

基于VCG机制的个人数据交易机制设计与分析

梁焯然

贵州大学管理学院, 贵州 贵阳

收稿日期: 2024年3月5日; 录用日期: 2024年3月19日; 发布日期: 2024年7月18日

摘要

个人数据作为数字生活的核心要素, 支撑着现代产品与服务的创新与发展。在当前的商业环境中, 个人数据的所有权与利用成为立法与实践关注的焦点。本文旨在通过设计新型的数据交易流程与分配机制, 探索用户对其个人数据的所有权及其在数据市场中的价值实现。基于对现有数据定价模型的分析, 提出了基于VCG机制的个人大数据定价模型, 该模型综合考虑了信息熵、权重值、数据参考指数及成本等因素对数据价值的影响, 实现了对数据价值的动态评估和正向定级。通过改进经典VCG机制引入交割支付规则, 使其满足预算平衡约束, 并激励交易者提供真实信息。通过理性分析与仿真实验得出该新型交易模式在交易速度和资源配置效率上均优于传统模式。

关键词

大数据交易, 机制设计, 定价方法, VCG机制

Design and Analysis of Personal Data Transaction Mechanisms Based on VCG Mechanism

Zhuoran Liang

School of Management, Guizhou University, Guiyang Guizhou

Received: Mar. 5th, 2024; accepted: Mar. 19th, 2024; published: Jul. 18th, 2024

Abstract

As a core element of digital life, personal data supports the innovation and development of modern products and services. In the current business environment, the ownership and utilization of personal data have become a focus of legislative and practical attention. This paper aims to explore users' ownership of their personal data and its value realization in the data market by de-

signing a new type of data transaction process and distribution mechanism. Based on the comparative analysis of existing data pricing models, a personal big data pricing model based on the VCG mechanism is proposed, which comprehensively considers the impact of information entropy, weight value, data reference index, and cost on data value, and achieves dynamic assessment and positive grading of data value. The delivery payment rule is introduced by improving the classical VCG mechanism to make it satisfy the budget balance constraint and incentivize traders to provide real information. Rational analyses and simulation experiments conclude that this new trading model outperforms the traditional model in terms of trading speed and resource allocation efficiency.

Keywords

Big Data Transactions, Mechanism Design, Pricing Methodology, VCG Mechanism

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在当今科学和商业社会中,数据被认为是极其重要的资产。据国际数据公司(IDC)预测,到2025年,全球将有超过557亿台联网设备,其中75%将连接到物联网平台,预计这些设备将产生73.1 ZB的数据[1]。这些数据主要来自于视频监控和工业物联网应用程序。几乎所有公司都在积极采用人工智能(AI)或机器学习(ML)技术,以从这些有价值的数据中获得竞争优势。数据对于这些AI或ML算法的推动至关重要,特别是大规模的数据集。因此,数据交易作为一种方便且具有前景的数据共享方式至关重要,可以释放AI或ML的潜力。数据交易与一般数据共享的概念有所不同。在移动群体感知系统中,工作人员通常会共享他们的感知数据并获得回报[2]。应用程序用户必须无意识地或被迫地向应用程序服务提供商披露其数据记录,例如网络浏览和在线购物订单,以获得访问其服务的权限。数据共享的概念已经扩展到数据交易,使得数据所有者和数据消费者能够主动决定是否参与,并进一步指定他们所需的数据类型,以及通过披露数据所期望获得的回报数量。这种方式有助于解决数据隐私问题和不对称信息问题,并提高数据交易的质量和合规性[2]。因此,在数据交易中加强隐私保护、促进信息对称和确保数据质量与合规性是至关重要的。最近的研究聚焦于数据交易方法的提出。数据交易涉及数据所有者和数据消费者之间的交易,包括原始数据样本、范围计数和聚合统计结果等各种数据集[3]。为了协商数据交易过程,常常引入数据交易平台或数据代理来支持数据交易消息或交易数据的传输。为了保护数据隐私,一些研究采用加密算法对数据进行加密,并向数据消费者披露加密数据;其他研究提出了差分隐私或其变体等隐私保护方案,以控制数据隐私水平。还有一些研究认为数据参与者或数据平台不可信,因此提出了基于区块链技术的去中心化数据交易平台。然而,几乎所有现有的研究都忽视了物联网数据交易中数据所有者和数据请求的动态特性[4]。在物联网交易中,交易是由各种智能物联网设备收集的众包数据来完成的。数据所有者并不总是能够提供数据,因为智能设备有时被其所有者占用,或者由于资源限制(如电池电量不足和间歇性连接)[5]而无法提供数据服务。同时,每个数据请求都是根据数据消费者的需求生成的,这是事先未知的。因此,我们认为数据所有者和数据请求都具有动态属性。在本文中,我们对物联网数据交易场景中数据所有者和数据请求的动态特性进行了建模。例如,当智能手机的拥有者决定进行数据的收集及售卖时,智能手机可以暂时充当数据收集器和数据所有者。因此,我们声称智能设备(即数据所有者)是间歇性的,并且由于资源、移动性或设备所有者占用有限,只能在特定的时间段(称为活动时间)内

交易其数据。此外，数据请求由数据消费者随机生成，然后提交到运行在边缘服务器或云服务器上的数据交易平台[6]。解决众包数据交易问题存在几个主要的技术挑战。

首先，需要允许数据所有者动态加入或离开数据交易过程，并根据应用程序的需求随机生成数据请求[7]。这种不确定和不可预测的数据请求使得数据交易过程相当复杂。其次，数据交易平台很难在数据所有者和数据请求之间进行有效匹配，因为数据包的估值和活跃时间都是数据所有者的私人信息。最后，理性和战略性的数据所有者不愿意提供他们有价值的数据或如实报告他们的私人信息，除非得到适当的补偿。为此，我们提出了一种基于真实在线拍卖的数据交易算法，该算法包含两个关键组件，解决了如何将动态数据所有者与随机生成的数据请求相匹配以及如何确定数据包的交易价格这两个子问题。一方面，构建了一个在线拍卖模型来描述动态数据交易过程[8]。另一方面，每个数据包的交易价格是根据数据交易者个体对于该数据包的价值进行价格生成的，该值是根据数据所有者和数据采购厂商根据往期数据及个人评估得到的真实预估价值。严格的理论分析和广泛的模拟都证明了我们提出的在线数据交易算法的理想性质，例如个体理性、激励相容性和预算平衡性。

本文的主要贡献如下：

- 1) 本文将数据所有者对于自身数据的需求纳入考虑并据此提供一个动态的数据交易机制和分配机制。
- 2) 在考虑到目前数据交易市场并不成熟的交易体系及第三方指导价难以根据用户、企业及价值释放等多方面对价格进行干预[9]，在本文中我们引入 VCG 机制进行商品及交易金额的分配，并以交割支付的支付规则根据交易者个人对本次交易的社会福利贡献比率进行最终交易额的确定，在一定程度上极大地透明化数据交易市场交易金额并揭露各交易者对各数据包的估值，以便第三方机构及市场能更快摸清各类数据的价值。
- 3) 本文通过严格的理论分析和大量的仿真实验，证明了该算法并不会在出现单个或多个投机交易者时导致分配规则和支付规则偏离制度设计的初衷，并会给予投机者惩罚。
- 4) 本文通过鼓励数据产生者即用户主动与数据采购商交易数据、产品市场强制公开数据来源(数据采购商处)等方式[10]，使大众对于自身数据去向更加了解、透明化公众数据流向并采取大众监督的方式使数据来源公开透明，防止数据流入非法渠道。交易流程如图 1 所示。

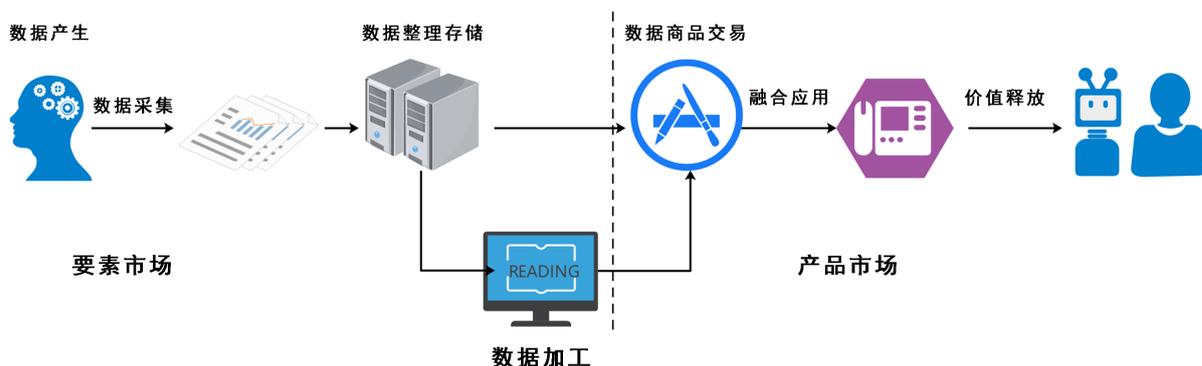


Figure 1. Data transaction flowchart

图 1. 数据交易流程图

本文以数据元组作为数据度量的基本单位，用它来评估个人数据的价值，以校准其在数据市场上的价格。结合前人的研究工作和我们的调查研究，我们可以确定影响一个数据元组值的参数。包括数据成本、价值权重、信息熵、信用评级和数据参考指数。

- 1) 数据成本：交易平台收集、整理和分析数据，以供客户交易。数据产品的成本包括固定成本和可

变成本。由于数据产品的固定成本较低，可以在数据扩展时忽略。因此，数据产品的成本主要通过产生、存储和共享数据的可变成本在形成数据商品时确定，并在交易平台上通过反馈机制得到数据的价值。

2) 价值权重：对于一个数据元组，其价值权重与其价值量和价格呈正相关。为了准确地反映每个数据元组的值，重要的是设置一个称为权重值的属性。权重越大，其值越高。

3) 信息熵：信息熵是信息理论中用来量化信息源中的不确定性的指标。它描述了一组数据中包含的平均信息量。当数据的分布更加随机或不均匀时，信息熵较高，表示其中的信息具有更大的不确定性和复杂性。相反，当数据的分布更加集中或均匀时，信息熵较低，表示其中的信息具有较少的不确定性和复杂性。信息熵的计算基于数据的概率分布，它用于数据压缩、通信和密码学等领域，有助于理解和分析信息的特性和结构。

4) 信用评级：个人信用等级越高，数据提供的可信度越高。

5) 数据参考索引：个人提供的数据元组越多，引用(出售)的数据元组越多，则数据参考索引，其数据值越大。

2. 研究设计

2.1. 信息熵

个人数据在本文中是指个人的基本数据，即个人未经处理的原始数据。为了便于描述，数据销售者(提供者)被称为用户。因为本文讨论的是个人数据定价，所以“用户”指的是个人用户。假设基本的销售单位是一个数据包，它由 n 个数据元组组成。

根据 Shannon 的理论，信息是能够降低不确定性的内容。信息量是一个相对的量，它与事件发生的可能性相关。换句话说，信息量等于选择概率的对数的概率。

设 $X \in \mathbb{R}$ 为离散型变量，概率分布为 $p(x) = P(X = x), x \in \mathbb{R}$ 。则 X 的熵 $H(X)$ 可定义为：

$$H(X) = -\sum_x p(x) \log_2 p(x) \quad (1)$$

熵也被称为自信息，用来描述随机变量的不确定度量。它表示信息源 X 提供的平均信息量；无论发送什么符号。随机变量的熵越大，其不确定度越大，正确估计其值的可能性就越小。随机变量的不确定度越大，用于确定其值的信息内容就越大。假设一个数据包中有 n 个元组，每个元组共有 k 个数据项，则 x_{ij} 表示第 i 个元组的第 j 个数据项。

则一个数据包中所有元组的总熵为：

$$H(X) = \sum_{i=1}^n H(x_i) = -\sum_{i=1}^n \sum_{j=1}^k p(x_{ij}) \log_2 p(x_{ij}) \quad (2)$$

根据等式(2)计算整个数据包 $H(X)$ 的熵。设 q 表示数据包的信息内容的大小，并将 $H(X)$ 分配给 q 。第 i 个数据元组的信息内容权重由 q_i/q 确定，并满足以下约束：

$$\sum_{i=1}^n \frac{q_i}{q} = 1 \quad (3)$$

2.2. 数据参考指标

h 指数(h-index)是一种用于评估学术研究人员产出和影响力的指标。它由物理学家 Jorge E. Hirsch 于 2005 年提出，旨在综合考虑学者的论文数量和被引用次数[4]。**h** 指数的计算方法如下：假设一个学者的论文按被引次数从高到低排列，如果他的第 n 篇论文被引用了至少 n 次，而其后的 n 篇论文被引用次数

均不超过 n 次，那么他的 h 指数就是 n 。 h 指数的作用体现在以下几个方面。首先，它可以客观评估学者的学术产出和影响力。相比于仅仅关注论文数量或总被引次数的指标， h 指数综合考虑了学者的论文质量和影响力。较高的 h 指数意味着学者在学术界的影响力更大，其研究成果得到更广泛的引用和认可。

基于 h -index 的定义，我们将用户数据包的产出等价于论文发布，则用户的数据包等价于一篇发布的论文。在本文中创建 B -index 用以衡量数据提供者即用户的权威性，用户成交的数据元组数量越多，其 D -index 越大，用户的权威性越高，用户 y 的 D -index 为 D_y 。

2.3. 数据价值

个人数据分类主要以数据表为单位。首先，根据经验设置权重值，权重值可分为 n 个级别，从 1 到 n 。权重值越大，数据的重要性就越高。假设第 i 个数据元组的权重值的 w_i ，则该数据包的权重值为 W 。

3. 竞价模型

集合 $M = Y \cup K$ ，为交易主体，其中 Y 为数据包提供者集合， K 为采购厂商集合，其中数据包提供者即用户 y 的竞标信息为 $s_y(T_y, H(x), D_y, W, O_y, n_y)$ ，其中 T_y 表示用户 y 所售卖的该份数据的类型； $H(X)$ 为用户 y 提供的数据包中所有元组的总熵； D_y 为用户 y 在该平台的权威度； W 为该数据包通过经验设置的权重值以表示该份数据的价值； O_y 表示用户 y 对该数据包的报价。则数据出售主体的集合为：

$S = (s_1, s_2, s_3, \dots, s_y)$ 。采购厂商的竞价信息为： $b_k(T_k, H(X)_k, D_k, W, E_k, r_k, n_k)$ ，其中 T_k 表示采购厂商 k 所需购买的数据类型； $H(x)_k$ 为采购厂商 k 对所购买数据包中熵的要求； D_k 为采购厂商 k 对所购买数据包的用户的权威度的要求； W 为采购厂商 k 对所采购数据包的权重值的要求； E_k 为采购厂商 k 自身的企业信息； r_k 为采购厂商 k 对所采购数据包的单位报价； n_k 为采购厂商 k 所要购买的数据包的数量。则所有采购厂商的集合为： $B = (b_1, b_2, b_3, \dots, b_k)$ 。则所有采购厂商的竞标信息 l_B 展开式为：

$$l_B = \{b_1(T_1, H(x)_1, D_1, W, E_1, r_1, n_1), b_2(T_2, H(x)_2, D_2, W, E_2, r_2, n_2), \dots, b_k(T_k, H(x)_k, D_k, W, E_k, r_k, n_k)\} \quad (4)$$

3.1. 分配规则

针对此问题，建立以社会效益最大化为目标的数学规划模型如下：

$$\max z = \sum_{y \in Y} \sum_{k \in K} r_k n_{ky}^c - O_y n_{ky}^c$$

s.t.

$$\sum_{y \in Y} O_y n_c \leq \sum_{k \in K} r_k n_c \quad (5)$$

$$\sum_{y \in Y} T_y \subseteq T_k \quad (6)$$

$$\sum_{y \in Y} n_y \geq n_k \geq n_{ky}^c \quad (7)$$

$$\sum_{y \in Y} D_y \geq D_k \quad (8)$$

$$\forall y \in Y, \forall k \in K \quad (9)$$

3.2. 支付规则

基准支付采用经典 VCG 支付规则，且结合 3.1 的分配规则，在双方竞价结构下，克服了传统单边

VCG 机制只反映了一方的竞价行为及其信息而存在的合谋可能，使得采购方的报价限定了数据包的单位价格上限，用户方报价限定了数据包单价的下限，通过各方的竞价博弈，这确保最终交易价格成为平衡用户和采购方竞争的均衡结果。

1) 用户方获得的支付：

规划目标 z 表示社会最大效益值， $z_c^{y/Y} - z$ 表示在第 c 次交易中除去用户方 y 的报价后系统重新分配的社会福利减小量，则在这个系统中某用户 y 与数据采购商 k 在第 c 次成交量为 n_{ky}^c 的交易中获得的 VCG 支付为：

$$P_{cn_{ky}^c}^{yk} = \sum_{k \in K} \sum_{y \in Y} O_y n_{ky}^c + (z_c^{y/Y} - z) \tag{10}$$

2) 数据采购商的支付：

下式中 $z - z_c^{k/K}$ 表示若数据采购商 k 不购买用户 y 的数据包，系统重新进行交易匹配后的社会福利减小量；则在这个系统中数据采购商 k 与用户 y 第 c 次成交量为 n_c 的交易中获得的 VCG 支付为：

$$P_{cn_{ky}^c}^{yk} = \sum_{k \in K} \sum_{y \in Y} r_k n_{ky}^c - (z - z_c^{k/K}) \tag{11}$$

3) 交易成功的判决条件：

平台进行供需方撮合与前置条件判定完成后，对任意一个采购厂商来说需要最后进行检测交易是否成功，交易成功的条件为：

$$\begin{aligned} & \sum_{k \in K} \sum_{y \in Y} r_k n_{ky}^c - (z - z_c^{k/K}) - P_{cn_{ky}^c}^{yk} \\ &= \sum_{k \in K} \sum_{y \in Y} r_k n_{ky}^c - (z - z_c^{k/K}) - \left(\sum_{k \in K} \sum_{y \in Y} O_y n_{ky}^c + (z_c^{y/Y} - z) \right) \geq 0 \end{aligned} \tag{12}$$

即，数据采购商给出的支付价格需不小于数据商品的最终价格，若：

$$\sum_{k \in K} \sum_{y \in Y} r_k n_{ky}^c - (z - z_c^{k/K}) - P_{cn_{ky}^c}^{yk} < 0 \tag{13}$$

则数据采购商 k 无法获得计划的采购。

3.3. 交割支付

在经典 VCG 支付规则下，数据采购厂商和用户的实际支出与收入如上文所示，各交易人的支出与收入是相互独立的，由于在支付规则中加入了 VCG 机制，目标是使得说真话是交易人的最优策略。但在双边竞价结构下，该支付方式是否存在预算不平衡问题无法保证，本节对经典 VCG 支付规则进行改进以最终形成的交割支付，能确保支付规则生成的交易价格使得交易平台满足预算平衡约束，即所有数据采购厂商支付的采购价格之和大于或等于所有用户得到的价格之和，以此保证平台无需进行额外补贴。由于考虑到交易者在社会福利上的贡献，通过给予“奖励”使得数据采购厂商降低支付总额，用户方增加收入总额，但是不能保证数据采购厂商的支付总额不小于用户方的收入总额，使得交易成功率大幅降低。

为解决以上问题，在此提出按数据采购厂商和用户的 VCG 支付收益比例来分配创造的社会福利，从而修正支付规则。即按照每位交易人对社会福利增量的贡献所占总福利的比重进行社会福利的再分配。按 3.2 的报价原则，记市场在每次交易中实现的总福利为 A ，且 A 等于：

$$A = \sum_{y \in Y} \sum_{k \in K} r_k n_{ky}^c - (z - z_c^{k/K}) \tag{14}$$

由所有 K 个采购厂商分别退出交易时计算的目标函数变化的加和，即所有采购厂商对社会福利的贡献值记为：

$$B = \sum_{k=1}^K \sum_{y \in Y} r_k n_{ky}^c - (z - z_c^{k/K}) \quad (15)$$

由 Y 位用户分别退出交易时计算的目标函数变化的加和，即所有用户对社会福利的贡献值记为：

$$\Gamma = \sum_{y=1}^Y \sum_{k \in K} O_y n_{ky}^c + (z_c^{y/Y} - z) \quad (16)$$

取值：

$$\lambda = B / (B + C) \quad (17)$$

$$\eta = C / (B + C) \quad (18)$$

λ 、 η 分别表示采购厂商和用户在基准支付中各自实现的收益比上两者实现收益的总和，并且 $\lambda + \eta = 1$ ，假设基准支付下采购厂商 k 的收益占有所有采购厂商 K 收益的比例为：

$$\alpha_k = (z - z_c^{k/K}) / B \quad (19)$$

基准支付下用户 y 的收益占有所有用户 Y 收益的比例为：

$$\beta_y = (z - z_c^{y/Y}) / C \quad (20)$$

令所有成功交易采购厂商获得的总收益为 λA ，所有成功交易用户获得的总收益为 ηA 。设：

① 采购厂商 k 对采购物资最终的实际支付为

$$\sum_{k \in K} \sum_{y \in Y} r_k n_{ky}^c - \alpha_k \lambda A \quad (21)$$

② 用户 y 对所有出售数据包的最终收入为：

$$\sum_{k \in K} \sum_{y \in Y} O_y n_{ky}^c + \beta_y \eta A \quad (22)$$

这一支付规则的修正解决了上述提到的在两方竞标模式中基准支付的不足并达到了以下目标：① 满足预算平衡约束；② 根据交易人的社会福利贡献率分配市场收益，并准确刻画出每个交易人的市场贡献。

4. 机制性质

引理 1 如果用户低于自己的真实成本进行报价或数据采购商高于自己对数据商品的估值进行报价，报价人面临负效用风险即亏损。

证明： 设数据采购商 k 对产品的估值为 v_k ，数据采购商 k 对产品的报价为 v'_k ，采购数量为 n_k 。其中 $v'_k > v_k$ ，则采用上述高报价策略时，数据采购商 k 的效用为：

$$U_k = \sum_{k \in K} v_k n_k - (v'_k n_k - \alpha_k \lambda A) \quad (23)$$

由于 $v'_k > v_k$ ， $\alpha_k \lambda A > 0$ ，所以式(23)可能小于零。在密封竞标下，数据采购商 k 无法得知其他采购商的报价而因此无法估计自己的收益，这导致数据采购商如果采用高于自己的估价 v_k 进行报价时，无法保证收益非负，则采购厂商面临负效用风险。

同理可证当用户方的实际报价低于自己真实成本时，也将会面临负效用风险。

定理 1 该机制满足参与理性约束。

证明： 交割支付规则部分已经分析，由于对于数据采购商厂商，其成交价格是将数据采购商厂商的报价

减去该数据采购厂商对社会福利增量乘以他对社会福利的贡献所占总福利的比重。假设数据采购商 k 的竞标单价为 r_k ，在用户 y 与数据采购商 k 在第 c 次成交量为 n_{ky}^c ，对货物的真实估值为 v_k ， λ 表示所有采购厂商在基准支付中实现的收益比上两方交易者实现收益的总和，数据采购商 k 的收益占有所有采购厂商 K 收益的比例为 α_k ，那么采购厂商 k 的实际支付为 $\sum_{k \in K} \sum_{y \in Y} r_k n_{ky}^c - \alpha_k \lambda A$ 。若 $r_k > v_k$ ，则无法保证

$\sum_{k \in K} \sum_{y \in Y} r_k n_{ky}^c - \alpha_k \lambda A \leq \sum_{k \in K} \sum_{y \in Y} v_k n_{ky}^c$ ，因此数据采购厂商作为理性人，将不会报出比对产品估值 v_k 更高的价格。即必有采购人的报价不大于他对产品的估值： $r_k \leq v_k$ 。

所以数据采购厂商 k 的收益：

$$\sum_{k \in K} \sum_{y \in Y} v_k n_{ky}^c - (r_k n_{ky}^c - \alpha_k \lambda A) \geq 0 \quad (24)$$

同理，对于用户方，其所获得的最终支付为用户方的报价加上该用户对社会福利增量乘以他对社会福利的贡献所占总福利的比重。假设用户 y 的货物单位报价为 O_y ，用户 y 与数据采购商 k 在第 c 次成交量为 n_{ky}^c 。用户 y 所供给货物的单位成本为 C_y ， η 表示所有用户在基准支付中实现的收益比上两方交易者实现收益的总和，用户 y 的收益占有所有用户 Y 收益的比例为 β_y ，那么用户在交易完成后所获得的实际收入为：

$$\sum_{k \in K} \sum_{y \in Y} O_y n_{ky}^c + \beta_y \eta A \quad (25)$$

若 $O_y < C_y$ 则无法保证：

$$\sum_{k \in K} \sum_{y \in Y} O_y n_{ky}^c + \beta_y \eta A \geq \sum_{k \in K} \sum_{y \in Y} C_y n_{ky}^c \quad (26)$$

因此用户作为理性人，将不会报出比产品估值 C_y 低的报价，即必有用户的实际报价不小于成本： $O_y \geq C_y$ 。

所以用户 y 的实际收入为：

$$R_y = \sum_{k \in K} \sum_{y \in Y} O_y n_{ky}^c + \beta_y \eta A - \sum_{k \in K} \sum_{y \in Y} C_y n_{ky}^c \geq 0 \quad (27)$$

综上，我们所设计的支付规则是符合参与理性(IR)约束的，即交易双方作为理性的竞标者选择诚实报价其收益一定非负。

定理 2 该机制满足激励相容约束

证明： 设 v_k 为数据采购厂商 k 对产品的每单位价格的真实估值， v'_k 为数据采购厂商 k 在本次交易中的报价，用户 y 与数据采购商 k 在第 c 次成交量为 n_{ky}^c ， λ 表示所有数据采购厂商在基准支付中实现的收益比上两方交易者实现收益的总和，数据采购厂商 k 的收益占有所有数据采购厂商 K 收益的比例为 α_k 。当数据采购厂商 k 的报价等于其真实估值时，即 $v'_k = v_k$ ，称他为诚实交易人。下面将证明当数据采购厂商 k 采取高报策略时，将面临负效用风险；数据采购厂商 k 采取低报策略时，将面临失去真实报价时可赢得交易的风险。对于用户方而言情况类似。

接下来将证明数据采购厂商 k 采取诚实报价是占优策略，具体内容如下：

① 当数据采购厂商 k 采取高报策略时，即 $v'_k > v_k$ 。当采取高报策略可以赢得交易时，由于交易者的目标均是在保证收益非负的情况下尽可能赢标，由引理 1 可知，此时面临负效用风险；当采取高报策略仍然不能赢标时，其收益为零，与采取说真话策略时的收益相同，采用高报策略未能改善其所获得的收益。

② 如果数据采购厂商 k 采取低报策略，即 $v'_k < v_k$ ，如果数据采购厂商 k 采用诚实报价的策略失标时，

他采用低报策略仍然会失标。若采取低报策略并赢标，则此时数据采购厂商 k 的效用为：

$$U'_k = \sum_{y \in Y} \sum_{k \in K} v_k n_{ky}^c - (v'_k n_{ky}^c - \alpha'_k \lambda A') \quad (28)$$

但同时也面临失标的风险。由于数据采购厂商 k 采取低报的策略，他即使赢标，对市场社会福利的贡献必定下降， $\alpha_k \lambda A > \alpha'_k \lambda A'$ 。因采购商采取低报策略并赢标而导致的效益变化值为：

$$U'_k - U_k = \sum_{y \in Y} \sum_{k \in K} v_k n_{ky}^c - (v'_k n_{ky}^c - \alpha'_k \lambda A') + \alpha_k \lambda A \quad (29)$$

其中 $\sum_{k \in K} \sum_{y \in Y} v_k n_{ky}^c - v'_k n_{ky}^c$ 为低报策略所带来的收益增加值， $\sum_{k \in K} \sum_{y \in Y} \alpha_k \lambda A - \alpha'_k \lambda A'$ 为低报策略所带来的收益减小值。总之，低报策略虽有增加采购厂商直接收益的效应，但是由于数据采购厂商 k 的低报导致社会福利降低而导致数据采购厂商 k 所获得的“奖励减少”，在密封拍卖竞标环境中采购厂商无法保证低报带来的收益一定大于低报带来的损失。并且低报策略相较于诚实报价而言降低了赢标概率，会增加失标风险。

由于在此系统中，采购厂商的交易前提为实际支付价格 \leq 支付意愿。当取等号时代表采购厂商收益为零，但仍然会选择交易。因为成功交易会使得采购厂商和用户方投入资源获得了不低于社会平均回报的收益补偿，而失去交易使得资源回报为零。所以成功交易对采购厂商和用户方而言总是有意义的。当处于密封竞标的博弈结构下，数据采购厂商 k 因没有其他采购厂商的报价而无法确定其绝对占优策略，只有采购厂商的报价不低于用户方的报价时，才能保证中标，低报并不能明确提升其收益，同时采用低报策略还将降低赢标概率，而失去交易会使得数据资源失去本应获得的社会平均收益回报，所以作为理性竞标者，此时低报并不是占优策略。

所以，因为采购厂商处于密封竞标的博弈结构中，在没有其他采购厂商的报价信息而无法计算自身收益，理性的交易人会在收益增加和尽可能赢标之间进行权衡，采取高报策略和低报策略均不是占优策略，因此诚实报价是竞标人的占优策略。对于用户方而言同样如此。

定理 3 该机制满足预算平衡约束

证明：假设一批次交易者中，数据采购厂商数量为 K 个，总支出为 Ψ ；用户数量为 Y 个，总收入为 θ ； A 表示市场在每次交易中实现的总福利 λ 、 η 分别表示采购厂商和用户方在基准支付中各自实现的收益比上两者实现收益的总和，并且 $\lambda + \eta = 1$ 。则：

$$\Psi = \sum_{k=1}^K \sum_{y \in Y} r_k n_{ky}^c - \alpha_k \lambda A \quad (30)$$

其中， r_k 表示数据采购厂 k 对数据包的单位报价，用户 y 与数据采购商 k 在第 c 次成交量为 n_{ky}^c ，数据采购厂 k 的收益占有所有数据采购厂 K 收益的比例为 α_k 。则有 $\sum_{k=1}^K \alpha_k = 1$ ，由此可知式(30)可变形为：

$$\Psi = \sum_{k=1}^K \sum_{y \in Y} r_k n_{ky}^c - \lambda A。$$

同理可知 Y 个用户的总收入为：

$$\theta = \sum_{y=1}^Y \sum_{k \in K} O_y n_{ky}^c + \beta_y \eta A \quad (31)$$

其中用户 y 单位报价为 O_y ，用户 y 与数据采购商 k 在第 c 次成交量为 n_{ky}^c 。用户 y 的收益占有所有用户 Y 收益的比例为 β_y 。则有 $\sum_{k=1}^K \beta_y = 1$ ，由此可知式(31)可变形为： $\theta = \sum_{y=1}^Y \sum_{k \in K} O_y n_{ky}^c + \eta A$ 。则：

$$\begin{aligned}
\Psi - \theta &= \sum_{k=1}^K \sum_{y \in Y} r_k n_{ky}^c - \alpha_k \lambda A - \sum_y \sum_{k \in K} O_y n_{ky}^c + \beta_y \eta A \\
&= \sum_{k=1}^K \sum_{y \in Y} r_k n_{ky}^c - \lambda A - \sum_y \sum_{k \in K} O_y n_{ky}^c + \eta A \\
&= \sum_{k=1}^K \sum_{y=1}^Y r_k n_{ky}^c - O_y n_{ky}^c - A(\lambda + \eta) = A - A(\lambda + \eta) \\
&= 0
\end{aligned} \tag{32}$$

由上式可知该机制满足预算平衡约束，无需第三方进行补贴。

5. 投机报价策略交易仿真

根据第 3 节将用户和数据采购商的报价相互比较，以确定成功交易者及最终交易价格。现分析不同情形下当用户和数据采购商采取投机报价策略时的结果。假设数据采购商真实支付意愿和用户真实成本服从正态分布。不失一般性，现以图 2 中的情况为例，分析用户和数据采购商的投机报价会导致的结果，并可类推到多个投机交易者报价情况。假设数据采购商和用户对某一数据商品的估值和成本分别服从正态分布 $N(\mu_1, \sigma_1^2)$ 和 $N(\mu_2, \sigma_2^2)$ 。通过随机生成服从正态分布的数据采购商估值和用户成本，通过开发一个 Python 软件工具，根据第 3 节的交易规则计算每次的交易价格及匹配成功对数。在接下来的仿真中，设定数据采购商报价的期望均值为 $\mu_1 = 9.0$ ，用户报价的期望均值为 $\mu_2 = 8.6$ 。在一个有活力的市场中数据采购商的期望报价应高于用户的期望成本，以确保该市场能形成有效的供需关系，再进一步在每次的仿真中随机产生参与报价的数据采购厂商和用户的数量每次单边最多为 200 人，由于交易重复不断进行，每次交易人数和该市场总交易人数有区别，如此每次交易的人数有限，参与交易者不需花太多时间去等待交易的完成，且该人数限制可在实践中调整。基于多次仿真来分析不同投机策略下该市场的交易价格 (Trading Price) 和成功率 (Successful Rate) 变化趋势。

投机交易者仿真

部分数据采购商可能通过采取投机报价，报出比其对该份数据商品真实估值低的价格来尽可能提高其收益，即使低报价会导致其面临交易难以达成的风险。现针对数据商品竞拍的仿真中，研究数据采购商的投机报价对其成功率和最终交易价格的影响。假设数据采购商估值期望中值为 $\mu_1 = 9.0$ ，标准差为 $\sigma_1 = 0.5$ ，假设用户成本期望中值为 $\mu_2 = 8.6$ ，标准差为 $\sigma_2 = 0.5$ 。在每次交易中，随机选择一名数据采购商低报，低报的程度系数 (Underbidding Rate) 为 β ，比如，当 $\beta = 0.9$ ，表示该数据采购商的报价在其真实估值 (诚实报价策略) 的基础上上浮 10% 作为最终报价。图 2 显示了数据采购商的低价策略在两百次交易中对平均成功率，图 3 则显示了单个采购商的低报策略在不同交易人数中对平均交易价格的影响。从图 2 中可以看出，当投机报价数据采购商使用低报策略时，该数据采购商的成功率 (即赢得交易概率) 显著下降。根据仿真分析，当低报系数等于 0.55 时，该数据采购商赢得交易的概率接近于 1。根据图 3，投机报价数据采购商的低报策略对平均交易价格的影响非常有限，平均而言使交易价格平均下降 1% 或 2% 这表明尽管低报策略可能会略微增加数据采购商的效用但增加有限，并且当低报程度过高时，不会对平均交易价格产生影响。考虑到失去交易机会的风险显著增加，使得低报策略不值得选择。上述分析也适用于试图使用提高报价策略来提高其效用的用户，类似的用户的低报策略对其效用的增加作用非常有限且失去交易机会的风险显著增加。

6. 结论

在现有数据交易模式和交易流程下，鉴于大多数用户对个人信息价值的认知存在欠缺，针对此问题，

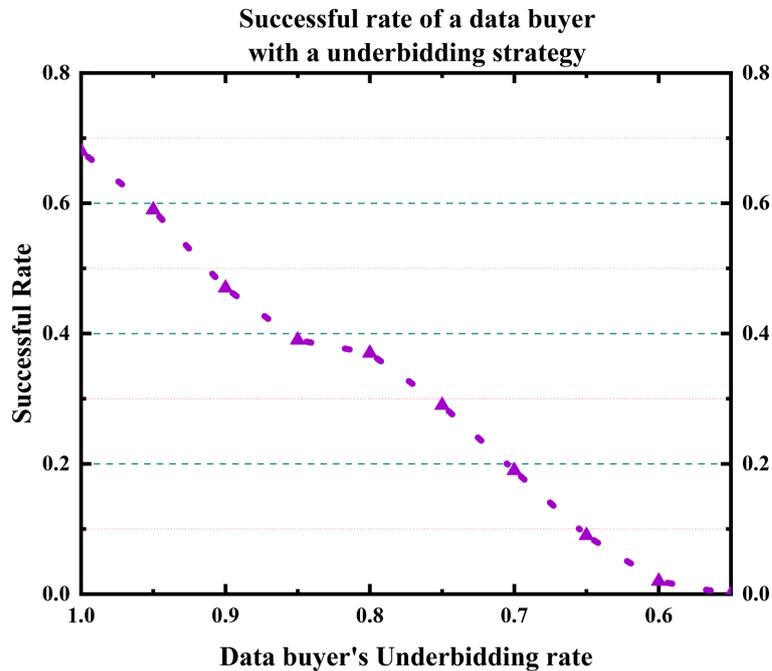


Figure 2. Trend chart of trade success rates at different levels of speculation
图 2. 不同投机程度下交易成功率趋势图

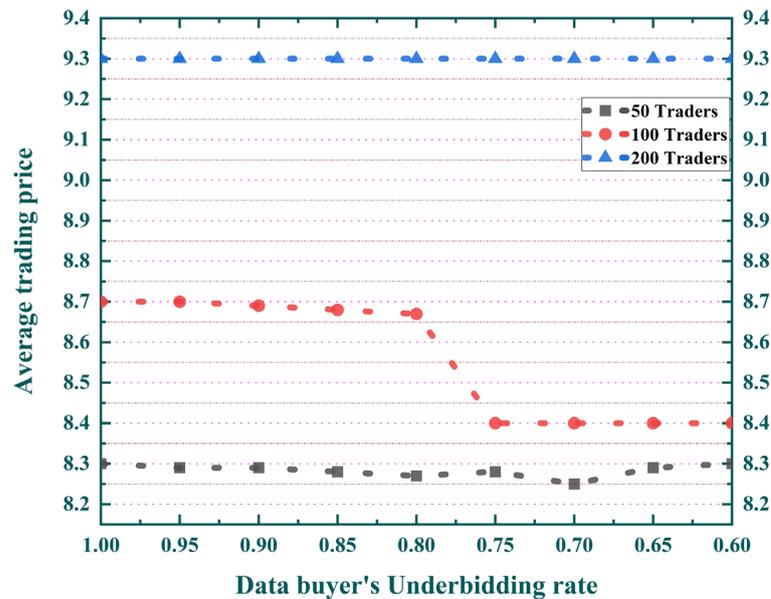


Figure 3. Trend charts of transaction prices at different levels of speculation for individual data procurement vendors
图 3. 单个数据采购厂商不同投机程度下的交易价格趋势图

本文提出了一套激励用户进行数据交易的机制，一方面可以使得用户更积极地参与到数据交易的过程中以加快数据要素市场的建设，另一方面可以使民众加强对于个人隐私数据的重视和保护。本文以大数据定价规则不确定、市场信息不对称、用户对于数据交易意愿过低、数据商品特殊性导致数据定价困难的问题为背景，结合拍卖理论和机制设计理论中的经典模型设计一个基于组合拍卖的定价模型，适用于不同应用场景和买卖双方的不同要求。引入了改良后的 VCG 机制所包含的分配规则和支付规则，通过“惩

罚”投机报价交易者的形式用买卖双方报价所带来的社会福利值对其进行奖励，能较好地揭示多方交易者对数据商品的真实估值，从而达成各交易方真实信息的暴露，通过构建交易模型，基于社会福利最大化原则实现市场资源的最优配置。本文提出的交易机制能够及时反映市场各类交易信息，快速达成交易匹配，形成以社会效益最大化为目标的数据交易系统，从而实现市场资源的合理配置，通过仿真分析新交易机制显著提升了市场交易效率。

参考文献

- [1] Hu, D., Li, Y., Pan, L., *et al.* (2021) A Blockchain-Based Trading System for Big Data. *Computer Networks*, **191**, Article ID: 107994. <https://doi.org/10.1016/j.comnet.2021.107994>
- [2] Samani, E., Kohansal, M., Mohsenian-Rad, H. (2021) A Data-Driven Convergence Bidding Strategy Based on Reverse Engineering of Market Participants' Performance: A Case of California ISO. *IEEE Transactions on Power Systems*, **37**, 2122-2136. <https://doi.org/10.1109/TPWRS.2021.3114362>
- [3] Heckman, J.R., Boehmer, E.L., Peters, E.H., *et al.* (2015) A Pricing Model for Data Markets. <http://hdl.handle.net/2142/73449>
- [4] 周乐欣, 滕可, 吕凡. 平台型物流采购双向竞标交易模式创新研究[J]. 管理学报, 2019, 16(4): 624-632.
- [5] 郭鑫鑫, 王海燕, 孔楠. 信息不对称下个人健康数据交易双边定价策略研究[J]. 管理工程学报, 2022, 36(4): 129-139.
- [6] 赖明辉, 薛巍立, 田歆, 李丽丽. 整车运输协作问题迭代拍卖机制设计[J]. 系统工程理论与实践, 2018, 38(12): 3174-3186.
- [7] 周乐欣, 宋山梅, 李露. 大数据条件下物流采购竞价交易模式创新研究[J]. 贵州大学学报(社会科学版), 2018, 36(2): 63-68.
- [8] 彭安华, 肖兴明, 岳睿. 基于交互双层模糊规划的敏捷供应链构建与优化[J]. 中国机械工程, 2014, 25(17): 2314-2319+2350.
- [9] Cong, R.G. and Wei, Y.M. (2012) Experimental Comparison of Impact of Auction Format on Carbon Allowance Market. *Renewable & Sustainable Energy Reviews*, **16**, 4148-4156. <https://doi.org/10.1016/j.rser.2012.03.049>
- [10] Lai, M.H., Cai, X.Q. and Hu, Q. (2017) An Iterative Auction for Carrier Collaboration in Truckload Pickup and Delivery. *Transportation Research Part E: Logistics and Transportation Review*, **107**, 60-80. <https://doi.org/10.1016/j.tre.2017.09.006>