

基于博弈论和拍卖的数据定价综述

张小伟¹, 江东¹, 袁野²

1. 东北大学计算机科学与工程学院, 辽宁 沈阳 110819; 2. 北京理工大学计算机学院, 北京 100081

摘要

在大数据时代, 随着数据爆炸式的增长, 将数据视为一种商品, 建立一个高效的数据交易市场, 通过数据交易市场为数据所有者提供利益补偿, 为数据需求者提供数据或服务, 使得数据能够在数据所有者和数据需求者之间充分地自由流动显得尤为重要。然而如何为数据设定合理的价格是必须考虑的。对基于博弈论和拍卖的数据定价进行了研究, 调查了该分类下不同的数据定价模型, 并将其分为不同的类型, 综合比较各个模型的优劣。将常见的数据交易市场进行分类, 指出不同的数据交易框架在实现过程中的优点和挑战。对已有的数据定价研究进行总结, 以便数据定价领域的学者能更轻松地掌握该领域的研究现状及重点。

关键词

数据定价; 数据交易市场; 博弈论; 拍卖

中图分类号: TP311

文献标识码: A

doi: 10.11959/j.issn.2096-0271.2021039

A survey of game theory and auction-based data pricing

ZHANG Xiaowei¹, JIANG Dong¹, YUAN Ye²

1. School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China

2. School of Computer Science & Technology, Beijing Institute of Technology, Beijing 100081, China

Abstract

In the era of big data, with the explosive growth of data, regarding data as a commodity and establishing an efficient data trading market is a important thing. By data trading's way, profit compensation is provided for data owners, and raw data or services are provided for data consumers, so that data can flow fully freely between data owners and data consumers. However, how to set a reasonable price for the data is necessary. Data pricing based on game theory and auctions was investigated. Different data pricing models under this category were investigated. These models were divided into different types, and the advantages and disadvantages of each model were compared comprehensively. Then, common data trading markets were classified, and the advantages and challenges of different data transaction frameworks in the implementation process were pointed out. A summary of existing data pricing research was made, so that scholars in the field of data pricing can more easily grasp the current research status and the key of data pricing.

Key words

data pricing, data trading market, game theory, auction

1 引言

近年来,随着物联网(Internet of things, IoT)的发展,基于IoT的智慧世界系统(智慧电网、智慧交通、智慧城市)发展迅速,来自各种设备或应用程序的日志文件、数据库文件、视频、图片、文档等数据呈爆发式增长^[1-5]。数据作为一种新的生产要素,促进及引领了新一轮的科技创新和经济增长。然而,在大数据繁荣发展的背后有一个问题不容忽视,即“数据孤岛”。不同的企业甚至不同的部门之间的数据无法流通,如统计局和医院等机构长期存储大量个人信息数据,滴滴和美团等企业掌握了大量的用户信息及出行信息,若无法将数据以合理的价格销售,不仅无法最大化数据拥有者的收益,还会为维护和存储数据耗费大量财力。近年来发展迅速的机器学习和深度学习需要大量的数据进行训练以提高测试精度,但是由于数据收集困难,导致训练精度无法提升,因此数据需求者需要一个渠道来获得大量且可靠的数据。基于上述需求,Oliveira A C等人^[6]提出了数据即服务(data as a service, DaaS)的概念。在DaaS中,数据被认为是一种商品,可以通过购买数据,并将其处理之后获得知识,进而帮助企业、政府做出决策。因此,在数据拥有者和数据需求者之间搭建一个数据交易平台,越来越受到人们的关注。

要建立一个合理有效的数据交易市场,需要解决以下几个问题。第一个问题是市场的参与者是谁?数据的参与者可以分为3类,分别是数据拥有者、数据消费者、交易中间人。数据拥有者通常为政府、私人企业(如微软、雅虎、微博等)或者个人;数据消费者即数据需求者,可能为数

据处理算法开发人员、数据分析师等;交易中间人即第三方交易平台,如现有的贵阳大数据交易所、中关村数海大数据交易平台等。按照交易中间人职能的不同,可以将其分为中间商和服务提供商,中间商负责连接数据拥有者和数据消费者,此时交易的是原始数据,如中关村数海大数据交易平台;而服务提供商一般会对数据拥有者的原始数据进行处理,为数据消费者提供处理之后的数据,如贵阳大数据交易所。第二个问题是如何为确定交易的数据设定合理的价格?在设定相应的数据定价模型时,必须考虑相应的市场结构以及数据需求者和数据拥有者的需求,确保双方的收益。第三个问题则与数据交易市场相关,即选择何种数据交易市场才能确保市场参与者的需求得到满足,实现公平、隐私保护等,如选用基于分布式的数据交易市场能够解决隐私泄露的问题。为了解决上述问题,笔者对大数据交易市场进行了全面的调查,以帮助希望了解该领域的新来者对这一复杂的学科和新兴的研究领域有一个大致的了解。

2 数据定价及相关概念

2.1 数据的定义及特性

在大数据时代,每天都有接近25 TB的数据产生^[7]。中国信息通信研究院发布的《大数据白皮书(2020年)》预测,2030年全球数据生产量将达到612 ZB,而2035年时将达到2 142 ZB,全球数据量将迎来爆发式的增长。

在从庞大的数据集中提取数据的价值时,经常用数据挖掘和商业智能(business intelligence, BI)来描述大数据的处理过程,并且往往需要通过这些手段来发现数

据的价值。数据的价值与数据的交易价格关系密切。

给数据一个确定的价格,将数据视为一种可以自由交易的商品,可以提高数据市场的流动性,从而创造更大的价值。然而相比于传统的实物商品,数据商品在生产、分配和消费的流程中显著降低了成本^[8],分别表现在商品的搜索成本、生产成本、复制成本、运输成本和跟踪与验证成本这5个方面。并且,数据的特性导致数据的复制成本接近0,即边际成本接近0,这导致无法使用边际成本等于边际收益的方式使数据产品的收益最大化,这与传统的实物产品不同^[9-10]。同样,数据产品的一个明显特征为非竞争性,即一个数据消费者消费一件数据商品并不会减少和降低其他数据消费者可以得到的产品数量和质量。但是数据同样具有排他性,即数据消费者需要支付一定的费用来获得数据的使用权。

2.2 常见的数据定价方式

Muschalle A等人^[11]介绍了数据市场中可以参考的收费策略,主要包括如下几种。

- 通过控制版本来收费,如企查查等应用,用户可以免费使用基础版本查询信息,但是如果需要更高级的信息,则需要购买专业版。

- 基于使用时间的收费,像拨号上网那样,根据使用时长收取费用,但是该方式的缺陷是对于一个无法明确自己需求的数据需求者或者不熟悉购买方式的数据需求者而言,他需要花费大量时间,导致花费太高,从而可能抑制数据需求者的购买欲望。

- 以固定费用获取一定量数据的永久使用权或应用程序接口(application programming interface, API)的调用权。

Kantere V等人^[12]对该方法进行了详细的研究。

- 以统一的定价对产品进行收费,如参考文献[13]中的方式,但是该方式缺乏灵活性。

- 组合定价,即固定的基础费用加额外的单位费用。如各大电信运营商的收费分为基础的套餐费用和在此基础之上的额外单位费用。

数据交易市场中常见的数据定价方式可以分为:基于博弈论和拍卖的数据定价^[14-18]、基于查询的数据定价^[19-21]、基于模型的数据定价(model-based pricing, MBP)^[22-23]。其中,基于博弈论和拍卖的数据定价更注重价格发现的过程,利用数据交易市场的不同特点,使用不同的方式确定适当的数据价格。

2.3 数据定价模型的标准

由前文可知,有多种多样的方法来设计和实现数据产品的定价。针对如何衡量这些定价模型的优劣以及一个优秀的定价模型应该满足什么标准,Pei J^[24]将数据定价的标准分为6类,具体如下。

- 诚实:参与数据交易的各方均有动力“说真话”,报出自己真实的估价。

- 收益最大化:无论是数据拥有者还是数据消费者,他们参与交易最直接的动力是使自身的利益最大化。

- 公平:数据的卖家能够根据自身的贡献量获得公平的收入份额。

- 无套利定价:无法通过多个市场之间的价格差异来获得收益。

- 隐私保护:能够保护参与交易的各方的信息。

- 计算效率:由于数据商品自身的特性,如数据的时效性或多卖家多买家之间的适当匹配机制,需要高速有效的计算方法。

本文将沿用这6个标准来衡量下述研究的优劣。

大多数模型追求的是实现上述6个标准中的部分标准,而无法满足全部要求。如Goldberg A V等人^[25]的研究表明,在拍卖这一价格发现过程中,无法同时实现诚实、竞争性和无嫉妒性。因此,如何在诸多标准之间做出取舍也是当前数据定价研究需要关注的主要内容。此外,数据商品还具备其他特性,这些特性要求数据定价模型实现某些特殊的标准,具体如下。

- 某些数据可能是对时间敏感的,同一数据在不同时间段的价值可能会完全不同,因此对于该类数据的定价要求是实现较高的计算效率。

- 某些数据的价格可能与数据中包含的信息有相关性,而公布数据的价格可能会泄露数据的信息,因此对于该类数据的定价要求是实现无套利定价。

- 数据的价值并不取决于数据量,而是取决于其提供的信息量,对于该类数据的定价要求是实现诚实、公平、收益最大化等。

3 基于博弈论和拍卖的定价

博弈论和拍卖均是关于决策制定的研究,二者均需多方参与,如数据的拥有者、数据的消费者以及中间人,并且每个参与者做出的决策都会影响其余参与者做出的决策,即参与者之间相互影响,最终做出在当前情况下的最优决策。本节研究的是如何使用博弈论和拍卖的方式为数据设置合理价格。

3.1 基于博弈论的定价

博弈论又被称为对策论或赛局理论

等,属于运筹学的范畴,是一门有关“互动行为”的科学^[26-28]。博弈论是一种方法,目前已被应用于多种学科,如经济学、计算机科学、国际关系、军事战略等。博弈是指由两个或两个以上的理性人或组织,参加一系列具有竞争或对抗性质的行为。参与博弈的各方会受到一系列的环境约束(即规则),各自拥有不同的目标或者利益,为了实现自己的目标或将利益最大化,参与者必须综合考虑所有对手的全部可行行动方案,并在其基础上做出最有益于自己的决策。

为了理解博弈的含义,下面给出一些博弈方面的术语。

- 效用函数:效用函数是参与博弈的各方能够从该次博弈中获得的效用水平,反映了参与者对结果的期望。经济学中规定效用函数必须是可量化的,它可以是连续函数,也可以是离散函数。效用的取值可以为正值,也可以为负值。参与博弈的每一方都有自己的效用函数,但是彼此并不一定了解对方的效用函数。

- 参与博弈的主体:博弈的参与者是一个能够选择合适的决策从而最大化自身效用的主体,该主体可以是个人,也可以是团体,如国家、企业、组织等,以下简称参与者。

- 参与者可选择的策略或者策略的组合:策略是指参与者选择自身行动的规则,该规则将决定参与者在什么情况下选择什么行动,如“敌进我退,敌驻我扰,敌疲我打,敌退我追”是一种策略,这里,“敌”与“我”是参与博弈的双方,“进”“退”“驻”“扰”“疲”“打”“退”“追”是8种不同的行动,由战略规定于何时采取何种策略^[29]。

在基于博弈论数据定价的过程中有诸多参与者,根据当前的研究,可将参与者分为3类:数据拥有者(卖家)、数据消费者(买家)、中间人。通常情况下均假定参与

博弈的各方均为“理性人”，即参与人的每一个经济活动都是利己的，试图以最小的经济成本获得最大的收益。

根据不同的博弈类型，可将现有常见的博弈论定价模型分为3类：基于非合作博弈(non-cooperative game)的定价模型、基于Stackelberg博弈的定价模型、基于讨价还价博弈的定价模型。

3.1.1 基于非合作博弈的定价模型

非合作博弈是指一种参与人之间不可能组成联盟或者达成一种具有约束力的协议的博弈类型^[30]。

Luong N C等人^[14]设计了一个定价模型来评估IoT传感数据的价格。该模型假设了一个IoT数据交易市场，在市场中卖家为了销售各自的数据形成了竞争关系。该市场可以建模为一个非合作博弈模型，其中数据拥有者扮演的是卖家的角色，卖家分别拥有各自的定价策略。令 (V, π) 表示拥有 n 个参与人的博弈，其中 V_i 表示第 i 个参与人可选择的定价策略空间， $V = V_1 \cdot V_2 \cdot V_3 \cdots V_n$ 表示各个参与人的定价策略 V_i 的笛卡儿积， $\pi = (\pi_1(v), \dots, \pi_n(v)) \in R^n$ 表示参与人 i 所能得到的效用函数构成的向量。将 v_i 设定为参与人 i 的定价策略， $v_i \in V_i$ ，可以得到一个由 n 个参与人构成的关于策略的 n 维向量 $v = (v_1, v_2, v_3, \dots, v_n)$ 。同时，在博弈论中， v_i 的最优策略选择依赖于其他参与人的策略选择。因此，设定向量 v_i' 表示除不选择定价策略的参与人 i 以外的所有参与人构成的策略组合，即 $v_i' = (v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n)$ 。这3个因素之间的联系是参与人 i 使用给定的定价策略 v 来获得相应的报酬 π 。一个给定的策略向量 $v^* = (v_1^*, v_2^*, v_3^*, \dots, v_n^*) \in V$ 是纳什均衡的，没有任何一个销售者可以在其他销售者不改变自身策略的情况下，通过改变自身的策略

提高自己的收益，即

$$\forall i, v_i \in V_i : \pi_i(v_i^*, v_i^*) \geq \pi_i(v_i, v_i^*) \quad (1)$$

式(1)表明鉴于纳什均衡状态，卖家没有动机改变自身的决策，因为那样会使自身的预期收益变低。式(1)表明纳什均衡是指在给定别人策略的情况下，做出相应最优的选择。但是应当注意并不是所有的博弈均存在纳什均衡，而且有时一个博弈中会存在多个纳什均衡。因此，使用非合作博弈对数据集进行定价的充分必要条件是该博弈有且只有一个纳什均衡。

由于参与博弈的各方均为“理性人”，参与者会将将自己的战略建立在假定对手会将其收益最大化的基础上。因此，完全竞争市场^[31-33]中的均衡就是非合作博弈均衡。在这种市场中，每个参与者都会根据别人的价格策略来决定自己的价格策略。在完全竞争市场中，数据拥有者为了使自身的收益最大化，希望在销售自身数据时价格尽可能高。数据消费者为了使自身的效用最大化，希望在得到数据时尽可能地压低价格。在传统的商品市场中，使用边际成本等于边际收益的方式使数据产品的收益最大化，但是数据商品的特性(即边际成本接近0)导致无法使用上述方法实现利润最大化。故Li Z N等人^[34]提出了一种定价方法，该方法将所有需求者和供应商都考虑在内，在不降低价格的情况下得出适当的价格。但是，由于将动态创建大量物联网数据交易，这通常会成为集中式定价系统的性能瓶颈，因此该方法对于物联网数据交易是不切实际的。

3.1.2 基于Stackelberg博弈的定价模型

非合作博弈的纳什均衡成立的前提是参与博弈的卖家彼此知道对方的策略，并且同时宣布自己的策略。然而，在现实

生活中,这个前提成立的情况较少,参与者无法计算其纳什均衡,因此他们无法为数据设定合理的价格。在现实中更常见的情形是:一个卖家(追随者)等待其他卖家(领导者)先宣布自己的定价策略,然后追随者在领导者做出定价策略的情况下,做出对应策略的优化,从而确定相对最优的定价策略,这种博弈模式被称为Stackelberg博弈^[35]。

在Stackelberg博弈中,卖家1(领导者)首先确定自己的定价策略 v_1 ,卖家2(追随者)在观测到 v_1 后,确定自己的定价策略 v_2 。因为卖家2是在观察到 v_1 之后确定自己的定价策略的,所以该博弈是完美信息动态博弈。卖家1率先行动,无法观察到 v_2 ,因此,卖家2的定价策略应该是一个由 V_1 决定的 V_2 ,是一个映射函数 T ,即 $V_1 \rightarrow V_2$ 。Haddadi S等人^[36]证明,相较于使用其他模型,使用Stackelberg博弈模型,领导者可以获得更大的收益。Lyu X Y等人^[15]指出,在数字市场中参与Stackelberg博弈的各方均可实现利益最大化,尤其是对于领导者而言,可将其效用最大化。

Mei L J等人^[16]讨论了针对纯捆绑定价和分开定价两种情况,在数据拥有者和中间商之间搭建Stackelberg博弈模型(其中数据拥有者为领导者,中间商为追随者)、实现双方利益最大化的条件下,数据拥有者应该采取何种定价方式才能使得自身利益最大化。参考文献[16]虽然涉及数据定价的问题,却忽略了数据交易市场信息不完全以及市场竞争的客观条件。

Liu K等人^[37]搭建了一个两阶段的Stackelberg博弈模型来解决数据需求者的定价和购买问题。在最简单的情况下,市场包含多个数据拥有者 $i = \{1, 2, \dots, N\}$ 、一个数据需求者、一个中间商。Liu K等人^[37]假设:中间商可以获得数据拥有者的名单以及数据需求者和数据拥有者的交易相关信

息。第一阶段,每一个数据拥有者对自己拥有的数据设定合理的价格,中间商作为领导者,将定价策略的集合展示给数据需求者;第二阶段,数据需求者作为追随者,根据给出的定价策略空间,决定自己的购买策略;最后,中间商根据数据拥有者的服务质量以及数据需求者的购买意愿来决定由数据拥有者中的赢家与数据需求者发生交易, (p_i, x_i) 表示以单价 p 从数据拥有者 i 处购买了数量为 x 的数据。

Xu C Z等人^[38]在数据拥有者、服务提供商和数据需求者之间构建了一个3层的Stackelberg博弈模型。在该模型中,第一层的Stackelberg博弈发生在数据拥有者和服务提供商之间,数据拥有者决定原始数据的价格,服务提供商根据定价策略做出相应的购买策略。在第二层,服务提供商在购买原始数据之后需要对数据进行加工,并且提供 N 种不同精度版本的数据供数据需求者选择,并为其设置合理的价格。在第三层,数据需求者选择要购买的数据版本,并决定购买的数量。

使用Stackelberg博弈进行数据定价^[16,37-41]的缺陷在于:在这种模型中,每个数据拥有者需在主供应商宣布价格后宣布自己的价格,但在数据交易环境中,确定主要数据拥有者的方式困难且效率低下,使得该模型在数据交易市场中的实施难度较大。

3.1.3 基于讨价还价博弈的定价模型

讨价还价博弈指的是由两名或者更多的参与人就如何分配一个物品达成协议,为了达成这种协议所有参与人需要进行谈判。假设在一个简单的数据交易市场中,只有当数据拥有者和数据需求者对某种数据商品的销售价格达成一致时,交易才会发生。

在讨价还价模型中, r_i 表示数据拥有

者的保留价格,这是数据拥有者销售数据所能接受的最低售卖价格; r_c 表示数据需求者的保留价格,这是数据需求者购买数据所能接受的最高支付价格;数据拥有者和数据需求者分别报出自己的价格策略 p_o 和 p_c 。数据拥有者想要确定最优的定价策略 p_o^* 来使自身的收益 $\pi_o(p_o, r_o)$ 最大化,即 $\pi_o(p_o^*, r_o) \geq \pi_o(p_o, r_o), \forall p_o$ 。同样地,数据需求者想要制定最优的出价策略 p_c^* 来使自身的收益 $\pi_c(p_c, r_c)$ 最大化,即 $\pi_c(p_c^*, r_c) \geq \pi_c(p_c, r_c), \forall p_c$ 。在讨价还价博弈中,如果 $p_c^* \geq p_o^*$,则数据以销售价格 p 成交, $p = kp_c^* + (1-k)p_o^*, 0 \leq k \leq 1$ 。最终得到了讨价还价博弈的纳什均衡解 (p_o^*, p_c^*) 。

Jung K等人^[17]使用差分隐私(differential privacy, DP)的方式给原始数据或查询出来的结果增加噪声,这是因为DP提供了一个令人信服的方式来衡量隐私(如隐私的损失 ϵ)。近年来有很多面向数据拥有者的基于隐私损失方面的定价策略^[42-48]。Jung K等人^[17]提出了一种公平协商的方式,采用Rubinstein讨价还价模型来确定数据的价格以及隐私损失的价值,保证了公平交易。

讨价还价博弈适用于复杂谈判条件下的谈判,且其最后的解是合作博弈最终的解决方案,因此其常被用于诸多领域的资源分配,如无线体域网(wireless body area network, WBAN)^[46]、无线传感器网络(wireless sensor network, WSN)^[47]、频谱分配^[48]。但是在讨价还价博弈中,需要供需双方通过谈判达成协议,而谈判过程通常是耗时和浪费资源的,故将该模型应用于数据交易市场存在一定困难。

3.2 基于拍卖的定价

拍卖是流行的数据交易机制之一。一般来说,拍卖是一种经济驱动的方案,其

目的是通过买卖双方的竞价过程分配商品,并建立相应的价格^[49]。当前对拍卖理论的研究^[50-51]以及拍卖理论在诸多领域的应用^[52-55]已经有了很好的探索。在信息不对称的经济环境中,拍卖是一种形式简单却又具备完整定义,能够确保公平和效率,以及卖方的收益最大化的方案,因此在解决大数据交易问题方面拍卖机制显示出巨大的潜力。在对大数据交易市场中拍卖理论的相关工作做出总结之前,首先介绍拍卖机制的基本概念。

- **投标方:** 投标人,指在市场上投标并以购买商品为目的的人,即买方。在大数据市场中,一般由数据需求者充当买方。

- **拍卖商:** 拍卖商扮演的是代理角色,与上述中间商的作用相似,负责运行拍卖流程,确定获胜者,并进行支付和分配。

- **卖方:** 指希望通过售卖数据增加自身收益的个人或团体,如数据拥有者。

- **投标人的估价:** 在拍卖过程中,投标方和卖方都对他们需要或出售的每一单位商品进行估价。此外,估价可以高于或低于最终清算价格,这是由拍卖商在拍卖过程中决定的。

- **结算价格:** 拍卖中最终成交的价格(买卖双方同意达成交易的价格),即落锤价。拍卖价格有两种类型,即要价和竞价。卖方提出一个要价(即出售该商品的价格),而投标方可以提出一个投标价格(表示他们为获得该商品所愿意提交的价格)。结算价格则是由拍卖商根据社会福利最大化等优化目标来确定的最终交易价格。

根据参与拍卖的投标方和卖方人数的不同,可以将目前基于拍卖定价方式的数据市场的研究分为单边拍卖(单个卖方、多个投标方)和双边拍卖(多个卖方、多个投标方)。基于拍卖的大数据交易市场的典型框架如图1所示^[56]。

3.2.1 单边拍卖

在单边拍卖中，卖方提交竞价到拍卖商，拍卖商根据不同的拍卖机制，确定最后的胜利者。其市场结构如图2所示，通常数据拥有者是“资源优势方”，而最终的价格是由其和数据需求者即投标方共同决定的，投标方被称为“信息优势方”。

(1) 第k价格密封拍卖

第一价格密封拍卖和第二价格密封拍卖是重要的两种第k价格密封拍卖的方式。在第一价格密封拍卖中，投标方以密封的形式独立出价，获胜的投标方是出价最高的人，需要支付最高的价格来获得商品，

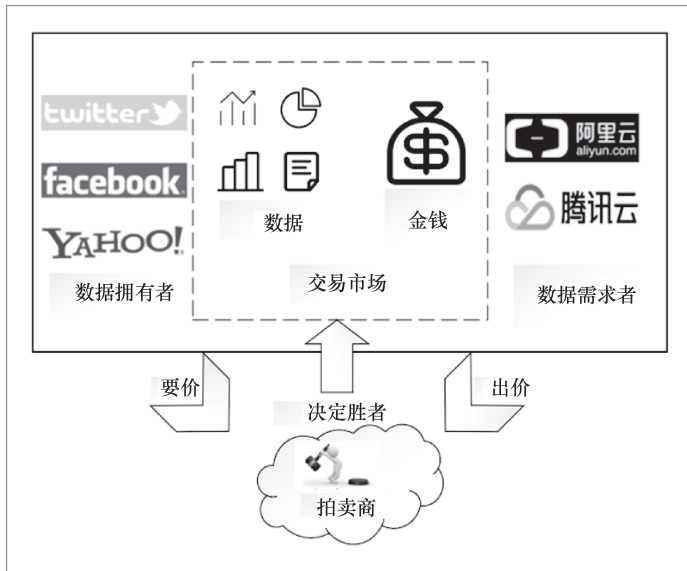


图1 基于拍卖的大数据交易市场的典型框架^[56]

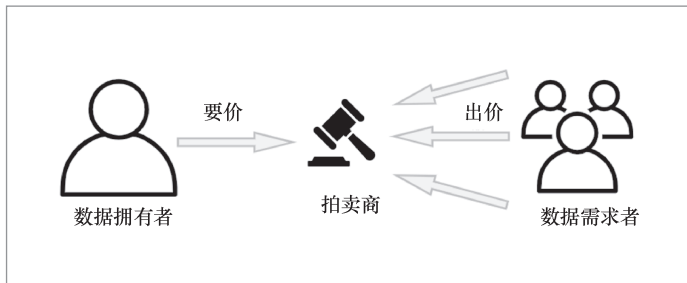


图2 单边拍卖的市场结构

最终的成交价为 $p_F = \max_{p \in P} p$ ，其中P是投标方的竞价集。第二价格密封拍卖又被称为Vickrey拍卖，投标方同样以密封的形式独立出价，获胜的投标方是出价最高的人，获胜者需要支付第二高的价格来赢得拍卖，即 $p_F = \max_{p \in P} p$ ，其中 p_i 是获胜者所出的最高价格。

虽然第一价格密封拍卖机制在现实中的应用很多，但是它也有许多缺点。首先，投标方很难计算出应如何报价；其次，卖方或者拍卖的设计方很难预测拍卖中会发生什么，难以保证投标方之间不结成联盟，信息不对等的情况会阻碍公平的实现；最后，拍卖的获胜者可能会出现“赢者诅咒”的现象。

在第二价格密封拍卖中，每个投标方的占优策略是使出价等于自己对这件商品的完全估价，也就是说，对于投标方而言，此时诚实才是最好的竞拍策略。因此投标方报出的价格是真实且可信的，并且保证了最终的拍卖品归属于出价最高者，因此该密封拍卖是具有帕累托效率的。理论上讲，第二价格密封拍卖是理想化的，满足以下条件。

- 强动机保证：第二价格密封拍卖是占优策略激励相容的。
- 高性能保证：第二价格密封拍卖机制能够使社会福利最大化。
- 计算高效：第二价格密封拍卖可以在输入量的多项式时间内实施。

但是，在市场中实现第二价格密封拍卖是困难的，如投标方之间形成联盟、拍卖商和投标方之间串通、拍卖商的品德败坏等问题都会影响这种拍卖方式的有效性。

(2) VCG拍卖

VCG (Vickrey-Clarke-Groves) 拍卖是Vickrey拍卖的广义形式，是将单物品拍卖形式的Vickrey拍卖一般化为多物品拍卖的形式^[57]。假设市场中有M个商品需要

出售,表示为 $T=\{t_1,t_2,\dots,t_M\}$, t_i 表示第 i 件商品,以及市场中有 N 个投标方,表示为 $B=\{b_1,b_2,\dots,b_N\}$, b_i 表示第 i 个投标方。VCG拍卖背后的原理为获胜者需补偿其获得物品所造成的其他 $N-1$ 个投标方的社会价值损失。例如,在Vickrey拍卖中,获胜者对其他投标方造成的社会价值的损失等于出价第二高的竞价(假定投标方都是真实出价),而这恰好是Vickrey拍卖的支付规则。因此,在VCG拍卖中,根据VCG拍卖规则,如果 $v_i(t_j)$ 是最高的,则投标方 b_i 获得商品 t_j ,投标方 b_i 需要支付的价格为:

$$P=V_{N\setminus\{b_i\}}^M - V_{N\setminus\{b_i\}}^M\{t_j\} \quad (2)$$

其中, $v_i(t_j)$ 表示投标方 b_i 对商品 t_j 的出价, V_N^M 表示由 M 件商品创造的社会价值。VCG拍卖的结果是贝叶斯纳什均衡的^[58]。

尽管理论上VCG拍卖能够使投标方“说真话”,且能够实现社会福利最大化,但是VCG拍卖机制在实际的运用中存在一些障碍,如偏好获取困难,计算福利最大化结果或者近似福利最大化结果十分困难,以及VCG拍卖机制的收益和激励可能表现不好。

(3) 组合拍卖

在大数据市场中,数据需求者对数据的需求是多样化的,单一的数据可能无法满足其需求,且一般需要大量的数据才能达到其要求。在这种市场需求下,组合拍卖的方式被提出。在这种方式下,卖家提供灵活的可捆绑销售的多种物品的商品组合^[59],投标方提出自己的需求及报价,而拍卖商根据竞价中包含的约束条件和卖方的物品分配能力,找到最优的分配,从而确定拍卖中最终胜利的卖家。与上述密封竞价等标准的拍卖方式相比,组合拍卖的优势是经济效率高、买家收益最大化、卖家收益最大化等。然而,在组合拍卖中,计算拍卖最终的获胜者是一个NP问题,这意味

着在现实世界中很难找到最优的分配。

上述各种拍卖方式在实际应用中均有优缺点,而第一价格密封拍卖和第二价格密封拍卖由于具有简单和隐私保护等特点,相较于其他的拍卖方式,在实际中应用得更加广泛^[60]。参与拍卖的各方会根据需求的不同采取不同的拍卖方式。

首先,为了尽可能地保护隐私,以及促使数据拥有者将数据分享出来,需要一定的奖励机制。而差分隐私提供了衡量隐私的方式,如隐私的损失可以被量化,因此隐私被视为一种可替代的商品,其价格可以通过拍卖的方式确定^[61-63]。

其次,在数据交易市场中,“聪明的”投标方可能会操纵拍卖的过程,从而导致不真实的竞价。特别地,投标方可以通过创建多个身份来影响拍卖的结果,这被称为假名投标(false-name bids)^[64]。为了解决这个问题,An D等人^[56]提出了一种多轮假名验证拍卖(multi-round false-name proof auction, MFPA)的方式,从而构建了一个真实的数据交易市场。

在初始的基于拍卖的数据定价方面的研究中^[54,65-66],大多数拍卖中由数据拥有者充当拍卖商的角色,在卖家自身运营的平台上进行销售,在拍卖过程中,来自投标方的信息向卖家开放。另外,随着数据拥有者的增加,由每个数据拥有者自行维护拍卖平台的方式效率低下、不方便。因此,在网络环境中建立第三方数据交易平台是一种更加现实的方法。Gao W C等人^[67]提出由客观的第三方中间商平台充当拍卖商,这样不仅保证了客观性,也防止拍卖商和数据拥有者或数据需求者之间达成某种协议。此外,Gao W C等人^[67]解决了信息物理系统(cyber-physical system, CPS)中的隐私保护问题。他们提出了一种通用的隐私保护拍卖方案(privacy-preserving auction scheme, PPAS),其中拍卖商和

中间平台两个独立实体组成了一个不可信的第三方交易平台。通过实现同态加密和一次性填充,可以确定拍卖过程中的赢家,并对所有竞价信息进行伪装,并且他们设计的系统不基于某种特定的拍卖方式,故适用于各种拍卖规则。

3.2.2 双边拍卖

双边拍卖是现实世界中常用的拍卖方式之一,它被广泛应用于证券交易所^[68]、智能电网^[54,69]等场景。在双边拍卖中,多个数据拥有者和数据需求者同时向拍卖商提交自己的要价和出价,其市场结构如图3所示。

当且仅当数据拥有者的要价 p_o 小于等于数据需求者的出价 p_c 时,拍卖商决定最后的结算价格 p 。拍卖的过程通常如下,每一个卖家(即数据拥有者)对于自己所要销售的物品都有一个边际成本 MC ,卖家的收益是产品的价格 P 与边际成本的差值 $(P-MC)$,卖家的目标是获得尽可能多的收益,即价格尽可能高,但是同时又需要与市场中的其他卖家竞争,若竞争失败,则卖家的收入为0。每一个买家(即数据需求者)对于产品都有一个边际收益 MR ,买家的收益是边际收益与产品价格的差值 $(MR-P)$,买家的目标是获得尽可能多的收益,即价格尽可能低,但是又需要与市场中的其他买家竞争,若竞争失败,则买家的收益为0。对

于卖家而言,每一轮的要价都必须低于上一轮;而对于买家而言,每一轮的出价都必须高于上一轮。市场中买家和卖家的数量足够多时,产品价格的形成如图4所示。

Jiao Y T等人^[70]首先根据数据量大小对大数据分析性能的影响定义了数据成本和效用,然后提出真实、合理、计算效率高的贝叶斯利润最大化拍卖模型。通过求解利润最大化拍卖模型,得到最优服务价格和数据量,从而解决服务商的利润最大化问题。参考文献[70]是一篇引入服务提供商角色的文献。服务提供商(即中间商)负责收集数据拥有者的数据,并对数据拥有者进行隐私补偿,同时利用自身的专业性对收集的大量数据进行处理,满足数据消费者的需求,其为买家提供的是服务而不是原始数据。但是该参考文献假定市场中的服务提供商是垄断者,这在现实场景中较难实现,不满足客观需求。而Cao X Y等人^[71]着眼于数据交易与市场交易效率的协调,提出了一种迭代拍卖机制来协调交易,并以社会福利最大化为目标。在参考文献[71]中,数据拥有者与数据消费者直接发生交易,交易的是原始数据。然而上述激励机制忽视了数据消费者之间既有偏好又有复杂的利益冲突关系这一事实。Cai H等人^[72]提出了DTPCI这种双边拍卖机制来应对市场偏好的多样性、数据消费者之间复杂的利益冲突关系以及交易双方的战略选择这三大挑战。上述研究大多基于集中式的数据交易市场,存在集中式数据交易市场常有的问题,如数据拥有者的隐私泄露、数据丢失等问题。而Li Z N等人^[34]则聚焦于基于分布式的数据交易市场,并且提出了具有稀疏信息的经纪人来管理和调整交易市场,采用了一种迭代双边拍卖的方案。该方案由经纪人解决计算资源交易的分配问题,并设计了一个特定的价格规则来诱导计算资源的买家和卖家以真实的方式投标,从而实现了社会福利最大化,并

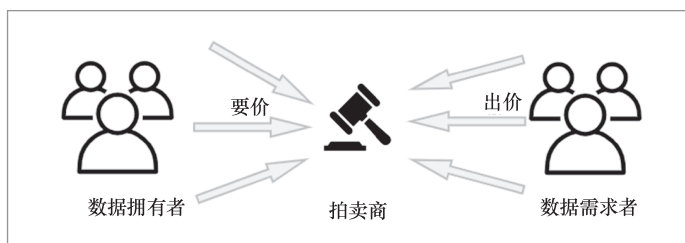


图3 双边拍卖的市场结构

且实现了安全的数据交易。

关于拍卖市场研究的进一步分析见表1。

4 数据交易市场

数据定价和数据交易是一个相辅相成的过程。因为数据具有商业价值，所以大数据市场以及数据交易市场成为实现数据共享的最有效的市场。第3节总结了一些常见的基于博弈论和拍卖的数据定价模型，本节将系统地研究数据交易平台，并基于现有的研究将数据交易市场划分为集中式数据交易市场和分布式数据交易市场。

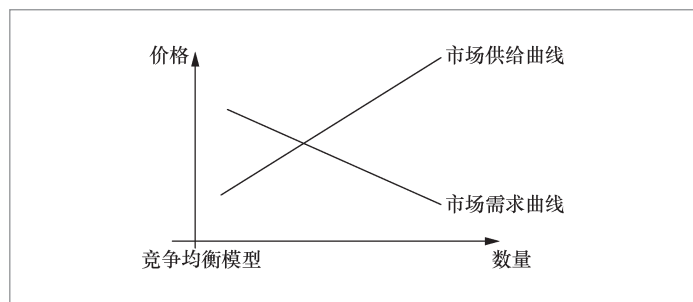


图4 双边拍卖市场价格的形成

4.1 集中式数据交易市场

纵观关于数据交易的研究，根据是否有第三方参与，可以将数据交易划分为有中间商的数据交易市场^[16,21,43,75]和无中间

表1 不同拍卖机制在数据定价中的应用

参考文献	定价模型	市场结构			概述	适用场景
		卖家	中间商	买家		
[73]	VCG拍卖	一个或多个	一个	多个	服务提供商可以由卖家自身充当，也可以由多个卖家选定一个中间商充当，服务提供商根据收到的投标价格提供不同粒度的组合服务，买家根据服务提供商提供的组合，在满足质量约束和最小化社会成本的情况下，选择自己需要的组合	最小化买家的成本、“讲真话”
[10]	组合拍卖	多个	无	多个	设计一个数据市场和一个强大的实时匹配机制，有效地购买和出售机器学习任务的训练数据	公平、真实的零遗憾机制
[67]	可适用多种拍卖规则	一个或多个	一个	多个	提出了一种通用的隐私保护拍卖方案，其中拍卖商和中间平台两个独立实体组成了一个不可信的第三方交易平台。通过同态加密和一次性填充，可以确定拍卖过程中的赢家，并对所有竞价信息进行伪装，解决了CPS中的隐私保护问题	隐私保护拍卖、安全性
[70]	双边拍卖	多个	一个	多个	首先根据数据量大小对大数据分析性能的影响定义了数据成本和效用，然后提出真实、合理、计算效率高的贝叶斯利润最大化拍卖。通过求解利润最大化拍卖，得到最优服务价格和数据量，解决了服务商的利润最大化问题。服务提供商收集卖家的数据，并对卖家进行隐私补偿，同时利用自身的专业性对收集的大量数据进行处理，以满足买家的需求，为买家提供的是服务而不是原始数据	中间商收益最大化
[71]	双边拍卖	多个	无	多个	提出了一种迭代拍卖机制来协调交易，以社会福利最大化为目标。其中，卖家与买家直接发生交易，交易的是原始数据	社会福利最大化
[74]	第二价格密封拍卖+VCG拍卖	一个或多个	一个	多个	因为在多赢家拍卖策略中，传统的VCG拍卖可能会减少卖家的收益且容易受到“合谋”攻击，所以修正了第二价格密封拍卖，使在多赢家拍卖策略中能够解决上述问题。修改了多赢家拍卖策略中的优化问题，并且拍卖商可以根据标准选择最终的赢家	“讲真话”、社会福利最大化

商的数据交易市场^[10,54,65-66,76],根据为买家提供的服务可以划分为买家提供原始数据的数据交易^[10,16,21,71]和为买家提供处理后的数据(如对数据加入噪声从而保护数据所有者的隐私,或经预处理直接获得满足客户需求的数据)的数据交易^[43,75]。关于集中式数据交易市场的进一步分析见表2。

在没有中间商作为中介的框架中,数据拥有者和数据需求者直接进行数据交易,这种交易方式可以避免为第三方支付交易费用,可以让交易双方更了解双方的需求,从而降低交易价格。然而在没有第

三方平台的交易市场中,数据拥有者需要自己搭建数据交易市场,并且由于市场中存在多个数据拥有者,卖方之间的数据可能会形成“孤岛”;而且,对于买家而言,由于信息不对称,从多个数据拥有者的交易平台中分别购买数据会消耗大量的精力和金钱,且存在购买数据与需求不匹配的可能,从而降低买家购买的欲望。为了解决上述难题,现有的关于数据交易市场的框架大多存在一个或多个中间商充当第三方平台。但是该框架在解决上述难题的同时也带来了新的问题,如第三方平台的信任

表2 集中式市场中各种定价模型的应用

参考文献	定价模型	市场结构			提供的服务	概述	适用场景
		卖家	中间商	买家			
[10]	基于拍卖的数据定价	多个	无	多个	原始数据	设计一个数据市场和一个强大的实时匹配机制,以有效地购买和出售机器学习任务的训练数据	公平、真实的零遗憾机制
[31]	Stackelberg 博弈	多个	无	多个	原始数据+处理后的服务	对原始数据和提供的服务提供捆绑销售服务;使用Stackelberg博弈模型来确保服务,然后和数据消费者协商数据的价格	收益最大化
[16]	Stackelberg 博弈	一个	少数	多个	原始数据	针对纯捆绑定价和分开定价两种情况,讨论了在数据销售者和市场代理商之间搭建Stackelberg博弈模型(其中数据销售者为领导者,中间商为追随者)、实现双方利益最大化的条件下,数据销售者应该采取何种定价方式使得自身利益最大化	利益最大化
[71]	基于拍卖的数据定价	多个	无	多个	原始数据	提出了一种迭代拍卖机制来协调交易,以社会福利最大化为目标,其中,数据拥有者与数据消费者直接发生交易,交易的是原始数据	社会福利最大化
[21]	基于查询的数据定价	多个	多个	一个	原始数据	传统的基于查询的数据定价集中于无套利定价方面,本模型提供了简洁的定价函数,并且实现了收益最大化	中间商收益最大化、无套利定价
[22]	基于模型的定价	一个或多个	一个或多个	一个或多个	拥有不同精度的机器学习模型	提出了一种基于模型的定价框架MBP,这个框架不销售原始数据,销售的是不同精度的机器学习模型,该框架最大的属性是无套利	无套利定价
[43]	讨价还价模型	多个	一个	多个	加入噪声后的原始数据	引入了一个差分隐私数据市场框架PRIVATA,这个框架通过讨价还价的定价机制,决定隐私 ϵ 的单位价格和单位价值,从而保证数据消费者和数据拥有者之间的地位公平	隐私保护、公平
[75]	非合作博弈	多个	多个	多个	对数据进行处理后提供服务	引入了一个数据-价格比率的参数,证明了当该参数大于1并且服务提供者有限时,市场中存在唯一有意义的纳什均衡解	充分竞争市场

危机以及数据产品接近0的复制成本,进而导致数据丢失、数据被非法篡改、用户隐私泄露等潜在的问题。

4.2 分布式数据交易市场

传统的数据交易是在一个集中式的市场中进行的,这就需要可靠的权威机构负责数据交换。在这项工作中,为了解决潜在的问题,将区块链技术^[77]应用于数据交易市场,从而构建去中心化的数据交易市场。去中心化的数据交易市场可以摆脱单点性能瓶颈和单点故障问题,并且允许参与数据交易的多方之间建立去中心化的信任和依赖,其吸引了越来越多的关注。区块链是一种去中心化的分布式数据存储技术,其特点是防篡改、可跟踪性和公共可访问性,其中共识机制确保了同步。为此,目前的研究设计了一个联合体区块链网络,以创建P2P(peer-to-peer)交易市场,提高数据所有者节点、服务提供商节点、数据消费者节点和矿工之间数据交易的安全性。在数据市场中,区块链的引入可以解决集中式数据市场的弊端,为实现数据交易的去中心化提供了另一种可能。

数据交易属于在线的商业交易,参与者之间存在固有的不信任问题,这是因为没有任何先验的交易在前。而这种初始的不信任可能会导致数据交易陷入僵局,参与者没有动力进行交易。因此Delgado-Segura S等人^[78]提出了一个公平的比特币数据交易的基础协议,建立了一个公平的数据交易市场。因为该协议是原子性的,所以该交易过程可以随时完成或终止,以确保数据提供商和数据消费者都没有遭受损失。但是该数据交易市场的缺点是不能辨别虚假信息,且一次只能考虑一个交易流程。

Liu K等人^[37]利用区块链技术建立物

联网数据市场体系,提出了一个边缘/云计算辅助的、区块链增强的框架,以应对物联网数据市场中安全、信任和效率的挑战;并在其中搭建了两阶段的Stackelberg博弈,以解决数据需求者的定价和购买问题,为物联网数据市场设计了最优定价机制,从而最大化赢家和消费者的利润。

此外,Chuang I H等人^[76]提出了一种信任感知物联网数据经济系统(trust-aware IoT data economic system, TIDES)。Yu Y等人^[79]提出了一种基于比特币的用于物联网数据交易的加密货币LRCoin。Zhou J Y等人^[80]在区块链上结合数据嵌入和相似度学习提出了分布式数据销售,这种方法权衡了数据检索的有效性和数据索引的泄漏风险之间的关系。汪靖伟等人^[81]提出搭建一个基于区块链系统的数据交易市场分别需要考虑如下几个设计目标:去中心化、公平性、隐私性、有效性和经济激励。

尽管基于区块链的数据定价可以解决大多数集中式数据定价市场带来的潜在问题,但是基于区块链的数据定价也面临许多难题,如固有的物理约束、有限的计算和存储资源限制了使用更复杂的应用程序的机会。为了解决计算和存储资源有限的问题,Li Z N等人^[34]通过引入边缘云计算来增强基于区块链的数据定价,提供了强大的计算资源和可扩展性。其中边缘提供了有限的低时延的计算和存储资源,而云计算可以提供强大的计算和存储资源,但是缺点是时延较高。而Li Z N等人^[34]将二者结合,并且采用了非合作博弈的模型,更加符合现实的数据交易市场。

尽管基于区块链的去中心化数据交易系统可以支持链上数据的安全性,但是需要更多努力来确保用户收集并记录在区块链上的链下数据的可信性,即分布式数据市场也带来了隐私和安全问题。

5 总结

本文研究了数据定价的相关问题。具体来说,首先分析了数据作为商品相较于传统的实物产品的特性与区别,并且回顾了常见的数据产品定价策略和定价准则。对于数据定价以及建立一个高效的数据交易市场,明确了其重要性,对不同的市场结构、数据定价策略以及数据定价模型进行了分类,并指出了各种类型的优势和局限性。然后研究了博弈论和拍卖策略,详细介绍了不同的方案、交易平台及相关问题,并对现有的基于博弈论和拍卖的数据定价模型进行了回顾,分析了这些研究对数据定价的贡献及局限性。最后讨论了传统的集中式数据交易市场和近年热门的基于区块链的数据定价方式,并讨论了两种数据交易市场的优劣,分析了这两种交易平台未来的改进方向。本文旨在让大家了解常见的数据定价的方式和方法,以及数据定价领域存在的尚未解决的难题,以期进一步推动大数据交易的发展。

参考文献:

- [1] LIN J, YU W, ZHANG N, et al. A survey on Internet of things: architecture, enabling technologies, security and privacy, and applications[J]. IEEE Internet of Things Journal, 2017, 4(5): 1125–1142.
- [2] SUN Y C, SONG H B, JARA A J, et al. Internet of things and big data analytics for smart and connected communities[J]. IEEE Access, 2016, 4: 766–773.
- [3] STANKOVIC J A. Research directions for the Internet of things[J]. IEEE Internet of Things Journal, 2014, 1(1): 3–9.
- [4] WU J J, ZHAO W. Design and realization of WInternet: from net of things to Internet of things[J]. ACM Transactions on Cyber-Physical Systems, 2016, 1(1): 1–12.
- [5] YANG X Y, REN X B, LIN J, et al. On binary decomposition based privacy-preserving aggregation schemes in real-time monitoring systems[J]. IEEE Transactions on Parallel and Distributed Systems, 2016, 27(10): 2967–2983.
- [6] OLIVEIRA A C, FETZER C, MARTIN A, et al. Optimizing query prices for data-as-a-service[C]//Proceedings of the 2015 IEEE International Congress on Big Data. Piscataway: IEEE Press, 2015: 289–296.
- [7] LIANG F, YU W, AN D, et al. A survey on big data market: pricing, trading and protection[J]. IEEE Access, 2018, 6: 15132–15154.
- [8] GOLDFARB A, TUCKER C. Digital economics[J]. Journal of Economic Literature, 2019, 57(1): 3–43.
- [9] CHAWLA S, DEEP S, KOUTRISW P, et al. Revenue maximization for query pricing[J]. Proceedings of the VLDB Endowment, 2019, 13(1): 1–14.
- [10] AGARWAL A, DAHLEH M, SARKAR T. A marketplace for data: an algorithmic solution[C]//Proceedings of the 2019 ACM Conference on Economics and Computation. New York: ACM Press, 2019: 701–726.
- [11] MUSCHALLE A, STAHL F, LÖSER A, et al. Pricing approaches for data markets[C]//Proceedings of the 2012 International Workshop on Business Intelligence for the Real-Time Enterprise. Heidelberg: Springer, 2012: 129–144.
- [12] KANTERE V, DASH D, GRATSIAS G, et al. Predicting cost amortization for query services[C]//Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data. New York: ACM Press, 2011: 325–336.

- [13] NIYATO D, ALSHEIKH M A, WANG P, et al. Market model and optimal pricing scheme of big data and Internet of things (IoT)[C]//Proceedings of the 2016 IEEE International Conference on Communications. Piscataway: IEEE Press, 2016: 1–6.
- [14] LUONG N C, HOANG D T, WANG P, et al. Data collection and wireless communication in Internet of things (IoT) using economic analysis and pricing models: a survey[J]. *IEEE Communications Surveys & Tutorials*, 2016, 18(4): 2546–2590.
- [15] LYU X Y, ZHANG R T, YUE J. Competition and cooperation between participants of the Internet of things industry value chain[J]. *Advances in Information Sciences & Service Sciences*, 2012, 4(11): 406–412.
- [16] MEI L J, LI W, NIE K. Pricing decision analysis for information services of the Internet of things based on Stackelberg game[C]//Proceedings of the 2nd International Conference on Logistics, Informatics and Service Science. Heidelberg: Springer, 2013: 1097–1104.
- [17] JUNG K, PARK S. Privacy bargaining with fairness: privacy–price negotiation system for applying differential privacy in data market environments[C]//Proceedings of the 2019 IEEE International Conference on Big Data. Piscataway: IEEE Press, 2019: 1389–1394.
- [18] JANG B, PARK S, LEE J, et al. Three hierarchical levels of big–data market model over multiple data sources for Internet of things[J]. *IEEE Access*, 2018, 6: 31269–31280.
- [19] KOUTRIS P, UPADHYAYA P, BALAZINSKA M, et al. Query–based data pricing[J]. *Journal of the ACM*, 2015, 62(5): 1–44.
- [20] LIN B R, KIFER D. On arbitrage–free pricing for general data queries[J]. *Proceedings of the VLDB Endowment*, 2014, 7(9): 757–768.
- [21] CHAWLA S, DEEP S, KOUTRIS P, et al. Revenue maximization for query pricing[J]. *arXiv preprint*, 2019, arXiv:1909.00845.
- [22] CHEN L, KOUTRIS P, KUMAR A. Towards model–based pricing for machine learning in a data marketplace[C]//Proceedings of the 2019 International Conference on Management of Data. New York: ACM Press, 2019: 1535–1552.
- [23] CHEN L, KOUTRIS P, KUMAR A. Model–based pricing: do not pay for more than what you learn![C]//Proceedings of the 1st Workshop on Data Management for End–to–End Machine Learning. New York: ACM Press, 2017: 1–4.
- [24] PEI J. A survey on data pricing: from economics to data science[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2020.
- [25] GOLDBERG A V, HARTLINE J D. Envy–free auctions for digital goods[C]//Proceedings of the 4th ACM Conference on Electronic Commerce. New York: ACM Press, 2003: 29–35.
- [26] CRETTEZ B. On Hobbes’s state of nature and game theory[J]. *Theory and Decision*, 2017, 83(4): 499–511.
- [27] WU Q, WANG F, ZHOU L, et al. Method of multiple attribute group decision making based on 2–dimension interval type–2 fuzzy aggregation operators with multi–granularity linguistic information[J]. *International Journal of Fuzzy Systems*, 2017, 19(6): 1880–1903.
- [28] SOLMEYER N, DIXON R, BALU R. Characterizing the Nash equilibria of a three–player Bayesian quantum game[J]. *Quantum Information Processing*, 2017, 16(6): 146.
- [29] 张维迎. 博弈论与信息经济学[M]. 上海: 上海人民出版社, 2012.
- ZHANG W Y. *Game theory and information economics*[M]. Shanghai:

- Shanghai People's Publishing House, 2012.
- [30] NASH J F. Non-cooperative games[M]. Princeton: Princeton University Press, 2016.
- [31] NIYATO D, HOANG D T, LUONG N C, et al. Smart data pricing models for the Internet of things: a bundling strategy approach[J]. IEEE Network, 2016, 30(2): 18–25.
- [32] ZHENG Z Z, PENG Y Q, WU F, et al. An online pricing mechanism for mobile crowdsensing data markets[C]// Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing. New York: ACM Press, 2017: 1–10.
- [33] LI C, LI D Y, MIKLAU G, et al. A theory of pricing private data[J]. ACM Transactions on Database Systems, 2014, 39(4): 1–28.
- [34] LI Z N, YANG Z Y, XIE S L. Computing resource trading for edge-cloud-assisted Internet of things[J]. IEEE Transactions on Industrial Informatics, 2019, 15(6): 3661–3669.
- [35] SIMAAN M, CRUZ J B. On the Stackelberg strategy in nonzero-sum games[J]. Journal of Optimization Theory and Applications, 1973, 11(5): 533–555.
- [36] HADDADI S, GHASEMI A. Pricing-based Stackelberg game for spectrum trading in self-organised heterogeneous networks[J]. IET Communications, 2016, 10(11): 1374–1383.
- [37] LIU K, QIU X Y, CHEN W H, et al. Optimal pricing mechanism for data market in blockchain-enhanced Internet of things[J]. IEEE Internet of Things Journal, 2019, 6(6): 9748–9761.
- [38] XU C Z, ZHU K, YI C Y, et al. Data pricing for blockchain-based car sharing: a Stackelberg game approach[C]// Proceedings of the 2020 IEEE Global Communications Conference. Piscataway: IEEE Press, 2020: 1–5.
- [39] KANG X, ZHANG R, MOTANI M. Price-based resource allocation for spectrum-sharing femtocell networks: a Stackelberg game approach[J]. IEEE Journal on Selected Areas in Communications, 2012, 30(3): 538–549.
- [40] YAO H, MAI T, WANG J, et al. Resource trading in blockchain-based industrial Internet of things[J]. IEEE Transactions on Industrial Informatics, 2019, 15(6): 3602–3609.
- [41] RAWAT D B, SHETTY S, XIN C. Stackelberg-game-based dynamic spectrum access in heterogeneous wireless systems[J]. IEEE Systems Journal, 2014, 10(4): 1494–1504.
- [42] ZHANG T, ZHU Q. On the differential private data market: endogenous evolution, dynamic pricing, and incentive compatibility[J]. arXiv preprint, 2021, arXiv:2101.04357.
- [43] JUNG K, LEE J, PARK K, et al. PRIVATA: differentially private data market framework using negotiation-based pricing mechanism[C]// Proceedings of the 28th ACM International Conference on Information and Knowledge Management. New York: ACM Press, 2019: 2897–2900.
- [44] NIU C, ZHENG Z, WU F, et al. Unlocking the value of privacy: trading aggregate statistics over private correlated data[C]// Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2018: 2031–2040.
- [45] YAO L, JIA Y, ZHANG H, et al. A decentralized private data transaction pricing and quality control method[C]// Proceedings of the 2019 IEEE International Conference on Communications. Piscataway: IEEE Press, 2019: 1–5.
- [46] MOULIK S, MISRA S, GAURAV A. Cost-effective mapping between wireless body area networks and cloud service providers

- based on multi-stage bargaining[J]. IEEE Transactions on Mobile Computing, 2016, 16(6): 1573–1586.
- [47] MAO Y, CHENG T, ZHAO H, et al. A strategic bargaining game for a spectrum sharing scheme in cognitive radio-based heterogeneous wireless sensor networks[J]. Sensors, 2017, 17(12): 2737.
- [48] AZIMI S M, MANSHAEEI M H, HENDESSI F. Cooperative primary-secondary dynamic spectrum leasing game via decentralized bargaining[J]. Wireless Networks, 2016, 22(3): 755–764.
- [49] MCAFEE R P. A dominant strategy double auction[J]. Journal of Economic Theory, 1992, 56(2): 434–450.
- [50] 李三希, 王泰著. 拍卖理论研究述评[J]. 中国科学基金, 2021, 35(1): 2–3.
- LI S X, WANG T M. A review of auction theory[J]. Bulletin of National Natural Science Foundation of China, 2021, 35(1): 2–3.
- [51] 许永国. 拍卖经济理论综述[J]. 经济研究, 2002(9): 84–91, 95.
- XU Y G. An overview of auction theory[J]. Economic Research Journal, 2002(9): 84–91, 95.
- [52] AN D, YANG Q, YU W, et al. Sto2Auc: a stochastic optimal bidding strategy for microgrids[J]. IEEE Internet of Things Journal, 2017, 4(6): 2260–2274.
- [53] IOSIFIDIS G, GAO L, HUANG J, et al. An iterative double auction for mobile data offloading[C]//Proceedings of the 2013 11th International Symposium and Workshops on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks. Piscataway: IEEE Press, 2013: 154–161.
- [54] AN D, YANG Q, YU W, et al. SODA: strategy-proof online double auction scheme for multimicrogrids bidding[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2017, 48(7): 1177–1190.
- [55] DAS S R, SUNDARAM R K. Auction theory: a summary with applications to treasury markets1[Z]. 1996.
- [56] AN D, YANG Q Y, YU W, et al. Towards truthful auction for big data trading[C]//Proceedings of 2017 IEEE 36th International Performance Computing and Communications Conference. Piscataway: IEEE Press, 2017: 1–7.
- [57] AUSUBEL L M, MILGROM P. The lovely but lonely Vickrey auction[J]. Combinatorial Auctions, 2006, 17: 22–26.
- [58] LUCIER B, PAES LEME R, TARDOS É. On revenue in the generalized second price auction[C]//Proceedings of the 21st International Conference on World Wide Web. New York: ACM Press, 2012: 361–370.
- [59] BLUMROSEN L, NISAN N. Combinatorial auctions[J]. Algorithmic Game Theory, 2007, 267: 300.
- [60] SUZUKI K, KOBAYASHI K, MORITA H. Efficient sealed-bid auction using hash chain[C]//Proceedings of the International Conference on Information Security and Cryptology. Heidelberg: Springer, 2000: 183–191.
- [61] GHOSH A, ROTH A. Selling privacy at auction[C]//Proceedings of the 12th ACM Conference on Electronic Commerce. New York: ACM Press, 2011: 199–208.
- [62] FLEISCHER L K, LYU Y H. Approximately optimal auctions for selling privacy when costs are correlated with data[C]//Proceedings of the 13th ACM Conference on Electronic Commerce. New York: ACM Press, 2012: 568–585.
- [63] DANDEKAR P, FAWAZ N, IOANNIDIS S. Privacy auctions for recommender systems[J]. ACM Transactions on Economics and Computation, 2014, 2(3): 1–22.
- [64] WANG Q, YE B, TANG B, et al. Robust large-scale spectrum auctions against false-name bids[J]. IEEE Transactions on Mobile Computing, 2016, 16(6): 1730–1743.
- [65] NADENDLA V S S, BRAHMA S K, VARSHNEY P K. Optimal spectrum auction design with 2-D truthful

- revelations under uncertain spectrum availability[J]. *IEEE/ACM Transactions on Networking*, 2016, 25(1): 420–433.
- [66] FENG Z N, ZHU Y M, ZHANG Q, et al. TRAC: truthful auction for location-aware collaborative sensing in mobile crowdsourcing[C]//*Proceedings of the 2014 IEEE Conference on Computer Communications*. Piscataway: IEEE Press, 2014: 1231–1239.
- [67] GAO W C, YU W, LIANG F, et al. Privacy-preserving auction for big data trading using homomorphic encryption[J]. *IEEE Transactions on Network Science and Engineering*, 2018, 7(2): 776–791.
- [68] YANG J. The efficiency of an artificial double auction stock market with neural learning agents[M]//*Evolutionary computation in economics and finance*. Heidelberg: Springer, 2002: 85–105.
- [69] LI D, YANG Q, YU W, et al. Towards double auction for assisting electric vehicles demand response in smart grid[C]//*Proceedings of the 2017 13th IEEE Conference on Automation Science and Engineering*. Piscataway: IEEE Press, 2017: 1604–1609.
- [70] JIAO Y T, WANG P, NIYATO D, et al. Profit maximization auction and data management in big data markets[C]//*Proceedings of the 2017 IEEE Wireless Communications and Networking Conference*. Piscataway: IEEE Press, 2017: 1–6.
- [71] CAO X Y, CHEN Y, LIU K J R. Data trading with multiple owners, collectors, and users: an iterative auction mechanism[J]. *IEEE Transactions on Signal and Information Processing over Networks*, 2017, 3(2): 268–281.
- [72] CAI H, ZHU Y M, LI J, et al. Double auction for a data trading market with preferences and conflicts of interest[J]. *The Computer Journal*, 2019, 62(10): 1490–1504.
- [73] WU Q, ZHOU M C, ZHU Q, et al. VCG auction-based dynamic pricing for multigranularity service composition[J]. *IEEE Transactions on Automation Science and Engineering*, 2017, 15(2): 796–805.
- [74] SOFIA D S, EDWARD A S. Auction based game theory in cognitive radio networks for dynamic spectrum allocation[J]. *Computers & Electrical Engineering*, 2020, 86: 106734.
- [75] GUIJARRO L, PLA V, VIDAL J R, et al. Competition in data-based service provision: Nash equilibrium characterization[J]. *Future Generation Computer Systems*, 2019, 96: 35–50.
- [76] CHUANG I H, HUANG S H, CHAO W C, et al. TIDES: a trust-aware IoT data economic system with blockchain-enabled multi-access edge computing[J]. *IEEE Access*, 2020, 8: 85839–85855.
- [77] NAKAMOTO S. Bitcoin: a peer-to-peer electronic cash system[Z]. 2008.
- [78] DELGADO-SEGURA S, PÉREZ-SOLÀ C, NAVARRO-ARRIBAS G, et al. A fair protocol for data trading based on Bitcoin transactions[J]. *Future Generation Computer Systems*, 2020, 107: 832–840.
- [79] YU Y, DING Y J, ZHAO Y Q, et al. LRCoin: leakage-resilient cryptocurrency based on Bitcoin for data trading in IoT[J]. *IEEE Internet of Things Journal*, 2018, 6(3): 4702–4710.
- [80] ZHOU J Y, TANG F Y, ZHU H, et al. Distributed data vending on blockchain[C]//*Proceedings of the 2018 IEEE International Conference on Internet of Things and IEEE Green Computing and Communications and IEEE Cyber, Physical and Social Computing and IEEE Smart Data*. Piscataway: IEEE Press, 2018: 1100–1107.
- [81] 汪靖伟, 郑臻哲, 吴帆, 等. 基于区块链的数据市场[J]. *大数据*, 2020, 6(3): 21–35.
- WANG J W, ZHENG Z Z, WU F, et al. Blockchain based data marketplace[J]. *Big Data Research*, 2020, 6(3): 21–35.

作者简介



张小伟 (1996-), 男, 东北大学计算机科学与工程学院硕士生, 主要研究方向为数据定价。



江东 (1996-), 男, 东北大学计算机科学与工程学院博士生, 主要研究方向为GPU加速图计算、数据定价。



袁野 (1981-), 男, 北京理工大学计算机学院教授、博士生导师, 主要研究方向为大数据管理与分析 (包括图数据管理、众包数据管理、不确定数据管理、数据隐私保护等)。优秀青年科学基金获得者。曾获中国电子学会科学技术奖 (自然科学) 一等奖、全国百篇优秀博士学位论文提名奖、中国计算机学会优秀博士学位论文奖、辽宁省优秀博士学位论文奖。中国计算机学会数据库专业委员会委员、大数据专家委员会委员, 中国计算机学会高级会员, IEEE、ACM高级会员。在ACM SIGMOD、VLDB、ICDE、KDD、SIGIR、IJCAI、VLDB Journal等重要学术会议和期刊上发表论文90余篇, 其中CCF A类40余篇。先后承担和参加国家自然科学基金重点项目和优秀青年科学基金项目、国家重点研发计划、863计划、973计划等多项重点项目。常年担任多个国际数据库会议 (共同/出版/宣传/本地) 主席/程序委员会委员, 包括SIGMOD、VLDB、ICDE等CCF A类会议。

收稿日期: 2021-05-31

通信作者: 袁野, yuanye@mail.neu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.60933001, No.61932004, No.62002054, No.61732003, No.61729201); 中央高校基本科研基金资助项目 (No.N181605012)

Foundation Items: The National Natural Science Foundation of China (No.60933001, No.61932004, No.62002054, No.61732003, No.61729201), The Fundamental Research Funds for the Central Universities(No.N181605012)