Strategy-Proof Data Auctions with Negative Externalities[*]

(Extended Abstract)

Xiang Wang[†], Zhenzhe Zheng[†], Fan Wu[†], Xiaoju Dong[†], Shaojie Tang[§], and Guihai Chen[†]
[†]Shanghai Key Laboratory of Scalable Computing and Systems
Department of Computer Science and Engineering
Shanghai Jiao Tong University, China
[§]Department of Information Systems, University of Texas at Dallas, USA

ABSTRACT

Data has appeared to be a new kind of commodity with distinctive characteristics, which make it fundamentally different from physical goods as well as traditional digital goods. Therefore, new trading mechanisms for data need to be designed. In this paper, we model the data market as an auction with negative externalities, and design practical mechanisms for data trading. Specifically, we propose a family of Data auctIons in CompetiTive mArkets, namely DICTA. DICTA contains two mechanisms, including DICTA-FUL and DICTA-PAR. DICTA-FUL is a direct revelation auction mechanism in full competition markets, achieving strategy-proofness and optimal social welfare. In the partial competition markets, we show that the allocation problem is NP-hard. Therefore, we present an approximation algorithm for winner determination. By carefully integrating this approximation allocation algorithm and a charging scheme, DICTA-PAR achieves both strategy-proofness and d-approximation, where d is the maximum degree of the underlying undirected graph of the competition graph.

General Terms

Algorithms, Theory, Economics

Keywords

Data Market, Auction, Externality

1. INTRODUCTION

In recent years, data has become a new kind of commodity that can be traded on the Internet. For example, Xignite

sells financial data, Gnip vends data from social networks, and Sabre trades consumers' booking and searching data on travel. To facilitate online data marketing, several plantforms have emerged, *e.g.*, Azure Data Marketplace, Infochimps, and Dataexchange. These centralized plantforms let data owners upload and sell their data, and let data consumers discover and purchase the data needed.

However, data as a kind of commodity is fundamentally different from physical goods, since data exhibits a distinctive characteristic, *i.e.*, once produced, the data can be duplicated for any number of copies with low or no cost. Digital goods are also in unlimited supply, but for traditional digital goods, such as electronic books, audio files, and pay-per-view movie, the negative externalities do not exist. However, data buyers may want to possess the data exclusively or to limit the distribution of data copies to their competitors. According to a recent survey on data market [3], out of all vendors in the research, 87% of offered data is in business contexts. Buyers, who are mostly companies, purchase their interested data in order to gain advantages in their business against their competitors. Such advantages can be undermined, if their competitors also get the same data, called negative externalities. Therefore, a buyer's valuation on the data not only depends on whether she can get the data set, but also on the data allocation to her competitors.

There indeed exist some works studying negative externalities in share-averse digital good auctions [1, 2]. However, there are huge differences between our auction model and theirs. A major difference is that, in share-adverse auctions, externality only depends on the number of buyers sharing the item, which is just like the complete competition scenario in our paper. However, in the partial competition scenario, each buyer can submit her set of competitors. Yet, another major difference is that they all assume the externality function being a common knowledge. However, in our model, the information about externality, specifically, tolerance bound and set of competitor, is private information.

In this paper, we conduct an in-depth study on designing strategy-proof data auctions. We propose a family of Data auctIons in CompetiTive mArkets (DICTA). DICTA contains two mechanisms, namely DICTA-FUL and DICTA-PAR. Specifically, DICTA-FUL is for an ideal but meaningful setting, *i.e.*, full competition markets, where any pair of buyers compete with each other. In this scenario, we propose a computationally tractable algorithm to calculate an optimal allocation, so that Vickrey-Clarke-Groves (VCG) mechanism can be applied to achieve both efficiency and strategy-proofness. We further consider a more practical

scenario, *i.e.*, partial competition markets, in which each buyer just competes with a subset of the buyers instead of all. In this setting, the VCG mechanism is no longer appropriate due to computational intractability of searching for the optimal allocation. Therefore, we turn to design an approximation algorithm that finds a sub-optimal allocation, and incorporate it with a simple but effective pricing scheme to achieve strategy-proofness.

2. RESULTS

We consider a one shot sealed bid data auction with a trusted auctioneer and a set of n buyers $\mathbb{N} = \{1, 2, 3, \cdots, n\}$. There is a single set of data that can be duplicated to any number of copies, and then be sold to different interested buyers. Each buyer $i \in \mathbb{N}$ is interested in a single copy of the data set.

Each buyer has a set of business competitors $S_i \subseteq \mathbb{N} \setminus \{i\}$, and can only tolerate up to t_i competitors sharing the same set of data. If the buyer i wins the data set and there are no more than t_i competitors winning at the same time, then she has a valuation v_i on the data set. Otherwise, the valuation of the data set to the buyer i becomes 0. The triple $\theta_i = (S_i, t_i, v_i)$ is the private information of buyer i, and is widely known as type in the literature. In the data auction, each buyer i proposes a bid $b_i = (\hat{S}_i, \hat{t}_i, \hat{v}_i)$, which can differ from her type.

After collecting the bids from the buyers, the auctioneer constructs a directed *competition graph* G, in which each vertex represents a buyer, and each edge $(i, j), i, j \in \mathbb{N}$ indicates that buyer j is in buyer i's competitor set. Then, the auctioneer determines a set of winners $\mathbb{W} \subseteq \mathbb{N}$ and calculates a payment p_i for each winner.

2.1 FULL COMPETITION MARKETS

In this section, we present data auction mechanism DICTA-FUL for the full competition markets, in which any pair of buyers compete with each other, *i.e.*, the competition graph is a complete graph.

In this scenario, we can design a polynomial time algorithm to compute the optimal allocation, and thus can apply the celebrated Vickrey-Clarke-Groves (VCG) mechanism to achieve strategy-proofness. Therefore, we focus on algorithm design of allocation rule in this section.

In the data set allocation algorithm, we first sort all the buyers in a non-increasing order of their declared valuations, and denote the sorted list by Γ.

$$\Gamma : \hat{v}_1' \geq \hat{v}_2' \geq \cdots \geq \hat{v}_n'.$$

If there exists a tie, we break it arbitrarily. We note that $\hat{v}_i'$ may not be equal to $\hat{v}_i$ after sorting, and we will apply the allocation algorithm to the buyers according to the order in the sorted list.

Since any pair of buyers compete with each other in the full competition markets, for each winner, the number of her winning competitors is equal to the number of all the winners minus one. Thus, we can traverse every possible number of winners from 1 to n. For each number $m \in \{1, 2, \cdots, n\}$, we pick top m buyers whose tolerance bounds are no less than $m - 1$ from the sorted list Γ, and calculate their social welfare. If the number of qualified buyers is less than m, we simply select all of them without filling out the quota. Finally, we locate the m achieving the maximal social welfare, and output the corresponding set of winners as $\mathbb{W}$.

2.2 PARTIAL COMPETITION MARKETS

In this section, we present data auction mechanism DICTA-PAR for partial competition markets, in which each buyer competes with a subset of the buyers. The previously studied full competition markets are special cases of the partial competition markets.

Due to the hardness of allocating in partial competition markets, we here present our computationally efficient algorithm for data set allocation achieving d-approximation ratio.

Same as before, we first sort all the buyers in a non-increasing order of their declared valuations, and denote the sorted list by Γ. If there exists a tie, we break it arbitrarily.

Following the sequence specified in Γ, we visit each buyer i one by one, and check whether she can be allocated the data set without violating the following two constraints.

▶ The first constraint is that allocating the data set to buyer i should not breach any of the previously selected winners' tolerance bounds, *i.e.*,

$$\forall j \in \left\{ k | k \in \mathbb{W} \wedge i \in \hat{S}_k' \right\}, \hat{t}_j' > \left| \hat{S}_j' \cap \mathbb{W} \right|.$$

▶ The second constraint is that the number of previously selected winning competitors of buyer i should not exceed her tolerance bound, *i.e.*,

$$\hat{t}_i' \geq \left| \hat{S}_i' \cap \mathbb{W} \right|.$$

If both the above constraints are satisfied, we allocate the data set to buyer i; Otherwise, we deny buyer i's bid.

Since DICTA-PAR follows a greedy allocation rule, we adopt the concept of critical bid to determine the payment for each of the buyers. Given the distinctive characteristics of the binary valuation function in our data auction, we achieve strategy-proofness by adopting critical bid.

3. CONCLUSIONS

In this paper, we have modeled the data trading market as an auction with negative externalities. We have studied two different but connected market scenarios, including full competition markets and partial competition markets. For full competition markets, we have designed DICTA-FUL to compute the optimal allocation, and integrate it with the celebrated VCG mechanism. Thus, DICTA-FUL achieves both strategy-proofness and optimal social welfare. For partial competition markets, we have shown that finding the optimal allocation is NP-hard, and also hard to approximate. In this scenario, we have designed DICTA-PAR, which is a combination of a d-approximation allocation algorithm and a carefully designed charging scheme.

REFERENCES

[1] J. Pei, D. Klabjan, and W. Xie. Approximations to auctions of digital goods with share-averse bidders. *Electronic Commerce Research and Applications*, 13(2):128–138, 2014.

[2] M. Salek and D. Kempe. Auctions for share-averse bidders. In *WINE*, volume 8, pages 609–620. Springer, 2008.

[3] F. Schomm, F. Stahl, and G. Vossen. Marketplaces for data: An initial survey. *ACM SIGMOD Record*, 42(1):15–26, 2013.