmsinapplicable; The system is more effective due to presence of exploratory posted prices under the linear market value model. To The system is more effective due to presence of Ellipsoid-Based Pricing Mechanism.

3. IMPLEMENTATION

Architecture:

The reserve price essentially imposes a lower bound on the posted price beyond the market value estimation, while the ordering between the reserve price and the market value is unknown. In addition, the impact of such a lower bound on the whole learning process has not been studied as of yet; and (3) the online mode requires our design of the posted price mechanism to be quite efficient. In other words, the data broker needs to choose each posted price and further update its knowledge about the market value model with low latency.

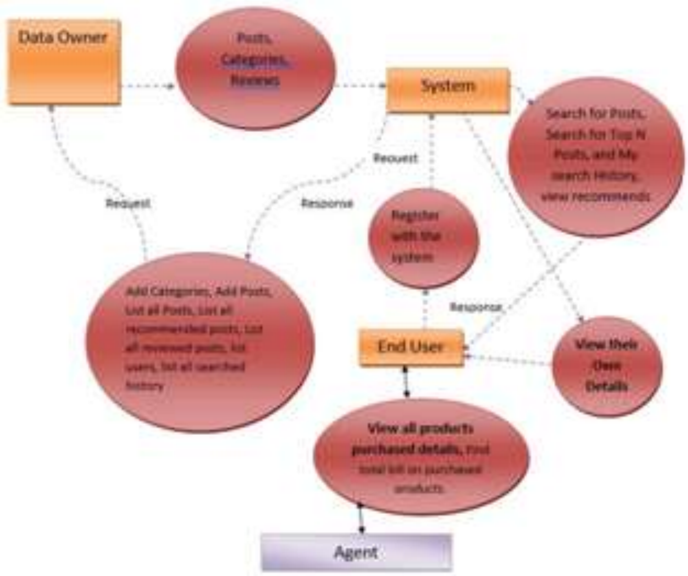
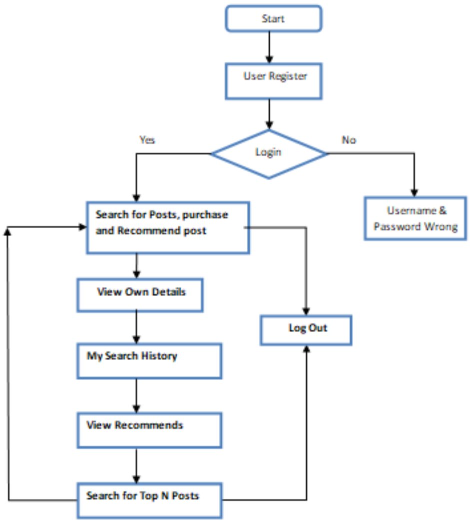


Fig-1: Architecture

ISSN: 2278-4632
Vol-13, Issue-08, August 2023

Flow Chart1: Remote User



4. REUSLUTS

MODULES:

User: Users Buying goods and the services from merchants who sell on the Internet. Since the emergence of the World Wide Web, Shoppers can visit web stores from the comfort of their homes and shop as they sitin front of the computer Consumers buy a variety of items from online stores. In fact, people can purchase just about anything from companies that provide their products online.

Data Owner: Merchants have sought to sell their products to people who surf the Internet. Before people buy anything online, get to know the seller people need to know their contact details for a reputable business should make this information easy to find. And also track the product details of customer mostly like, number of users view the product or purchase the product. A reputable business should also have good customer feedback - friends, family or other customers rate them highly.

Agent: Supplies the product items to multiple stores in a city. And also collects the data details from merchants which product is moving fast and users like mostly. Easily can track and maintain supply the demand product to the market by using advance methods like Weighted Frequent Itemset Mining.

5. CONCLUSIONS

In this paper, we have proposed the first contextual dynamic pricing mechanism with the reserve price constraint, for the data broker to maximize its cumulative revenue in online personal data markets. Our posted price mechanism features the properties of ellipsoid to perform online optimization effectively and efficiently and can support both linear and non-linear market value models, while allowing some uncertainty. We further have illustrated how to support two other similar application scenarios and extensively evaluated all three use cases over three practical datasets. Empirical results have demonstrated the feasibility and extensibility of our pricing mechanism as well as the functionality of the reserve price constraint.

REFERENCES

[1] C. Niu, Z. Zheng, F. Wu, S. Tang, and G. Chen, "Online pricing with reserve price constraint for personal datamarkets," in Proc. of ICDE, 2020, pp. 1978–1981.

[2] "Factual," https://www:factual:com/, 2008.

[3] "DataSift," https://datasift:com/, 2010.

[4] "Datacoup," https://datacoup:com/, 2012.

[5] "CitizenMe," https://www:citizenme:com/, 2013.

[6] "CoverUs," https://coverus:health/, 2018.

[7] M. Balazinska, B. Howe, and D. Suciu, "Data markets in the cloud: An opportunity for the database community," PVLDB,vol. 4, no. 12, pp. 1482–1485, 2011.

[8] A. Roth, "Technical perspective: Pricing information (and its implications)," Communications of the ACM, vol. 60, no. 12, p. 78, 2017.

[9] C. Li, D. Y. Li, G. Miklau, and D. Suciu, "A theory of pricing private data," Communications of the ACM, vol. 60, no. 12, pp. 79– 86, 2017.

[10] A. Ghosh and A. Roth, "Selling privacy at auction," in Proc. of EC, 2011, pp. 199–208.

[11] P. Koutris, P. Upadhyaya, M. Balazinska, B. Howe, and D. Suciu, "Query-based data pricing," in Proc. of PODS, 2012,pp. 167–178.

[12] "Toward practical query pricing with querymarket," in Proc. of SIGMOD, 2013, pp. 613–624.

[13] S. Deep and P. Koutris, "QIRANA: A framework for scalable query pricing," in Proc. of SIGMOD, 2017, pp. 699–713.

[14] [14]"The design of arbitrage-free data pricing schemes," in Proc. of ICDT, 2017, pp. 12:1–12:18.

[15] B. Lin and D. Kifer, "On arbitrage-free pricing for general data queries," PVLDB, vol. 7, no. 9, pp. 757–768, 2014.

[16] C. Niu, Z. Zheng, F.Wu, S. Tang, X. Gao, and G. Chen, "Unlocking the value of privacy: Trading aggregate statistics over private correlated data," in Proc. of KDD, 2018, pp. 2031–2040.

[17] C. Niu, Z. Zheng, S. Tang, X. Gao, and F. Wu, "Making big money from small sensors: Trading time-series data under pufferfish privacy," in Proc. ofINFOCOM, 2019, pp. 568–576.

[18] S. Shalev-Shwartz, "Online learning and online convex optimization," Foundations and Trends in Machine Learning, vol. 4, no. 2, pp. 107–194, 2012.

[19] M. C. Cohen, I. Lobel, and R. P. Leme, "Feature-based dynamic pricing," in Proc. of EC, 2016, p. 817.

[20] I. Lobel, R. P. Leme, and A. Vladu, "Multidimensional binary search for contextual decision-making," in Proc. of EC, 2017, p. 585.

[21] R. P. Leme and J. Schneider, "Contextual search via intrinsic volumes," in Proc. of FOCS, 2018, pp. 268–282

