

需求驱动的柔性数据供应链构建及定价

任洪润^{1,2}, 朱扬勇^{1,2,3}

1. 复旦大学计算与智能创新学院, 上海 200438;
2. 上海市数据科学重点实验室, 上海 200438;
3. 上海数据研究院, 上海 200438

摘要

随着数据要素深度融入经济活动, 数据产品的生产与流通体系建设加速推进, 其中数据供应链建设尤为关键。首先, 分析了数据供应链与实物供应链的差异, 归纳出数据供应链的“网络供应与低延迟”、“‘0/1’库存”和“需求驱动的结构性”三大特征。其次, 针对参与者迅速匹配与全链条协同定价两大挑战, 提出了一种柔性数据供应链构建与定价模型——FlexChain。FlexChain由架构师主导, 一方面利用多维适配度评分机制实现生态拓扑的动态实时重构, 另一方面立足于数据加工中算力与合规成本超线性增长的特质, 引入非线性二次成本函数, 构建两阶段Stackelberg博弈定价模型。最后, 通过理论分析证明了该博弈均衡的存在性与唯一性, 有效实现了数据供应链各方利润的最大化与终端出价的传导优化。该框架为动态高频下的数据产品生产流通体系构建提供了可行的路径。

关键词

数据供应链; 柔性制造; Stackelberg博弈; 数据供应链定价

中图分类号: TP311

文献标志码: A

doi: 10.11959/j.issn.2096-0271.BDR26251

Demand-driven construction mechanisms and pricing models for flexible data supply chains

Ren Hongrun^{1,2}, Zhu Yangyong^{1,2,3}

1. College of Computer Science and Artificial Intelligence, Fudan University, Shanghai 200438, China
2. Shanghai Key Laboratory of Data Science, Shanghai 200438, China
3. Shanghai Data Research Institute, Shanghai 200438, China

Abstract

As data elements have been deeply integrated into economic activities, the construction of production and circulation systems for data products has been accelerated, among which the development of data supply chains has been particularly critical. First, the differences between data supply chains and physical supply chains were analyzed, and three distinctive features of data supply chains were summarized, namely "networked supply and low latency", "'0/1' inventory", and "demand-driven structural flexibility". Second, to address the two major challenges of rapid participant matching and coordinated pricing across the entire chain, a flexible data supply chain construction and pricing

framework, termed FlexChain, was proposed. Dominated by an architect, FlexChain employed a multi-dimensional fitness scoring mechanism to realize dynamic real-time reconfiguration of the ecological topology on the one hand, and introduced a nonlinear quadratic cost function to construct a two-stage Stackelberg game pricing model on the other hand, based on the superlinear growth characteristics of computational and compliance costs in data processing. Finally, the existence and uniqueness of the game equilibrium were proved through theoretical analysis, by which the profit maximization of all parties involved in the data supply chain and the transmission optimization of the end-user bidding price were effectively achieved. This framework has provided a feasible pathway for the construction of production and circulation systems for data products under dynamic and high-frequency scenarios.

Key words

data supply chain, flexible manufacturing, Stackelberg game, data supply chain pricing

0 引言

随着数据生产与流通的持续发展, 数据产品生产的分工日渐明晰^[1], 出现了专注于原始数据生产、质量提升、数据标注及模型训练等环节的专业厂商^[2]。这些厂商之间需要协同合作, 以确保数据生产的高效、可靠^[3], 因此, 构建数据供应链已成为必然趋势。目前, 多个领域已出现数据供应链的雏形。例如, A公司(教育领域)与B公司(医疗领域)各自需要一份市场分析报告以了解行业动态, 但他们难以直接对接研究机构并获取多种数据源。此时, 两家公司可通过市场研究服务平台(如 MarketResearch.com)与 IDC、Gartner 等研究机构对接。市场研究服务平台还可协助研究机构接入来自 Nielsen、Euromonitor 等供应商的教育或医疗数据。在此链条中, A、B公司为数据消费者, 平台为数据构建者, 研究机构(作为制造商)是核心企业, 而多个数据源则是供应商, 整体上有效构成了数据供应链的基本框架。面对成千上万同类消费者, 平台不仅能满足个性化需求, 还能整合资源, 实现数据产品的实时生产与批量供给。虽然

许多数据产品生产流程均遵循该模式, 但目前尚无对其结构的正式研究。为充分释放数据价值, 并确保组织间数据安全合规流通, 探索其系统构建势在必行。

数据供应链基于网络传输, 传统供应链中(如库存管理^[4-5])的难点在数据供应链中不复存在。同时, 其建立与传输要求响应速度快, 且各环节之间的互动需保持透明公开。因此, 其建设面临新挑战: 一是在多样化生态环境中, 如何快速确定参与者并完成建链; 二是如何构建模型以确定全链条价格。针对上述挑战, 本文提出了一种柔性数据供应链构建与定价框架——FlexChain。本文的主要贡献如下。

(1) 设计了参与者选择方法。首次提出由数据供应链架构师负责构建和运营的柔性数据供应链。针对任意购买需求, 架构师能快速确定参与者(制造商、供应商)并完成数据生产与交付。这种模式降低了供应链对核心企业的依赖, 提升了系统的灵活性与透明度。

(2) 设计了数据供应链定价模型。在考虑供应商与制造商利润最大化目标及架构师获利的基础上, 构建了相应的定价模型。通过求解最优化问题, 找到了满足 Stackelberg 均衡的最优定价策略, 确定了最优数据价格与规模。

1 相关工作

近年来，随着数据要素市场的快速发展，大量研究聚焦于数据交易机制的设计。朱扬勇等^[6]指出，数据需经深度加工转化为数据产品，方能实现安全高效流转，这为交易向纵向供应链延伸奠定了理论基础。在此基础上，现有研究进一步探讨了端到端模型市场^[7]、单边与双边供给模式^[8-9]，以及移动感知交易平台^[10]等具体场景；近年来，有学者开始从供应链视角审视多方博弈问题，例如，聚焦价值确权与流转优化^[11]，或引入多阶段博弈以协调多源网络中的收益冲突^[12]。然而，上述研究多关注松散的单次数据交易，且其链条拓扑通常被假设为长期刚性不变，在面对高度动态的数据要素市场需求时，容易产生显著的性能退化。相比之下，本文提出的 FlexChain 框架由架构师主导，能够根据购买需求动态编排上游供应商与中游制造商，实现多层级协同生产交付。

数据定价模型的建立对规范数据要素市场具有重要意义。相关研究^[13]探讨了基于市场类型的定价分类。现有定价研究主要集中于两类场景：一是数据原材料与流

式数据供应场景^[14-15]，包括元组粒度定价、动态定价以及群智感知激励等问题；二是下游定制化数据产品场景^[16]，涉及联邦 Shapley 值定价和基于模型的定价等方法。在深度协作方面，已有研究采用拍卖、Stackelberg 博弈、Shapley 值、历史记录及质量定价等运筹学方法，探讨多层级动态利润分配。

为更清晰地展示本文所提方法与现有研究的本质差异，从组织模式、供应链柔性、交易方式和成本特征等关键维度进行了系统对比，具体见表1。

由表1可知，现有研究在多源数据加工与流通过程中存在以下两点显著局限：（1）在规模与结构特质方面，缺乏对多源数据流转中动态重组特征的量化刻画。本文所提方法设计了供应链构建机制，允许架构师针对变化的数据产品规格与时限，在生态网络中实时重构拓扑关系，打破对固定主体的依赖。（2）在成本函数构建方面，现有研究多假设数据复制、加工或传输的边际成本为零，呈现线性特征，这忽略了实际流通中数据清洗、去标识化、多源数据级联加工时，算力与合规成本随规模扩大而加速上升的客观规律。通过在博弈定价中引入非线性二次成本函数，能够更客观地拟合数据全生命周期生产加工的实际成本。

表1 现有相关研究与所提方法的系统对比

维度	单点数据交易机制	传统刚性供应链	柔性数据供应链(本文所提方法)
组织模式	偶发性、单次双边交易	依赖固定核心企业,链条拓扑长期稳定	架构师主导模式,按需求动态组网
供应链柔性	无系统级供应链概念,缺乏层级协同	无法适配变动需求,固定配置易导致整体性能退化	高度柔性,可基于适配度评分动态重构
交易方式	倾向单批次、小金额的离散交易	依赖大批量实物流动与复杂的物理库存管理	考虑多源数据计量量汇聚时的数量转化与加工损耗,满足线性加权关系
成本特征	多假设复制边际成本为零,忽略过程损耗	成本主要源于实物仓储建设与物流损耗	考虑深度加工时算力与合规边际成本超线性增长(采用二次成本函数)

综上所述，有必要将数据生产过程视为柔性供应链系统，并在梳理相关工作的基础上，设计能实现全链条利润协调的柔性定价模型。

2 问题描述

2.1 数据供应链的特征

传统物品供应链的诸多难题与物流息息相关^[17-18]，而数据供应链依托网络传输，具有以下独有特点。

(1) 网络供应与极低延迟。实体物品的流通依赖交通工具，运输风险与成本高昂^[19]。数据产品则通过互联网传输，周期短、效能高，能够支持单批次、极小金额的快速交易。除了实时生成的流式数据存在微小延迟风险外，静态数据的延迟风险几乎可忽略不计^[20-21]。

(2) “0”或“1”库存。库存管理在传统供应链中占据重要地位。由于数据复制的边际成本几乎为零^[22]，其库存呈现极端化特征：“0库存”适用于实时产生、随用随弃的流式数据；“1库存”则指需长期保存于存储系统中的静态数据集，因复制成本极低，通常全网仅需存储一份即可满足多用户访问需求。

(3) 需求驱动的结构柔性。传统刚性供应链一旦建立便长期保持稳定，而数据供应链的生存周期取决于单次或单批次的数据购买需求。由于数据要素具有零复制成本、随用随弃或按需汇聚的物理特质，架构师可以根据不断变化的产品规格、质量标准和时限要求，在网络生态中实时重构拓扑关系。这种“即建即拆、动态演进”的特质，构成了柔性数据供应链的核心内涵。

2.2 两个核心问题

数据供应链所具有的高度柔性特征，决定了其构建的核心不再聚焦于实物调度，而在于如何实现动态参与者选择与全链条定价。

(1) 参与者选择。传统模式下，供应链通常由核心企业选择有限主体，且构建后保持相对稳定。然而，面对庞大的数据市场主体和消费者不断变化的动态需求，传统核心企业难以有效对接海量潜在合作方，也缺乏足够资源用于评估选择最合适的伙伴。因此，如何在数据市场中快速识别参与者并实时重构供应链是一大挑战。

(2) 数据供应链定价。在数据供应链中，传统定价模型所倚重的运输成本和存储成本不再突出。因此，有必要基于数据产品零复制成本等物理特征，重新设计专属于数据供应链的成本函数与定价模型，以准确反映多层级级联加工下数据价值的累积与传递，这也是实现全链条协调的难点所在。

3 柔性数据供应链构建

本文所提的柔性数据供应链构建与定价框架（FlexChain）主要由两部分核心体系协同构成：一是由架构师主导的拓扑动态重构机制；二是两阶段Stackelberg博弈定价模型。

3.1 柔性的形式化定义

数据供应链需要随需求变化动态重组，这与一旦建立便趋于稳定的传统刚性供应链存在本质差异。为刻画这一动态特征，将数据消费者的购买需求定义为三元组 $T=(\tau, p^T, \Phi)$ ，其中 τ 为数据产品规格（含产品类型、质量标准、交付时限等）， p^T 为消

消费者支付的数据产品单价, Φ 为需求的总数据量。设 $M = \{m_1, \dots, m_j\}$ 与 $S = \{s_1, \dots, s_n\}$ 分别为候选制造商与供应商集合。给定购买需求 T , 架构师从中选定一个制造商 $m \in M$ 和一组供应商 $S_D \subseteq S$, 形成参与方配置 (m, S) 。在刚性数据供应链下, 初始选定的参与方配置 (m, S) 在交付周期内保持不变。当购买需求从 T 变为 T' 时, 由于固定配置无法适配新规格, 供应链整体性能 (各方利润之和 P) 出现显著退化, 即 $P(T', m, S_D) \ll P(T, m, S_D)$ 。相比之下, 柔性数据供应链允许架构师针对新需求, 通过最大化适配度评分函数 $\text{Compatibility}(\cdot | T)$ 动态求解并重构配置, 即 $(m^*(T), S^*(T)) = \arg\max_{m \in M, S_D \subseteq S} \text{Compatibility}(m, S_D | T)$ 。其中, m^* 和 S^* 分别表示供应链架构师针对消费者购买需求所选择的最优制造商与最优供应商。

基于上述设定, 提出柔性效率增益命题: 对任意需求变动 $T \rightarrow T'$, 柔性供应链在新需求下的性能严格不低于刚性供应链, 即 $P(T', m^*(T'), S^*(T')) \geq P(T', m^*(T), S^*(T))$, 等号仅在需求变动前后指向同一最优配置时成立。

证明: 柔性配置 $(m^*(T'), S^*(T'))$ 是基于新需求 T' 优化求解的 $\arg\max$ 全局最优解, 而刚性配置 $(m^*(T), S^*(T))$ 仅是新需求优化空间中的一个可行解。根据最优化理论, 最优解的目标函数值必然不小于任意可行解, 证毕。

3.2 供应链角色构成与构建流程

柔性数据供应链的构建与协同运营依赖四类核心主体的参与, 各角色的具体定义与职责如下。

(1) 数据消费者: 供应链流转的起点与价值终点, 负责根据实际业务场景提交

包含产品规格、单价及数据量的个性化购买需求, 并支付相应费用。

(2) 数据供应链架构师: 全链条的中立协同中枢, 负责解构消费者需求、动态编排上游候选主体、监督合规流转并执行多方价值结算。

(3) 数据产品制造商: 供应链的核心加工节点, 负责将多源原始数据加工转化为高附加值的数据产品。

(4) 数据供应商: 供应链的底层原材料源头, 负责提供合规、可流通的原始数据集或特征数据。

基于上述角色间的协作, 柔性数据供应链的动态构建流程可精炼为以下 3 个关键阶段。

阶段一: 需求解构。架构师接收消费者的个性化购买需求, 将其转化为标准化的加工生产约束与合规审计指标, 并向全网候选主体发布建链意向。

阶段二: 拓扑重构。架构师基于预存的候选方多维质量档案, 运用适配度评分机制, 从海量候选者中快速重组最优的制造商与供应商配置, 完成拓扑动态组网。

阶段三: 交付结算。上游数据流转与加工完成后, 架构师将最终数据产品交付数据消费者, 并根据既定博弈定价协议, 完成各生产节点利润的清算与分配。

3.3 柔性供应链参与者选择机制

柔性数据供应链构建的核心工程任务是, 在海量候选主体中快速、精准地识别并筛选出最适配的制造商与供应商。当数据消费者提交购买需求 T 后, 供应链架构师充当第三方中立的编排纽带, 发布生产与供应需求, 并利用多维适配度评分函数对报名参与的候选方进行客观评估。相关符号及含义见表 2。

(1) 数据产品制造商筛选

架构师重点评估有意愿参加的候选制造商*j*的数据产品制造能力 v_j^T 。同时，为筛选出具备合规和经济规模的优质主体，将市场公开透明的固定合规门槛费用（如上链存证费、安全审计费等固定资产投入） $\mathcal{F}_j^T$ 纳入考量。制造商*j*的供应链适配度评分 r_j^T 刻画为：

$$r_j^T(v_j^T, \mathcal{F}_j^T) = \theta_1 v_j^T + \theta_2 \mathcal{F}_j^T \tag{1}$$

其中， θ_1 和 θ_2 分别表示数据产品制造商*j*的制造能力 v_j^T 和支付给供应链架构师的费用 $\mathcal{F}_j^T$ 在供应链适配度评分中的权重， $\mathcal{F}_j^T$ 为固定值。

(2) 数据供应商筛选

面对上游众多的数据供应商，架构师

根据他们提供的数据质量估计 $\bar{q}_i$ 、拥有并可流通的数据规模 λ_i ，以及需支付给供应链架构师的费用 f_i 进行综合考量。数据供应商*i*的供应链适配度评分 r_i 刻画为：

$$r_i(\bar{q}_i, \lambda_i, f_i) = \eta_1 \bar{q}_i + \eta_2 \lambda_i + \eta_3 f_i \tag{2}$$

其中， η_1 、 η_2 和 η_3 分别表示数据供应商的数据质量 $\bar{q}_i$ 、拥有的数据规模 λ_i 和支付给供应链架构师的费用 f_i 在供应链适配度评分中的比重， f_i 为固定值。

4 柔性数据供应链定价

4.1 定价问题的形式化定义

数据供应链定价的目标是找到满足Stackelberg均衡的最优定价策略组 $\langle p^*, \varphi^* \rangle$ ，使得供应链参与者能够最大化各自的利润。需要注意的是，没有参与者愿意采用会导致自身利润下降的替代策略。

定义1：策略组 $\langle p^*, \varphi^* \rangle$ 。将基于供应链参与者利润最大化的最优定价问题建模为一个两阶段Stackelberg博弈模型，从而确定最优策略组 $\langle p^*, \varphi^* \rangle$ 。供应链中的每个参与者都试图通过控制策略组中的一个参数来最大化自己的利润：在数据消费者给定单位数据产品价格 p' 的情况下，数据产品制造商通过调整单位数据价格 p 来控制支付给数据供应商的费用，数据供应商则通过调整数据规模 φ_i 来影响数据产品制造商的利润。

定义2：数据供应商利润 Φ_i 。数据供应商*i*的利润是从数据产品制造商处获得的费用减去供应数据的成本及支付给供应链架构师的费用，可表示为：

$$\Phi_i(p, \varphi_i) = p\varphi_i\chi_i - C_i(\bar{q}_i, \varphi_i)\chi_i - f_i\chi_i \tag{3}$$

表2 相关符号及含义

符号	含义
p^T	单位数据产品价格
p	单位数据价格
φ_i, φ	数据供应商 <i>i</i> 提供的数据规模, 数据产品规模
$r_j^T(\cdot), \theta_1, \theta_2$	数据产品制造商 <i>j</i> 与供应链的适配度函数和参数
$r_i(\cdot), \eta_1, \eta_2, \eta_3$	数据供应商 <i>i</i> 与供应链的适配度函数和参数
$\mathcal{F}_j^T$	数据产品制造商 <i>j</i> 支付给供应链架构师的费用
v_j^T	数据产品制造商 <i>j</i> 的产品制造能力
f_i	数据供应商 <i>i</i> 支付给供应链架构师的费用
$\bar{q}_i$	数据供应商 <i>i</i> 提供的数据质量的估计
λ_i	数据供应商 <i>i</i> 拥有的数据规模
χ_i	数据供应商 <i>i</i> 是否被供应链架构师选中
χ_j	数据产品制造商 <i>j</i> 是否被供应链架构师选中
$C_i(\cdot), \alpha_i, \beta_i$	数据供应商 <i>i</i> 的成本函数和参数
$\Psi(\cdot)$	数据产品制造商的利润函数
$C^T(\cdot), \gamma, \delta$	数据产品制造商的成本函数和参数

其中, p 表示单位数据价格 (由数据产品制造商支付给数据供应商); φ_i 表示数据规模; $\chi_i \in \{0, 1\}$, $\chi_i = 0$ 表示数据供应商没有被架构师选中, $\chi_i = 1$ 表示被选中; f_i 表示数据供应商 i 支付给供应链架构师的费用。

数据供应商 i 的成本函数, 可采用二次成本函数^[23] $C_i(\bar{q}_i, \varphi_i)$ 表示:

$$C_i(\bar{q}_i, \varphi_i) = (\alpha_i \varphi_i^2 + \beta_i \varphi_i) \bar{q}_i \quad (4)$$

其中, $\bar{q}_i$ 表示供应链架构师对数据供应商提供的估计。

假设 1 (规模线性映射): 由于数据要素流转的多样性, 本模型聚焦于价值与规模呈线性映射的数据级联场景, 如多源静态数据集的合并打包、分布式群智感知数据的计量汇聚。在此场景下, 下游数据产品的规模与上游输入的数据规模满足加性关系, 即 $\varphi = \sum_{i=1}^N \varphi_i \chi_i$ 。对于非线性的复杂模型训练场景, 可通过引入规模系数进行等价转换, 使本模型不失一般性。

定义 3: 数据产品制造商利润。数据产品制造商通过向数据消费者出售数据产品而获得利润, 其利润函数 $\Psi(p, \varphi)$ 为数据产品制造商从数据消费者处获得的费用减去支付给数据供应商的费用、成本和支付给供应链架构师的费用, 表示为:

$$\Psi(p, \varphi) = p^T \chi_j \sum_{i=1}^N \varphi_i \chi_i - p \chi_j \sum_{i=1}^N \varphi_i \chi_i - C^T(\varphi) \chi_j - \mathcal{F}_j^T \chi_j \quad (5)$$

其中, $p^T \chi_j \sum_{i=1}^N \varphi_i \chi_i$ 表示数据产品制造商 i 从数据消费者处获得的费用, p^T 为数据消费者确定的单位数据产品价格; $\chi_j \in \{0, 1\}$, $\chi_j = 0$ 表示数据产品制造商没有被供应链架构师选中, $\chi_j = 1$ 表示数据产品制造商被选中。 $p \chi_j \sum_{i=1}^N \varphi_i \chi_i$ 表示数据产品制造商 j 支付给数据供应商 i 的费用。 $C^T(\varphi) \chi_j$ 表示数据产品制造商的成本。 $\mathcal{F}_j^T \chi_j$ 表示数据产品制

造商支付给供应链架构师的费用。

采用二次成本函数 $C^T(\varphi)$ 表示数据产品制造商 j 的生产成本:

$$C^T(\varphi) = \gamma (\sum_{i=1}^N \varphi_i \chi_i)^2 + \delta \sum_{i=1}^N \varphi_i \chi_i \quad (6)$$

鉴于数据供应链具有“0 库存”或“1 库存”及网络传输的特质, 传统供应链中与物理仓储、物流损耗相关的线性成本不再适用。此处采用二次成本函数的经济学含义在于: 数据供应商的成本主要源于数据采集、隐私去标识化, 以及高频网络带宽调度的计算与合规边际成本; 而制造商的二次成本 $C^T(\varphi)$ 则精准反映了多源数据融合加工时, 算力消耗与模型训练成本随数据规模呈超线性增长的真实物理特质。

4.2 Stackelberg 博弈与均衡求解

假设 2 (完全信息环境的合理性): 本定价博弈模型在完全信息框架下展开。其合理性在于: 柔性数据供应链引入了中立的架构师角色。架构师预先积累了候选供应商的多维档案, 并在建链时负责统一评估与核验供应商的数据质量估计与历史合规成本。制造商在进行定价决策时, 可通过架构师提供的数据资源信息和标准化接口, 获知受邀参与链条的供应商的基本成本参数与质量画像, 从而消除了传统双边交易中的严重信息不对称, 使得完全信息 Stackelberg 博弈具有现实可行性。

定义 4: Stackelberg 均衡。数据供应链定价的目标是找到满足 Stackelberg 均衡的最优策略组 $\langle p^*, \varphi^* \rangle$, 而求解最优策略组的过程被建模为一个两阶段 Stackelberg 博弈。在该博弈中, 数据产品制造商充当领导者, 数据供应商充当追随者。最终目标是在确保追随者利润也达到最大化的前提下, 最大化领导者的利润。

阶段1：数据产品制造商在最优定价策略组下的利润。

$$\Psi(p^*, \varphi^*) \geq \Psi(p, \varphi^*) \quad (7)$$

阶段2：数据供应商在最优定价策略组下的利润。

$$\Phi_i(p^*, \varphi_i^*) \geq \Phi_i(p^*, \varphi_i, \varphi_{-i}^*) \quad (8)$$

当且仅当上述不等式满足时，最优定价策略 $\langle p^*, \varphi^* \rangle$ 构成 Stackelberg 均衡。

其中， φ_{-i}^* 表示除了数据供应商 i 以外的所有被供应链架构师选中的供应商的最优策略。

因此，所有数据供应商的最优策略可以表示为 $\varphi^* = \varphi_{-i}^* \cup \{\varphi_i^*\}$ 。定义4表明，任何供应链参与者都不能通过偏离最优策略来提高自己的利润。

本小节采用逆向归纳法对 Stackelberg 博弈的最优定价策略组 $\langle p^*, \varphi^* \rangle$ 进行求解。在阶段2中，分析数据供应商 i 的最优策略 φ_i^* ，即给定单位数据价格 p 时，数据供应商 i 提供的最优数据规模 φ_i^* 。然后，在阶段1中，求解数据产品制造商 i 的最优策略 p^* ，即在给定单位数据产品价格 p^T 时数据产品制造商的最优单位数据价格 p^* 。证明了最优定价策略 $\langle p^*, \varphi^* \rangle$ 已形成唯一的 Stackelberg 均衡，数据供应链中的每个参与者都不会单方面偏离其最优策略。

定理1：阶段2中，给定单位数据价格 p ，数据供应商 i 的最优数据规模 φ_i^* 由式（9）确定。

$$\varphi_i^* = \frac{p - \beta_i \bar{q}_i}{2\alpha_i \bar{q}_i} \quad (9)$$

对式（9）进行敏感性分析可知，在单位数据价格 p 固定的情况下，最优数据规模 φ_i^* 与数据质量 $\bar{q}_i$ 呈负相关关系。这符合数据要素市场的边际成本特征：数据质量

$\bar{q}_i$ 越高，意味着供应商在数据清洗、脱敏、确权等维度的初始资源投入与校验成本越高（表现为式（4）中成本随 $\bar{q}_i$ 放大）。因此，理性的供应商会采取与自身质量水平相匹配的数据规模策略。值得注意的是，在相同的单位数据价格下，高质量供应商因边际成本更低而提供更大的数据规模，这符合数字经济中“高质量、高产出”的基本规律。

证明：计算式（3）中数据供应商 i 的利润函数 $\Phi_i(p, \varphi_i)$ 关于 φ_i 的一阶与二阶偏导，可得：

$$\frac{\partial^2 \Phi_i(p, \varphi_i)}{\partial (\varphi_i)^2} = -2\alpha_i \bar{q}_i < 0 \quad (10)$$

由此可知， $\Phi_i(p, \varphi_i)$ 关于 φ_i 的二阶偏导数严格小于0，即目标函数为严格凸函数，存在唯一全局最优解。根据一阶偏导数 $\frac{\partial \Phi(p, \varphi_i)}{\partial \varphi_i} = 0$ ，求解得出数据供应商 i 的最优数据规模 φ_i^* ，如式（9）所示。

定理2：阶段1中，根据阶段2确定的数据供应商 i 的最优策略 φ_i^* ，数据产品制造商确定的最优单位数据定价 p^* 可以由式（11）计算。

$$p^* = \frac{p^T A + B + 2\gamma BA - \delta A}{2A(1 + \gamma A)} \quad (11)$$

其中， $A = \sum_{i=1}^N \frac{1}{2\alpha_i \bar{q}_i}$ ， $B = \sum_{i=1}^N \frac{\beta_i}{2\alpha_i}$ 。

证明：将式（9）代入式（5），求解数据产品制造商的利润函数 $\Psi(p, \varphi)$ 关于 p 的一阶与二阶偏导，可得：

$$\frac{\partial^2 \Psi(p, \varphi)}{\partial p^2} = -2A - 2\gamma A^2 < 0 \quad (12)$$

因此， $\Psi(p, \varphi)$ 关于 p 的二阶偏导数严格小于0，即目标函数为严格凸函数，存在唯一全局最优解。根据一阶偏导数

$\frac{\partial \Psi(p, \varphi)}{\partial p} = 0$, 求解得出数据产品制造商的最优单位数据价格 p^* , 如式 (11) 所示。

根据式 (11), 可以得到数据产品制造商的最优价格 p^* 。然后, 将 p^* 代入式 (9) 可以得到每个数据供应商的最优策略 φ_i^* 。由此, 基于数据消费者提供的单位数据产品价格, 满足数据供应链参与者利润最大化的最优定价策略组 $\langle p^*, \varphi^* \rangle$ 得以确定。

接下来, 分析 Stackelberg 均衡的存在性和唯一性。

定理 3: 基于供应链参与者利润最大化定价模型所确定的最优定价策略组 $\langle p^*, \varphi^* \rangle$ 构成唯一的 Stackelberg 均衡。

证明: 根据式 (11), 最优单位数据价格 p^* 仅由单位数据产品价格 p^T , 数据供应商 i 提供的数据质量的估计 $\bar{q}_i$, 以及成本参数 α 、 β 、 γ 、 δ 计算得出。当数据供应商 i 持有最优策略时, 数据产品制造商 j 的利润 $\Psi(p, \varphi)$ 仅随 p 变化。根据定理 2, 数据产品制造商只有在 p^* 处能获得最大利润, 任何其他 $p \neq p^*$ 都会导致更低的利润, 因此式 (7) 成立。类似地, 当确定最优单位数据价格 p^* 和数据规模 φ_{-i}^* 时, 数据供应商 i 只能在 φ_i^* 处获得最大利润, 因此式 (8) 成立。

因此, 最优定价策略组 $\langle p^*, \varphi^* \rangle$ 形成了数据供应链中的唯一 Stackelberg 均衡, $\langle p^*, \varphi^* \rangle$ 为求解得出的满足数据供应链中各方利润最大化的最优策略组。

为了聚焦于柔性结构下的级联定价机制, 本文所设计的柔性数据供应链运行于数据可信流通域 (如采用隐私计算、数据沙箱、分布式账本等技术手段的市场环境)。在此环境下, 数据要素的使用权与所有权实现分离, 数据“可用不可见、控权不控量”。因此, 本模型隐去了解析后的数据二次转卖、地下分销等负外部性行

为对博弈均衡的冲击。全链条的利润流转严格基于单次按需构建的交付合同, 这保障了定价模型在合规要素市场中的稳健性。

5 结束语

柔性数据供应链赋予了架构师针对任意需求快速建链的能力, 其核心挑战在于如何统筹各方利益, 以确保终端出价有效传导, 实现全链条透明、公正的利润分配。未来拓展方向如下。(1) 信息不完全情境的拓展。本文所提定价博弈模型在完全信息框架下展开, 假定架构师全面掌握候选供应商的质量与成本档案。未来研究将逐步放松此假设, 引入信号博弈或贝叶斯 Stackelberg 博弈。当供应商故意隐瞒真实成本 (成本系数成为其私有信息时), 需要深入研究架构师如何设计一套激励相容的逆向选择机制, 以防范供应商的道德风险与利益冲突。(2) 有限理性参与者的动态演化。虽然在完全理性假设下已经证明了唯一的 Stackelberg 均衡解, 但在实际要素市场中, 供需各方往往表现出有限理性行为。下一步工作将结合演化博弈与复制动态方程, 分析在非严格最优策略下, 各参与方利润函数在长期反复博弈中的策略流转轨迹与动态演化收敛性。(3) 大规模非线性场景的实证与修正。当前研究聚焦于价值与规模呈线性映射的数据级联场景。未来拓展将进一步探讨大模型深度训练等非线性算力消耗场景, 通过引入非线性规模转换系数修正制造商的成本函数, 并引入真实数据交易样本验证修正模型的稳健性与实用性。

综上所述, 该机制有望在数字经济发展中产生广泛的应用价值。

参考文献:

- [1] Fernie J. Relationships in the supply chain[M]//Logistics and retail management insights into current practice and trends from leading experts. CRC Press, 2023: 23–46.
- [2] Murphy K, Schölkopf B, Shah N B, et al. Double or nothing: multiplicative incentive mechanisms for crowdsourcing[J]. Journal of Machine Learning Research, 2016, 17(1): 5725–5776.
- [3] Khanlarzade N, Farughi H. Modeling the Stackelberg game with a boundedly rational follower in deterioration supply chain-based interaction with the leader's hybrid pricing strategy[J]. Expert Systems with Applications, 2024, 237: 121302.
- [4] Wang M Z. Design and supply chain management of intelligent logistics system using cloud computing under Internet of Things[J]. International Journal of Grid and Utility Computing, 2023, 14(2/3): 216–228.
- [5] Altekarr V. Supply chain management[M]. 2nd. New Delhi: PHI Titles, 2023.
- [6] 朱扬勇, 叶雅珍. 从数据的属性看数据资产[J]. 大数据, 2018, 4(6): 65–76.
- Zhu Y Y, Ye Y Z. Defining data assets based on the attributes of data[J]. Big Data Research, 2018, 4(6): 65–76.
- [7] Liu J F, Lou J, Liu J X, et al. Dealer: an end-to-end model marketplace with differential privacy[J]. Proceedings of the VLDB Endowment, 2021, 14(12): 2747–2750.
- [8] Wei X D, Nault B R. Monopoly versioning of information goods when consumers have group tastes[J]. Production and Operations Management, 2014, 23(6): 1067–1081.
- [9] Stahl F, Schomm F, Vossen G, et al. A classification framework for data marketplaces[J]. Vietnam Journal of Computer Science, 2016, 3(3): 137–143.
- [10] An B Y, Xiao M J, Liu A, et al. Crowdsensing data trading based on combinatorial multi-armed bandit and Stackelberg game[C]//Proceedings of the 2021 IEEE 37th International Conference on Data Engineering (ICDE). Piscataway: IEEE Press, 2021: 253–264.
- [11] Liu W B, Wang E, Yang B, et al. Data pricing via uncertainty quantification for sparse mobile CrowdSensing[J]. IEEE Transactions on Services Computing, 2026, 19(2): 1062–1074.
- [12] Zhang C, Xu Y, Kang R Y, et al. Pricing and trading of data options on data-as-a-service platforms[J]. IEEE Transactions on Services Computing, 2026, 19(2): 1478–1491.
- [13] 任洪润, 朱扬勇. 基于数据市场类型的数据定价模型研究[J]. 大数据, 2023, 9(4): 116–138.
- Ren H R, Zhu Y Y. Research on data pricing model based on data market type[J]. Big Data Research, 2023, 9(4): 116–138.
- [14] Parra-Arnau J. Optimized, direct sale of privacy in personal data marketplaces[J]. Information Sciences, 2018, 424: 354–384.
- [15] 陈思莹, 张丹, 丁小欧, 等. 基于数据质量的公平数据定价[J]. 大数据, 2024, 10(2): 54–67.
- Chen S Y, Zhang D, Ding X O, et al. Fair data pricing based on data quality[J]. Big Data Research, 2024, 10(2): 54–67.
- [16] 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,

2017: 1–10.

[17] Khan S A R, Piprani A Z, Yu Z. Supply chain analytics and post-pandemic performance: mediating role of triple-a supply chain strategies[J]. International Journal of Emerging Markets, 2023, 18(6): 1330–1354.

[18] Waqas U, Abd R A, Ismail N W, et al. Influence of supply chain risk management and its mediating role on supply chain performance: perspectives from an agri-fresh produce[J]. Annals of Operations Research, 2023, 324(1): 1399–1427.

[19] Mohamadi N, Niaki S T A, Taher M, et al. An application of deep reinforcement learning and vendor-managed inventory in perishable supply chain management[J]. Engineering Applications of Artificial Intelligence, 2024, 127: 107403.

[20] Keshavarz-Ghorbani F, Pasandideh S H R. Optimizing a two-level closed-loop supply chain under the vendor managed inventory contract and learning: Fibonacci, GA, IWO, MFO algorithms[J]. Neural Computing and Applications, 2021, 33(15): 9425–9450.

[21] Modares A, Farimani N M, Dehghanian F. A new vendor-managed inventory four-tier model based on reducing environmental impacts and optimal suppliers selection under uncertainty[J]. Journal of Industrial and Management Optimization, 2024, 20(1): 188–220.

[22] Salas-Navarro K, Florez W F, Cárdenas-Barrón L E. A vendor-managed inventory model for a three-layer supply chain considering exponential demand, imperfect system, and remanufacturing[J]. Annals of Operations Research, 2024, 332(1): 329–371.

[23] Duan X M, Zhao C C, He S B, et al. Distributed algorithms to compute walrasian equilibrium in mobile crowdsensing[J]. IEEE Transactions on Industrial Electronics, 2017, 64(5): 4048–4057.

作者简介



任洪润（1995–），女，复旦大学计算与智能创新学院博士生，主要研究方向为数据科学和数字经济，近期研究重点为数据定价、数据生产模型。



朱扬勇（1963–），男，博士，复旦大学教授，上海数据研究院有限公司学术副院长，上海市数据科学重点实验室创始主任，《大数据》期刊编委会副主任。在数据领域发表论文200多篇，持有高价值专利15件，出版学术著作17本，主导编制了多个地方政府数据发展规划。主要研究方向为数据科学与数字经济。

录用日期: 2026-07-07
通信作者: 任洪润, renhr20@fudan.edu.cn