SWDPM：社会福利优化数据定价机制

Yi Yu城市计算实验室 上海人工智能实验室上海，中国
yuyi@pjlab.org.cn

2nd Shengyue Yao城市计算实验室 上海人工智能实验室上海，中国
yaoshengyue@pjlab.org.cn

3rd JuanjuanLi复杂系统管理与控制国家重点实验室，中国科学院自动化研究所北京，中国
juanjuan.li@ia.ac.cn

4th Fei-YueWang复杂系统管理与控制国家重点实验室，中国科学院自动化研究所，北京，中国 澳门系统工程研究院，澳门科技大学，澳门，中国
feiyue.wang@ia.ac.cn

5th Yilun Lin城市计算实验室 上海人工智能实验室上海，中国
linyilun@pjlab.org.cn

arXiv:2305.06357v1 [cs.GT] 8 May 2023

摘要—**用户拥有的数据涉及隐私顾虑，加之数据可以无限复制，阻碍了数据交易的发展，使数据所有者难以在数据披露后继续保有其专有权。传统数据定价模型依赖统一定价或订阅制模型。然而，随着隐私保护计算技术的发展，市场如今能够保护隐私，并通过逐步披露信息完成交易，从而为通过数据使用创造更大的社会福利奠定了技术基础。在本研究中，我们提出一种新方法，利用基于匹配的马尔可夫决策过程（MDP）对逐步披露信息的多轮数据交易进行建模，并提出一种社会福利优化数据定价机制（SWDPM），以寻找最优定价策略。据我们所知，这是首个对逐步披露信息的多轮数据交易进行建模的研究。数值实验表明，SWDPM可使社会福利提升3倍，并通过促进交易者之间更好的需求匹配和价格协商，将交易可行性提高至多54%、交易效率提高至多43%、交易公平性提高至多25%。**

索引词—**数据定价、社会福利、强化学习、数据价值、匹配**

I. 引言

在数字化新时代，产生了大量具有可观生产价值的无处不在的数据，这推动了数据协作平台和数据即服务（DaaS） [1]等新兴产业的快速发展。尤其是，大数据分析市场规模在2022年达到2718.3亿美元，预计到2029年将达到6555.3亿美元[2], 。其发展有望通过为数据供应方、买方及整个社会创造价值增量，从而提升社会福利 [3]。

政府、市场和学术界日益认识到数据价值，同时从不同角度开展了大量数据定价实践与研究。在大多数由政府主导的平台上，数据均可免费共享供公众使用 [4]。在数据市场中，数据交易主要采用统一定价或基于订阅的定价，以实现收入最大化 [5]–[7]。在学术领域，研究者从基于价值、基于协商和基于成本等方面研究了各种数据定价模型，并应用了经济学、营销学、机器学习等领域的方法 [8]–[10]。然而，值得注意的是，当前的数据定价模型主要基于这样一种假设：所有信息一经交易就会被披露，因此数据很可能只会被交易一次，或以相同方式被多次交易 [8]–[11]。本文将其称为完全披露假设。

隐私保护技术通过量化信息披露并提高交易过程中的可靠性，支持多轮数据交易 [12]–[16]。由于数据包含大量信息，只要确保每轮交易仅披露部分信息，就可以分多轮交易数据，从而将数据交易从单次、非连续的过程转变为一系列具有时间关联的交易。采用多轮交易方式，交易者可以利用各轮交易之间的相关性，获取目标信息并实现收入目标，从而拓展应用范围，并通过最大化数据效用释放其潜在价值。然而，均匀定价或基于订阅的定价等现有模型很难满足这些需求，因为它们大多因完全披露假设而缺乏对动态变化的把握或对可持续性的考量 [11]。

在多轮交易模式中，数据交易可以

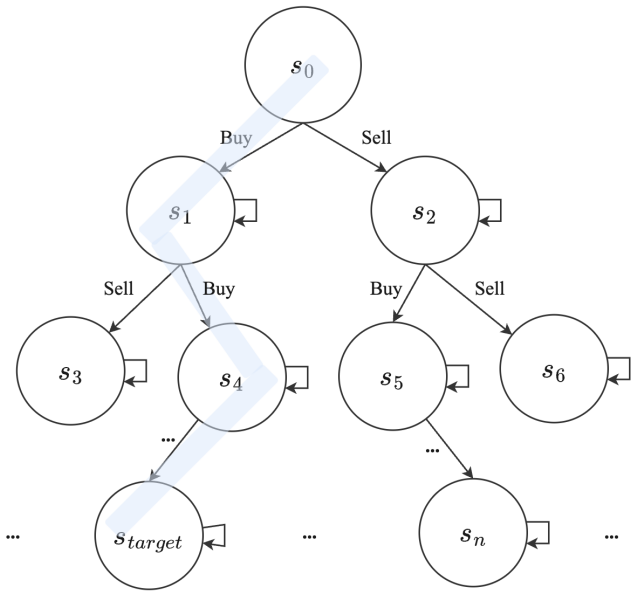

图 1. 将数据交易视为马尔可夫决策过程

建模为多智能体马尔可夫决策过程（MDP），其中每位交易者的状态可由信息量和货币来刻画，动作则可分为购买或出售信息，或维持现状，如图 1 所示。交易者当前的状态仅与前一状态和动作有关，体现了马尔可夫性质 [17],; 状态转移取决于其自身及其他交易者的动作，因此构成一种混合策略博弈。在这一过程中，定价对于确定信息披露与货币之间的交换比率至关重要，而交换比率与价值和波动性相关。合理定价能够激励交易者参与并释放数据价值，为所有买方和卖方带来剩余收益，最终提升社会福利；社会福利等于剩余收益之和 [18]。

将数据交易过程建模为马尔可夫决策过程后，我们提出了一种社会福利优化数据定价机制（SWDPM），该机制由一种匹配方法构成，并采用强化学习（RL）有效应对多轮交易模式。该机制不直接计算数据价值和波动性，而是通过学习交易者之间的互动来优化定价策略，从而激励交易。我们通过数值模拟验证所提机制的可行性。结果表明，与传统定价机制相比，该机制能够促进交易者之间更积极的合作，并在可行性、效率、公平性和社会福利增量方面表现更优。

针对上述研究空白，本文的主要贡献如下：（1）通过基于匹配的马尔可夫决策过程对数据交易过程进行建模，从而能够捕捉多轮交易模式中的定价动态以及交易者特征；（2）提出一种基于强化学习、以社会福利优化为目标的交易机制，以提升数据交易的可持续性，并通过最大化社会福利充分释放数据价值。

II. 相关工作

在数据定价领域，研究人员主要从两个方面开展研究：设计专门且复杂的模型，通过交易者的谈判估算波动性；采用经济定价方法，从基于价值和基于成本的视角估算数据价值。关于基于博弈论的模型，已有均匀、基于订阅和基于谈判的数据定价方法 [5]–[7], [19]–[22]。由于完全披露假设意味着数据可能只交易一次，现有基于谈判的定价模型在捕捉动态特征或将其最优解推广到更多场景方面存在不足 [11]。多数模型在可行性方面也存在缺陷，因为它们过于复杂，难以应用于现实。至于基于经济理论的模型，已有多种基于价值和基于成本的数据定价模型，例如基于数据质量 [8],、基于夏普利值 [9],、基于查询 [10], 以及基于数量的模型 [23]。然而，构建可持续且具有激励作用的定价机制仍是一项关键挑战，而这正是数字化时代所需要的 [24]。

在更广泛的定价研究中，近年来，马尔可夫决策过程和强化学习在机制设计领域受到了广泛关注。能源市场、保险行业和赞助搜索拍卖等多个领域，都将定价问题视为一系列决策问题 [25]–[29]。交易过程可以建模为马尔可夫决策过程或受约束的马尔可夫决策过程 [27], [28],，并引入了 Q-learning [25], 、深度确定性策略梯度（深度确定性策略梯度） [27],和长短期记忆网络（长短期记忆网络）[26]等多种算法，以支持定价机制的构建。大量数值模拟和实验验证表明，基于强化学习的机制在解决复杂定价问题时表现出较高的性能和鲁棒性，并通过设定适当的目标来实现公平性和收益最大化 [25], [28]。

由于完全披露假设的存在，主流数据定价研究缺乏对动态性和可持续性的考量，而在更广泛的定价研究领域，通过应用马尔可夫决策过程（MDP）和强化学习（RL），定价机制的动态性和可持续性得以实现。在隐私保护技术的支持下，可以利用马尔可夫决策过程（MDP）和强化学习（RL）为数据交易场景设计定价机制，从而在不失一般性的前提下，准确描述交易过程并提供激励性定价结果，进而释放数据价值。这些机制可以综合考虑公平性、可持续性和社会福利等多种目标。

III. 问题陈述

本研究聚焦于通过数据交易者之间的多轮交易释放数据价值，从而提升社会福利。为此，我们将数据交易过程建模为基于匹配的马尔可夫决策过程，具体内容将在

本节中详细阐述。提出 SWDPM 所用的变量和参数汇总于表 I。

表 I 变量和参数的符号说明

符号	含义
$\mathcal{T}$	交易者集合
s_t^i	交易者 i 在时间 t的状态
S_t	时间 t的状态集合
vt^i	交易者 i的目标信息量
v_t^i	交易者 i 在时间 t的信息量
c_t^i	交易者 i 在时间 t的货币金额
a_t^i	交易者 i 在时间 t采取的动作
A_t	时间 t的动作集合
dv_t^i	交易者 i 在时间 t披露的信息量
dc_t^i	交易者 i在时间 t 变更货币的金额
I_i	交易者 i的交易接受意向
I	交易接受意向集合
dc_{min}	最低价格
ΔA_t	时间 t的未解决动作集
Δa_t	ΔA_t中的未解决动作
Δdv_t^i	未交易披露信息量
p_a	状态转移概率（取 a 时）
r_t^i	交易者 i 在时间 t 的奖励
R_t	时间 t的系统奖励
$w(a_t^i)$	通过取 a_t^i 获得的交易者剩余
W	社会福利增量
Q	动作价值函数
Q^*	最优动作价值函数
$\tilde{Q}$	Q表
A_t^*	时间 t的最优动作集合
$\mathcal{D}$	历史交易数据
η	单位披露信息的标准价格
δ	折扣率
θ	公平性权重因子
λ	惩罚权重因子
α	学习率
γ	折扣因子
ξ	最大训练回合数
Φ_f	可行性评估指标
Φ_e	有效性评价指标
Φ_r	公平性评价指标

A. 基于匹配机制的数据交易马尔可夫决策过程

马尔可夫决策过程广泛用于表示动态系统中的随机决策过程；具体而言，系统动力学可表示为集合 (s, a, p_a, r_a) 的演化，其中 s 是系统状态， a 是在 s 采取的动作， $p_a = Pr\{s_{t+1} = s' | s_t = s, a_t = a\}$ 是系统在 s 时处于 s' 并于时间 t 采取 a 时，在时间 $t+1$ 转移到 s' 的概率，而 r_a 是在 s 采取 a 的奖励函数。数据交易过程中的集合 (s, a, p_a, r_a) 根据以下假设定义：

(1) 原始数据中披露信息的量可以进行衡量，并划分为多个部分，据此，多轮数据交易在理论上和实践中都是可行的。(2) 每个时间步内，交易者可以与多个交易者进行数据买卖，而一个代理每个时间步只能与另一个代理进行一次数据买卖。(3) 交易者的交易过程被视为单向过程。具体而言，交易者只会考虑购买（出售）数据，而当他们拥有的信息量低于（高于）其目标时，分别会进行相应操作。

(4) 信息量和用于交易数据的货币分别以最小单位 u_v和 u_c离散化。

根据这些假设，对于包含 N 个交易者的交易者群体 $\mathcal{T} = \{1, 2, ..., N\}$，MDP 的定义，包括集合 $(s,\ a,\ p_a,\ r_a)$ 的定义和匹配方法，详述如下：

a) 状态：对于 $\mathcal{T}$中的交易者 i，其在时间步 t、 s_t^i 的状态可以表示为一个由 $[vt^i, v_t^i, c_t^i]$组成的向量；其中，$v_t^i \in \mathbb{R}_{\geq 0}$ 表示交易者 i 在时间步 t所拥有的原始数据中已披露信息的累计信息量，而 $vt^i \in \mathbb{R}_{\geq 0}$ 表示交易者 i 期望拥有的目标累计信息量。此外， $c_t^i \in \mathbb{R}_{\geq 0}$表示交易者$i$ 在时间步 t所拥有的货币数量。请注意， vt^i 是在交易过程开始之前确定的固定值。此外，根据假设（3），当 $v_t^i < vt^i$时，交易者 i 会寻求花费其货币 c_t^i 购买数据，直到 $v_t^i = vt^i$；类似地，当 $v_t^i > vt^i$ 时，交易者 i会持续出售数据，直到其 v_t^i 达到vt^i。

b) 动作：交易者 i 通过采取动作 a_t^i在状态之间转换，该动作可表示为由 $[dv_t^i, dc_t^i]$构成的向量；其中，$dv_t^i \in [-v_t^i, \infty)$ 和 $dc_t^i \in [-c_t^i, \infty)$ 分别表示交易者 i 在时间步 t进行交易期间累计披露的信息量和累计变更的货币。此外，动作 a_t^i 可视为交易者 i出售（购买）一定量披露信息的价格提议，因此，可行的提议应符合交易者 i 的交易接受意向。在本研究中，交易接受意向表示为 I_i，其公式见 (1)：

$$I_i(dv_t^i, dc_t^i) = \begin{cases} 1, & dc_t^i(j) \geq dc_{min}(dv_t^i) \\ 0, & otherwise \end{cases} \tag{1a}$$

其中

$$dc_{min}(dv_t^i) \underrightarrow{R} \begin{cases} [-\eta \cdot dv_t^i, -(1-\delta) \cdot \eta \cdot dv_t^i], & dv_t^i(j) \geq 0 \\ [-\eta \cdot dv_t^i, -(1+\delta) \cdot \eta \cdot dv_t^i], & dv_t^i(j) < 0 \end{cases} \tag{1b}$$

其中， η 是一个常数，表示单位披露信息量的标准价格； u_v 和 $\delta \in (0,\ 1)$为折扣率。一般而言，如果针对披露信息dv_t^i提出的价格（dc_t^i）高于最低价格 dc_{min}，交易者 i 便可以接受该提议，即$I(i) = 1$。具体来说，当交易者 i 购买数据时，披露信息量越大，预期价格越低，因此从集合$[-\eta \cdot dv_t^i, -(1-\delta) \cdot \eta \cdot dv_t^i]$中随机选取 dc_{min} ；反之，当交易者 i 出售更多披露信息量时，预期价格越高，因此从集合$[-\eta \cdot dv_t^i, -(1+\delta) \cdot \eta \cdot dv_t^i]$中随机选取 dc_{min} 。

c) 匹配方法：给定 $\mathcal{T}$中的交易者动作集合 $A_t = \{a_t^1, ...a_t^N\}$及其交易接受意愿 $I = \{I_1, \ldots, I_N\}$，交易者之间将根据匹配方法自动执行交易。该方法是股票市场中广泛使用的做市商/接受者模型[30]。匹配方法由算法1描述。

算法1基于做市商/接受者模型的匹配方法

要求： 行动集合 A_t，交易接受意向集合 I
确保： 未解决动作集 $\Delta A_t = \{\Delta a_t^1, \Delta a_t^2, ..., \Delta a_t^N\}$

```
初始化 ΔA_t = A_t, m = ∞
当 ΔA ≠ ∅ 且 m ≠ 0 时
    对 ΔA 排序_t 采用随机顺序
    m=0
    对于所有 Δa_t 在 ΔA_t 中做
        获取 Δa_t 的交易者索引作为 i
        如果 dv_t^i = 0 和 dc_t^i = 0 然后
            ΔA = ΔA/{Δa_t}
            继续
        结束 if
        对于所有 Δa'_t 在 ΔA_t/{Δa_t} 中做
            获取 Δa'_t 的交易者索引，记为 j
            如果 dv_t^j · dv_t^i < 0 并且 |dv_t^j| ≤ |dv_t^i| 并且 |dc_t^j| ≤ |dc_t^i|
            以及 I_i(−dv_t^j, −dc_t^j) = 1 那么
                Δa_t = (dv_t^i + dv_t^j, dc_t^i + dc_t^j)
                m=m+1
                中断
            结束 if
        结束循环
    结束循环
结束 while
```

d) 状态转移概率：根据算法1中的匹配方法，给定 $\mathcal{T}$ 中交易者的状态集合、 A_t 中的动作集合，对于交易者 i，只有在所有交易均成功执行时，才能从 s_t 转移到 s_{t+1}，即 $\Delta A_t = \varnothing$，因此 $p_a = \prod_{i \in N} Pr\{s_{t+1}^i | s_t^i, a_t^i\}$；否则，交易提案将失败，所有交易者都停留在 S_t。尽管交易过程的建模十分复杂，且难以准确掌握 p_a，但可以通过蒙特卡洛方法优化动作。

e) 奖励： SWDPM 的目标是优化社会福利，因此奖励函数根据交易者剩余 [18]，、交易公平性 [5]，和成功率进行相应设计：

$$r_t^i = w(a_t^i) + \theta f(a_t^i) + \lambda |\Delta dv_t^i| \tag{2a}$$

其中

$$w(a_t^i) = |dc_t^i| - |dc_{min}(dv_t^i)| \tag{2b}$$

通过取 a_t^i 来表示交易者剩余，并且

$$f(a_t^i) = \frac{dc_t^i}{dv_t^i} - \frac{\sum_{i \in N} dc_t^i}{\sum_{i \in N} dv_t^i} \tag{2c}$$

通过采用 a_t^i体现公平性； θ 是公平性权重因子；而 $\lambda \Delta dv_t^i$ 表示未成功交易的惩罚，其中 Δdv_t^i 表示未通过匹配方法交易的信息量，可在 ΔA_t中找到， λ 是惩罚权重因子。

B. 使用 *Q-learning* 进行数据定价

给定数据交易过程的基于匹配的马尔可夫决策过程，可通过推导最优动作价值函数 Q^* 来求解最优定价。本研究采用集中式学习和集中式执行方案，因此通过经典的 Q-learning 方法推导最优系统动作价值函数 $Q^*(S_t, A_t)$。

通常，最优动作集合可通过式 (3) 得到：

$$A_t^* = \underset{A}{\operatorname{argmax}}\, Q^*(S_t, A) \tag{3}$$

其中， $Q*$ 可通过最优贝尔曼函数求得：

$$\begin{aligned} Q^*(S_t, A_t) =& \mathbb{E}_{S_{t+1} \sim p_a(\cdot|S_t, A_t)}[R_t + \\ & \gamma \max_A Q^*(S_{t+1}, A) | S_t, A_t] \end{aligned} \tag{4}$$

如在介绍状态转移概率时所述， $Q^*(S_t, A_t)$ 可通过蒙特卡洛 ˜ 方法近似，并通过更新表格 Q 中的值实现：

$$\begin{aligned} \tilde{Q}(S_t, A_t) =& \tilde{Q}(S_t, A_t) \\ & + \alpha \left(R_t + \gamma \max_A \tilde{Q}(S_{t+1}, A) - \tilde{Q}(S_t, A_t) \right) \end{aligned} \tag{5}$$

其中， γ 是折扣因子，表示数据价值随时间推移而损失，这与现实中随时间推移发生的折旧相符； $\alpha \in [0, 1]$ 是学习率，R_t 是系统奖励：

$$R_t = \sum_{i \in N} r_t^i \tag{6}$$

如图 2 所示，定价代理可在训练过程中通过多次迭代更新表格 ˜ Q ，从而近似最优系统动作价值函数。

在实际实现中，由于数据交易是实时进行的，定价算法的效率至关重要。为了优化定价流程，我们可以利用历史数据并采用预训练与微调策略，事实证明，这些策略能够降低强化学习 [31], [32]的复杂度。利用截至时间 $t\}$ 的历史交易数据 $\mathcal{D} = \{(S_0, A_0, S_0')$……，$(S_t, A_t, S_t')$˜ ˜，我们可以获得一个参考最优 Q'。随后将 Q' 传递给后续微调，以获得最优˜Q 和最优动作集合 A^*。预训练与微调策略的框架如图 2 所示，相应的伪代码见算法 2。

采用基于强化学习的定价机制与匹配方法，可帮助交易者通过多轮交易达到目标状态，从而充分释放潜在的数据价值，推动系统实现社会福利最优化。

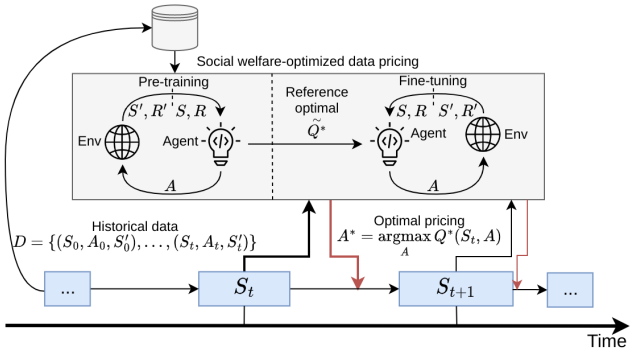

图 2. 基于预训练与微调强化学习的数据定价框架：参考最优 $\tilde{Q}'$ 使用历史数据进行预训练，并传递给微调阶段以获得最优动作集合 A^*

算法 2 预训练与微调方法 要求：历史数据 $\mathcal{D}$，当前状态 S_t，学习率 α，折扣因子 γ，最大训练回合数 ξ确保：最优动作集合 A^*1：预训练初始化 2：对于每个回合 $< \xi$ 执行 3：从中随机选取一个样本 (S_t A_t S_{t+1} $\mathcal{D}$, ,) 更新 $Q'(S_t, A_t)$ $= Q'(S_t, A_t)$ $+ \alpha(R_t +$4: $\gamma \max_A$ $\tilde{Q}'(S_{t+1}, A)$ $- \tilde{Q}'(S_t, A_t))$ 5：结束循环 6：为微调初始化 $S = S_t, Q = Q'$ 7：当 $\forall s \in S, v \neq vt$ 时执行8：采取 $A_t = \text{argmax}$ $_A Q$ S_t A S_{t+1} R_t (,), 观察并更新 $Q(S_t, A_t)$ $= Q(S_t, A_t)$ $+ \alpha(R_t +$9: $\gamma \max_A$ $\tilde{Q}(S_{t+1}, A)$ $- \tilde{Q}(S_t, A_t))$ 10：更新 $S_t = S_{t+1}$11：结束循环12：返回提取 $A^* = \text{argmax}_A Q(S_t, A)$

C. 评估指标

数据定价机制的性能可以从可行性、效率和公平性三个方面进行评估。

a) 可行性：在市场中，理性交易者只会接受收入超过成本的交易，这一点在数据交易场景中同样适用。可行机制可以优化定价，并提高交易者接受提案的概率。可行性指标据此设计：

$$\Phi_f = M_a / M_p \tag{7}$$

其中，M_a 表示一个时间步内测试中完成的动作数量；M 表示测试中的提案数量。计算复杂度也可以作为可行性的另一个方面，可在未来进行讨论。

b) 效率：通常，有效的定价机制能够为交易者和社会带来更高的单位投资收入，其中收入可以用社会福利表示增量，而投资则由披露信息产生。效率指标正是基于此设计的：

$$\Phi_e = W / V \tag{8}$$

其中，W 表示社会福利增量，即$W = \sum_{i \in N} w^i$ [18]；V表示披露信息量之和，即 $V = \sum_{i \in N} dv^i$。

c) 公平性：对相同工作给予同等报酬，可以激励交易者合作，从而为系统带来更高的社会福利增量。公平性通过一种感知视角进行分析 [33]:

$$\Phi_r = 1 - \sqrt{\frac{1}{N} \sum_{i \in N} \left(\phi^i - \bar{\phi}\right)^2} \tag{9}$$

其中，交易者 i在 $\mathcal{T}$中的 $\bar{\phi} = \frac{1}{N} \sum_{i \in N} \phi^i$ 和 $\phi^i = dc^i / d\overline{v^i}$ 。本文中的所有数据均为同质数据，而在未来的研究中，我们可以计算同质数据的公平性，并将所有公平性相加，以进行系统公平性评估。

IV. 数值模拟

本节通过实验验证我们提出的 SWDPM。由于缺少真实的数字资产交易数据，我们采用数值模拟来评估其可行性和性能。

A. 实验设置

该仿真基于典型的双边数据交易场景构建，如图 2所示的环境所示，其中包括四个数据买方和供应方，其偏好设置见表 II。请注意，对于所有交易接受意向，一旦在初始化时生成，在之后的时间或交易过程中均不会改变。为了评估 SWDPM 的性能，我们采用不同的初始状态进行了四轮仿真。交易者的初始状态 $s_0 = [vt_0, v_0, c_0]$ 从集合 $\{[10, 0, 10], [0, 10, 0], [12, 0, 12], [0, 12, 0], [10, 0, 9]\}$中随机选取。为降低计算复杂度，将最小单位 u_v 和u_c 设为 1，从而使状态空间和动作空间离散且有限。强化学习训练的超参数设置如表 II 所示。

表 II 数值模拟中的参数设置

参数	值
η	1
δ	0.2
γ	0.995
α	0.1
θ	-0.5
λ	-100
ξ	10^6

为与现实世界定价策略进行比较，我们采用了统一定价和基于订阅的定价，这两种定价方式常用于现实世界的数据交易市场 [11], [19],[20]。统一定价通常用于单边市场，在

这些市场中，价格由买方或卖方根据当前市场状态决定，而很少考虑谈判或可持续性；基于订阅的定价通常用于数字产品市场，在这些市场中，价格由提供商根据长期收入固定并决定，而很少考虑数据的动态性，例如 Spotify 中的定价。在这些策略中，交易能否成功仅取决于另一位交易者的意愿。

B. 评估结果

整体可行性、效率、公平性和社会福利增量等评估指标根据状态和动作计算得出，如图 3 和表 III 所示。

图 3 中，横坐标表示交易次数，纵坐标表示评估指标。如图 3 (1) 所示，SWDPM 可以显著提高可行性，因为定价代理能够通过训练学习交易者的偏好，并提供更合理的定价。可行性曲线的变化趋势与理论趋势一致：随着状态接近目标状态，可行的动作空间会变小，可行性也会自然下降。图 3 (2) 表明，SWDPM 也提高了效率，这说明更合理的定价能够为所有交易者带来更高的整体收入。图 3 (3) 中，公平性并非在每次交易中都有所提升，这可能是因为考虑到随机性，统一定价能够提供更公平的价格。但图 3 (4) 表明，SWDPM 带来的社会福利增量是统一定价或基于订阅的定价的 3 倍以上，这与我们的理论一致：通过多轮交易，数据的潜在价值能够得到充分挖掘。

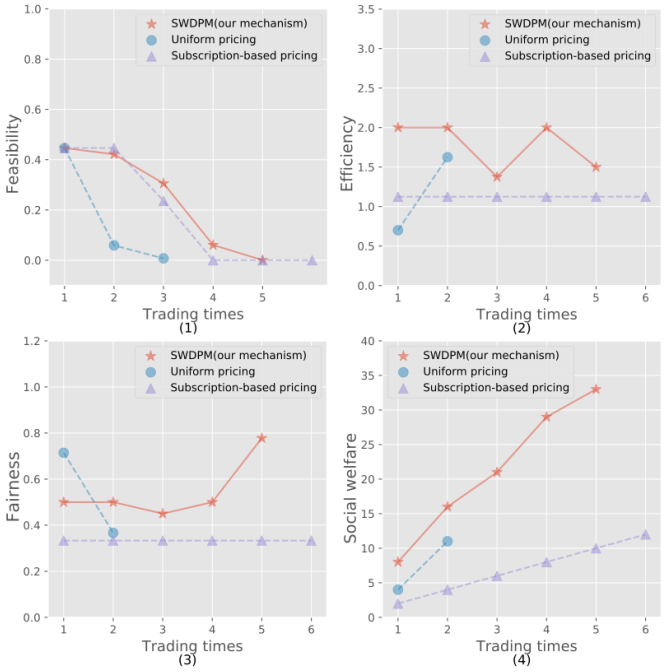

图 3. 初始状态设为 $\{[10,0,9],[0,10,0]\}$ 时，SWDPM 与对比方法的评估指标结果。

表 III 仿真评估指标结果

	方法	次数	Φ_f	Φ_e	Φ_r	$\sum W$
第1轮	**SWDPM**	**5**	0.261	1.8	0.533	**34**
	均匀	2	0.313	1.875	0.625	13
第2轮	**SWDPM**	**5**	0.235	1.875	0.533	**35**
	均匀	3	0.188	1.34	0.687	17
第3轮	**SWDPM**	**6**	0.094	1.896	0.528	**43**
	均匀	2	0.061	1.76	0.706	13
第4轮	**SWDPM**	**6**	0.094	1.917	0.528	**43**
	均匀	2	0.092	1.333	0.539	12

表 III 展示了单轮模拟中评估结果的平均值。与统一定价相比，SWDPM 实现了更多交易次数，这意味着通过训练捕捉到了交易特征和交易者偏好，从而能够持续激励交易。表 III 中的其他评估结果也与图 3 一致，表明 SWDPM 具有稳健性，能够在所有模拟轮次中稳定地促进数据交易并实现社会福利优化。

总体而言，模拟结果表明，在多轮交易条件下，考虑数据的动态特征至关重要。通过采用基于匹配的马尔可夫决策过程，并在强化学习中设定社会福利优化目标，SWDPM 能够捕捉交易过程中的动态变化，并通过持续激励交易过程实现社会福利优化。

V. 结论与未来工作

本文提出了一种支持多轮交易模式的新型机制 SWDPM。SWDPM 将交易建模为基于匹配的马尔可夫决策过程，并利用预训练与微调强化学习优化定价策略。与均匀定价和基于订阅的定价相比，数值模拟表明，SWDPM 能够提高社会福利通过促进需求与价格协商在交易者之间更好地匹配，从而提升可行性、效率和公平性。

我们提出的机制不仅可行，而且针对多轮交易模式实现了社会福利优化，能够通过提供公平合理的定价，为健康、可持续的数据交易市场发展提供理论支持。为进一步改进和拓展该机制，社区仍有一些工作有待未来研究。例如，在理论建模中，多个数据定价动作之间可能存在时间相关性，因此需要使用半马尔可夫决策过程将相关状态整合为一个状态，以提高模型的准确性。此外，SWDPM 的训练过程可以去中心化，以减轻计算负担，并促进其在交易者达到一定规模的更广泛市场中实施 [34]。此外，在实际应用方面，区块链研究中的智能合约和平台服务费为提升 SWDPM [35], [36], 作为撮合机制的实用性和效率提供了启示

在现实中，这可能计算成本高昂，而且难以自动、免费地完成。

致谢

本研究得到上海人工智能实验室、澳门特别行政区科学技术发展基金（0050/2020/A1）以及国家自然科学基金委员会（62103411）的资助。

参考文献

[1] Z. Zheng、J. Zhu 和 M. R. Lyu，“服务生成的大数据与大数据即服务：概述”，载于*2013 IEEE* 大数据国际大会，2013年6月，第403–410页。 [2] F. B. Insights，“大数据分析市场规模 [2022-2029] 预计到2029年将以13.4%的年复合增长率达到6555.3亿美元”，2022年11月。[在线]。可访问：https://www.globenewswire.com/en/news-release/2022/11/14/2554626/0/en/Big-Data-Analytics-Market-Size-2022-2029-Exhibits-13-4-CAGR-to-Reach-USD-655-53-Billion-in-2029.html[3] M. Tripathi and S. K. Inani，“信息与通信技术是否会影响经济增长？来自南亚区域合作联盟国家的实证证据，”《信息技术促进发展》, vol. 26, no. 4, pp. 773–787, Oct. 2020. [4] C. S. Lab，“哥本哈根城市数据交易所，”https://cphsolutionslab.dk/en/projekter/data-platforms/city-data-exchange, 2017. [5] A. Agarwal, M. Dahleh, and T. Sarkar，“数据市场：一种算法解决方案，” May 2019. [6] M. Zhang and F. Beltr´an，“数据定价方法综述，” Rochester, NY, Apr. 2020. [7] J. Chen, M. Li, and H. Xu，“向机器学习器出售数据：通过成本信号定价，” in 第*39*届国际机器学习会议论文集. PMLR, Jun. 2022, pp. 3336–3359. [8] H. Yu and M. Zhang，“基于数据质量的数据定价策略，”《计算机与工业工程》, vol. 112, pp. 1–10, Oct. 2017. [9] Y. Tian, Y. Ding, S. Fu, and D. Liu，“基于夏普利值的数据边界与数据定价，” *IEEE Access*, vol. 10, 14 300, 2022 14 288pp. – . [10] P. Koutris, P. Upadhyaya, M. Balazinska, B. Howe, and D. Suciu，“基于查询的数据定价，”《*ACM*学报》, vol. 62, no. 5, pp. 43:1–43:44, Nov. 2015. [11] J. Pei，“数据定价综述：从经济学到数据科学，” *IEEE*知识与数据工程汇刊, pp. 1–1, 2020. [12] C. Niu, Z. Zheng, S. Tang, X. Gao, and F. Wu，“小传感器创造大收益：河豚隐私保护下的时间序列数据交易，” in *IEEE INFOCOM 2019 - IEEE*计算机通信会议, Apr. 2019, pp. 568–576. [13] B. Zhou, J. Pei, and W. Luk，“社交网络数据隐私保护发布的匿名化技术简要综述，”《*ACM SIGKDD*探索通讯》, vol. 10, no. 2, pp. 12–22, Dec. 2008.[14] C. Li, D. Y. Li, G. Miklau, and D. Suciu，“隐私数据定价理论，”《*ACM*数据库系统汇刊》, vol. 39, no. 4, pp. 34:1– 34:28, Dec. 2015. [15] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, and G. Srivastava，“联邦学习安全与隐私综述，”《未来一代计算机系统》, vol. 115, pp. 619–640, Feb. 2021.[16] C. Li, Y. Yuan, and F.-Y. Wang，“一种基于区块链赋能的联邦学习的数据交易市场新框架，” in *2022*年*IEEE*第*25*届智能交通系统国际会议 *(ITSC)*, Oct. 2022, pp. 3392–3397. [17] Y.-H. Chang and M.-S. Lee，“将马尔可夫决策过程融入遗传算法以制定股票市场交易策略，”《应用软计算》, vol. 52, pp. 1143–1153, Mar. 2017. [18] C. Jiang and C. Wang，“高速铁路定价：对社会福利的影响，”《交通研究*E*辑：物流与交通评论》, vol. 155, p. 102484, Nov. 2021. [19] S. Wu and P. Pavlou，“信息服务的最优固定上限定价，”《信息系统协会学报》, vol. 20, no. 10, Oct. 2019. [20] C.-H. V. Lin, C.-C. J. Huang, Y.-H. Yuan, and Z.-s. S. Yuan，“面向订阅式物联网数据交易的全去中心化基础设施，” in *2020*年*IEEE*国际区块链会议 *(Blockchain)*, Nov. 2020, pp. 162–169. [21] J. Li, X. Ni, Y. Yuan, and F.-Y. Wang，“面向比特币交易动态确认博弈的新型广义二价拍卖机制，” *IEEE*服务计算汇刊, vol. 15, no. 3, pp. 1436–1447, May 2022.[22] J. Li, Y. Yuan, and F.-Y. Wang，“区块链挖矿中比特币交易排序的新型广义二价拍卖机制，” 决策支持系统, vol. 124, p. 113094, Sep. 2019. [23] S. Mehta, M. Dawande, G. Janakiraman, and V. Mookerjee，“如何出售数据集？数据变现的定价策略：第20届ACM经济学与计算会议（EC 2019），” *ACM EC 2019*——*2019*年*ACM*经济学与计算会议论文集, p. 679, Jun. 2019.[24] R. C. Fernandez, P. Subramaniam, and M. J. Franklin，“数据市场平台：通过交易数据资产解决数据问题，” *VLDB*基金会论文集, vol. 13, no. 12, pp. 1933–1947, Aug. 2020. [25] R. Maestre, J. Duque, A. Rubio, and J. Arevalo，“公平动态定价的强化学习，” in 智能系统与应用, ser. 智能系统与计算进展, K. Arai, S. Kapoor, and R. Bhatia, Eds. 查姆: 施普林格国际出版公司, 2019, pp. 120–135. [26] X. Kong, D. Kong, J. Yao, L. Bai, and J. Xiao，“基于长短期记忆网络和强化学习的需求响应在线定价，” 应用能源, vol. 271, p. 114945, Aug. 2020. [27] H. Xu, H. Sun, D. Nikovski, S. Kitamura, K. Mori, and H. Hashimoto，“负荷服务实体联合竞价与定价的深度强化学习，” *IEEE*智能电网汇刊, vol. 10, no. 6, pp. 6366–6375, Nov. 2019. [28] E. Krasheninnikova, J. Garc´ıa, R. Maestre, and F. Fern´andez，“保险业定价策略优化的强化学习，” 人工智能工程应用, vol. 80, pp. 8–19, Apr. 2019. [29] W. Shen, B. Peng, H. Liu, M. Zhang, R. Qian, Y. Hong, Z. Guo, Z. Ding, P. Lu, and P. Tang，“强化机制设计：在赞助搜索拍卖动态定价中的应用，” *AAAI*人工智能会议论文集, vol. 34, no. 02, pp. 2236–2243, Apr. 2020. [30] L. Harris，“做市商—吃单者定价对市场报价的影响，”南加州大学马歇尔商学院工作论文。见 *http://bschool. huji. ac. il/. upload/hujibusiness/Maker-taker. pdf*, 2013. [31] A. Nair, A. Gupta, M. Dalal, and S. Levine，“AWAC：利用离线数据集加速在线强化学习，” Apr. 2021. [32] T. Xie, N. Jiang, H. Wang, C. Xiong, and Y. Bai，“策略微调：衔接样本高效的离线与在线强化学习，” 神经信息处理系统进展, vol. 34, pp. 27 395– 27 407, 2021. [33] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel，“基于认知的公平性，” Nov. 2011. [34] J. Li, R. Qin, and F.-Y. Wang，“管理的未来：从道到智能组织与智能运营，”*IEEE*系统、人类与控制论汇刊：系统, pp. 1–11, 2022. [35] J. Li, Y. Yuan, and F.-Y. Wang，“利用排队博弈模型分析比特币交易手续费，”电子商务研究, vol. 22, no. 1, pp. 135–155, Mar. 2022. [36] J. Li, Y. Yuan, S. Wang, and F.-Y. Wang，“比特币区块链中的交易排队博弈，” in *2018*年*IEEE*智能车辆研讨会 *(IV)*, Jun. 2018, pp. 114–119.