Buying Data using Data: a Buyer's Perspective in a Data Market with Infinite Supply

Anonymous Author(s)

ABSTRACT

place holder place holder

ACM Reference Format:

Anonymous Author(s). 2020. Buying Data using Data: a Buyer's Perspective in a Data Market with Infinite Supply. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

place holder for greg.

2 DATA MARKET

Next, we layout our data market setting partially following

2.1 Data Market Model

A *data market* $\mathcal{M}$ is a composition of three sets of entities, namely *buyers* $\mathcal{B}$, *sellers* $\mathcal{S}$, and *mediator* m. We shall refer to the first two entities also as the *players* $\mathcal{A} = \mathcal{B} \cup \mathcal{S}$ in the market. Players may be individuals or groups (*e.g.,* companies or organizations) interested in trading datasets. We use $\mathcal{D}$ to denote the set of authorized datasets in $\mathcal{M}$. Usually, a dataset D is associated with a domain and several of other properties that represent it including *e.g.*, the features it contains. Each buyer $b \in \mathcal{B}$ is interested in a set of datasets $D_b \subseteq \mathcal{D}$ whereas each seller offers a set of products $D_s \subseteq \mathcal{D}$. The mediator is in charge of the transactions between the different players in the market.

RS: say something about the value of data (find references)

In what follows, each player $a \in \mathcal{A}$ has a different utility from a dataset, according to which they can set prices. A market is a temporal ecosystem. As such, we assume that the market has a finite horizon T and each interaction between the different entities takes place in a discrete timestamp $t < T$.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00

https://doi.org/10.1145/nnnnnnn.nnnnnnn

3 BUYING DATA IN A DATA MARKET

overview

3.1 "Naive" Strategies for Buying Data

3.2 Products Allocation Optimization Strategy

RS: revise:

Next, we define the *products allocation* problem with respect to a single player a before timestamp t in the market horizon T.

Let $\mathcal{D} = \langle D_1, D_2, \ldots, D_n \rangle$ be the set of relevant products for the player a. Generally, a player has a budget ? and a valuation value v_i for each relevant product D_i. At timestamp 0, the set of products $\mathcal{D}$, its corresponding valuations V and an initial ? are set. While the valuation values of products stays constant throughout the horizon, $\mathcal{D}$ and ? change with respect to the interaction in the market. We denote the products set available at time t and the current budget at time t as $\mathcal{D}^t$ and $?^t$, respectively. We assume that the only thing that changes the product set is that a player has purchased the product and thus, $\mathcal{D}^t \subseteq \mathcal{D}^{t'}, t' < t$ and $\mathcal{D}^0 = \mathcal{D}$. The budget changes with respect to the revenue from each purchased product and its cost.

In practice, the player can not simply purchase any product she wishes, even if she has a large enough budget for two primary reasons. First, a the "buying" player and the "selling" player may not be synchronized (*e.g.,* the seller demands a price that the buyer is not willing to pay). Second, when an auction is in order, the player may not win the auction for the product she wishes to purchase. Thus, at time t the player has to select a subset of products $\bar{\mathcal{D}} \subseteq \mathcal{D}^{t-1}$ that maximizes her future revenues. To simplify the notation, we denote that number of products available at time t as $m \leq n$. In what follows, let $C = \langle c_1, c_2, \ldots, c_m \rangle, c_i \in \mathbb{R}$ and $W = \langle w_1, w_2, \ldots, w_m \rangle, w_m \in \{0, 1\}$ represent a realization of costs and win indicators of the relevant products after timestamp t has completed, respectively.

In practice, a player has to allocate a subset of products at the beginning of time t. Thus, the player does not know the actual cost of products when the product allocation takes place. Accordingly, the player has to estimate the costs $\hat{C} = \langle \hat{c}_1, \hat{c}_2, \ldots, \hat{c}_m \rangle$ and wining indicators $\hat{W} = \langle \hat{w}_1, \hat{w}_2, \ldots, \hat{w}_m \rangle$. Using these estimations, the *products allocation* problem can be formalized as follows:

$$\begin{aligned} \underset{X}{\text{minimize}} \quad & \sum_{i=1}^{m} (v_i - \hat{c}_i) \cdot \hat{w}_i \cdot X_i \\ \text{subject to} \quad & \sum_{i=1}^{m} \hat{c}_i \cdot \hat{w}_i \cdot X_i \leq ? \\ & X_i \in \{0, 1\} \; i = 1, \ldots, m. \end{aligned}$$

Recalling that $\hat{w}_i \in \{0,1\}$ the player can actually decrease the size of $\mathcal{D}^{t-1}$ to the set of product she estimates she would win, *i.e.,* $win(\mathcal{D}^{t-1}) = \{p_i \in \mathcal{D}^{t-1} | \hat{w}_i = 1\}$. We denote the size of $win(\mathcal{D}^{t-1})$ as m_{win}

$$\begin{aligned} \underset{X}{\text{minimize}} \quad & \sum_{i=1}^{m_{win}} (v_i - \hat{c}_i) \cdot X_i \\ \text{subject to} \quad & \sum_{i=1}^{m_{win}} \hat{c}_i \cdot X_i \leq ? \\ & X_i \in \{0,1\} \; i = 1, \ldots, m_{win}. \end{aligned}$$

we begin with problem definition assuming auctions and than relax the auction such that each buyer is allocated to a seller and in case there is a match (the price the buyer is willing to pay is higher that the price limit set by the seller), the buyer gets the product.

3.2.1 *Estimating C.*

4 EVALUATION

bla bla bla

5 RELATED WORK