软件学报 ISSN 1000-9825, CODEN RUXUEW
2026,37(8):3309−3336 [doi: 10.13328/j.cnki.jos.007658] [CSTR: 32375.14.jos.007658]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

机器学习模型市场的收益分配策略综述*

何佳妮[1,2], 黄科满[1,2], 刘金飞[3], 卢 卫[1,2], 范 举[1,2], 杜小勇[1,2]

[1](数据工程与知识工程教育部重点实验室 (中国人民大学), 北京 100872)

[2](中国人民大学 信息学院, 北京 100872)

[3](浙江大学 计算机科学与技术学院, 浙江 杭州 310058)

通信作者: 黄科满, E-mail: keman@ruc.edu.cn

摘 要: 数据收益的合理分配是构建可持续数据市场的核心命题之一. 相较于传统生产要素, 数据的价值后验性、信息不对称性、零成本复制、外部性等特性, 给其收益分配策略的设计带来了多维度的挑战. 聚焦数据市场的重要分支——机器学习模型市场 (以下简称模型市场), 系统梳理该领域收益分配策略的研究进展, 揭示其“从同质化走向差异化、从短期走向长期”的发展趋势. 具体而言, 首先形式化定义模型市场的收益分配问题, 厘清收益分配的主体、模式与目标. 在此基础上整理“同质化分配-差异化补偿”的分配依据: 在同质化贡献度量上, 归纳基于 Shapley 值等指标的数据贡献度评估体系; 在差异化补偿指标上, 解析数据成本、数据多样性等差异化指标的度量方法, 进而揭示综合两个维度的混合策略. 进一步地, 针对模型市场在长期尺度下的动态特征, 分析不同主体的策略性行为对收益分配的影响及其应对措施. 最后总结当前研究面临的主要挑战, 明晰从差异化补偿以及长期动态的视角探讨收益分配策略优化的未来研究方向.

关键词: 模型市场; 收益分配策略; 数据贡献度量; 差异化补偿; 主体策略性行为

中图法分类号: TP18

中文引用格式: 何佳妮, 黄科满, 刘金飞, 卢卫, 范举, 杜小勇. 机器学习模型市场的收益分配策略综述. 软件学报, 2026, 37(8): 3309–3336. http://www.jos.org.cn/1000-9825/7658.htm

英文引用格式: He JN, Huang KM, Liu JF, Lu W, Fan J, Du XY. Survey on Revenue Allocation Strategies for Machine Learning Model Markets. Ruan Jian Xue Bao/Journal of Software, 2026, 37(8): 3309–3336 (in Chinese). http://www.jos.org.cn/1000-9825/7658.htm

Survey on Revenue Allocation Strategies for Machine Learning Model Markets

HE Jia-Ni[1,2], HUANG Ke-Man[1,2], LIU Jin-Fei[3], LU Wei[1,2], FAN Ju[1,2], DU Xiao-Yong[1,2]

[1](Key Laboratory of Data Engineering and Knowledge Engineering (Renmin University of China), Ministry of Education, Beijing 100872, China)

[2](School of Information, Renmin University of China, Beijing 100872, China)

[3](College of Computer Science and Technology, Zhejiang University, Hangzhou 310058, China)

Abstract: The fair allocation of data revenue is one of the core issues in building sustainable data markets. Compared with traditional production factors, data exhibit several characteristics, such as ex-post value, information asymmetry, costless replication, and externalities, which pose multidimensional challenges for designing revenue allocation strategies. This study focuses on the machine learning model market, which is an important branch of the data market. It systematically reviews the research progress of revenue allocation strategies in this domain, revealing a development trend from homogeneity to differentiation and from short-term to long-term. Specifically, the revenue allocation problem in the machine learning model market is first formalized, and the participants, allocation modes, and objectives are clarified. On this basis, the allocation basis of “homogeneous allocation-differentiated compensation” is organized. In terms of

* 基金项目: 国家自然科学基金 (62441230, 62172425); SMP-IDATA 晨星青年基金 (SMP2023-iData-002)
收稿时间: 2025-05-22; 修改时间: 2025-12-19, 2026-02-02; 采用时间: 2026-02-26; jos 在线出版时间: 2026-06-01
CNKI 网络首发时间: 2026-06-02

homogeneous contribution measurement, data contribution evaluation methods based on indicators, such as the Shapley value, are summarized. In terms of differentiated compensation, the measurement methods of differentiated indicators such as data cost and data diversity are analyzed, and a hybrid strategy integrating both dimensions is revealed. Furthermore, regarding the dynamic characteristics of the model market over the long term, the impact of strategic behaviors of different participants on revenue allocation and the corresponding response measures are analyzed. Finally, the main challenges in current research are summarized, and future research directions for optimizing revenue allocation strategies are clarified from the perspectives of differentiated compensation and long-term dynamics.

Key words: model market; revenue allocation strategy; data contribution measurement; differentiated compensation; strategic behavior of stakeholders

数据作为继土地、劳动力、资本和技术之后的第 5 生产要素, 其市场化配置已成为国家战略布局的关键领域之一. 数据市场着眼于在数据供需双方之间建立涵盖数据集和数据衍生品的交易机制 (包括直接交易和中介服务模式)[1]. 在数据供需双方的交易过程当中, 供方需要完成数据的准备与提供, 需方则是根据自身需求选择合适的数据产品; 中介环节需要通过模型训练等技术手段, 构建数据产品, 并且评估各方对目标任务的贡献以进行收益分配. 这一流程需要综合考虑数据质量、数据成本及场景适配性等多种因素, 并依托算法工具实现数据价值的跨主体传递.

与传统生产要素市场相比, 数据市场呈现出显著的双重特性: 一方面, 其价值创造高度依赖多方主体的持续协同 (如数据准备、模型训练等环节的链式协作)[2]; 另一方面, 数据的价值后验、信息不对称等特性带来的双向道德风险、数据权力结构失衡等问题[3], 对构建可持续协作形成了根本性制约. 正是这种协同必要性与协作脆弱性之间的矛盾, 使得传统市场机制在数据要素配置中面临失灵风险. 因此, 迫切需要为数据市场构建合理的收益分配策略, 以促进多方主体持续协同、维系数据市场持续运作. 公平合理的收益分配策略不仅能保障参与者的价值回报, 更是破解“数据孤岛”、建立可信协作关系的关键保障[4].

值得关注的是, 在大模型与机器学习技术的驱动下, 数据市场正经历显著的分化进程. 其中, 以模型作为数据产品形态的模型市场在近年来快速崛起[5,6]: 多方数据资源通过数据融合和模型训练形成可交易的机器学习模型等数据模型产品, 用户通过使用模型产品满足具体的业务需求.

如图 1 所示, 模型市场主要由数据供方、模型平台 (中介)、模型开发方和模型需方组成[6−8]. 模型需方根据自身需求和预算, 向模型平台提交需求. 模型平台通过分析需方的需求可行性与支付能力, 发布训练任务, 并招募数据供方提供相应的数据支持. 数据供方根据任务要求, 选择将数据资源出售给中介以获取相应的收益 (由中介分配报酬). 模型开发方在中介的调度下, 整合多方的数据资源, 基于已有的模型或重新训练以得到多个模型. 模型需方从中选择满足其需求的模型, 并支付相应的费用 (构成收益来源). 模型平台将所得收益在数据供方和模型开发方之间进行分配, 最终形成了涵盖“需求提出、数据支持、模型开发、模型交付、费用支付和报酬分配”的交易流程. 显然, 数据供方和模型开发方能否在交易流程中获得公平合理的收益, 是决定它们是否持续参与市场交易, 从而影响模型市场能否持续运行的关键. 因此, 设计合理的收益分配策略是保障模型市场发展的核心问题之一.

针对模型市场的收益分配策略, 业界已经开展了系列探索, 但现有工作往往关注单一因素和静态环境, 存在“收益分配的衡量维度单一化”“收益分配的策略设计静态化”等局限. 例如, Shapley 值、Banzhaf 值等被广泛地用于数据估值和在数据供方间进行收益分配[9−11], 通过利用统一的效用函数来评估参与方的贡献, 进行“无差异化”的收益分配. 然而, 这种分配策略没有考虑参与个体的差异性, 导致分配结果难以真实全面地反映各方的实际贡献价值. 此外, 现有的大多数收益分配策略都是基于静态环境的假设, 难以适应模型市场在长期发展中的动态变化. 具体而言, 在行为维度上, 没有考虑各主体会因为各自的利益而采取策略性行为, 例如部分数据供方可能会将同一数据集重复提交多次以换取额外的收益等利己行为 (行为静态性); 在时间维度上, 仅针对单次交易设计分配策略, 未考虑多轮交易之间的关系以从长期视角上优化各方收益分配 (时间静态性). 针对上述局限性, 已有研究着眼于引入“差异化”[12]和“动态化”[13]的视角来改进现有收益分配策略, 希望解决模型市场中的数据贡献度量、数据价值评估、多方动态博弈均衡等关键挑战[14]. 然而, 现有的研究相对碎片化, 缺乏从多维差异化和长期动态性视角对模型市场中收益分配策略进行体系化分析, 因此难以有效整合不同机制, 以应用于复杂的现实环境中.

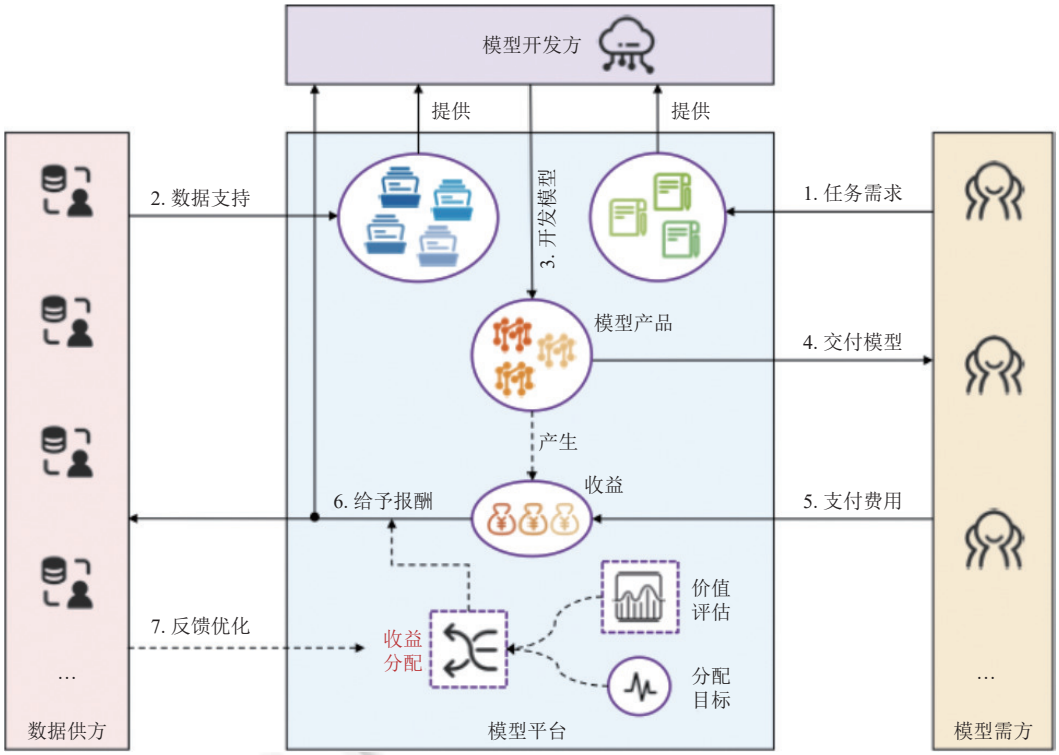


图 1 模型市场的收益分配

此外, 现有综述大多聚焦于广义数据市场的研究, 其研究视角与本文存在差异, 侧重的内容也不相同. Liang 等人[15]从数据定价、交易和保护的角度对数据生命周期及数据交易进行了全面的梳理. Zhang 等人[16]从数据搜索、数据产品化、数据交易、数据定价、收入分配以及隐私、安全和信任问题等方面对数据市场这一方向进行调研. 上述的研究都比较重视数据的整个生命周期, 尤其是对数据定价这一环节进行了深入的研究. 例如 Pei[7]从数据估值问题出发, 研究了各种动机的数据定价以及一系列定价模型的发展与基本原则. 刘枌等人[14]将大数据定价分为成本导向、市场导向、需求导向、利润导向和基于生命周期定价这 5 种定价. 蔡莉等人[17]从数据质量、信息熵、查询、博弈论、机器学习等角度对数据定价进行研究, 并对比了这几种方法的优劣势. 江东等人[18]在上述基础上, 从数据定价与交易的难点和相关准则出发, 将大数据在市场中的生命周期分成数据收集与集成、数据管理与分析、数据定价和数据交易这 4 个环节, 重点对数据定价思路和方法进行介绍, 分析了优势和短板.

从上述的分析可以看出, 现有对于数据市场的综述主要集中在数据定价的阶段, 对于后续的收益分配这一环节, 尤其是模型市场这一分支, 目前还没有综述进行详细的讨论. 同时对于收益分配的问题没有明确的形式化定义, 而是直接笼统描述, 比较分散. 因此, 本文明确将研究范围限定于模型市场, 将市场中的交易内容限定为模型及其信息, 并且重点关注交易后的收益分配环节. 进一步地, 本文还尝试从长期视角出发, 归纳总结模型市场中不同参与方的各类策略, 并在此基础上为收益分配的策略研究提出挑战和未来可能的发展方向.

在本文中, 如图 2 所示, 我们从以下角度进行梳理.

1) 通过对模型市场收益分配过程的分析, 从分配主体、分配模式和分配目标等方面给出收益分配策略的形式化定义, 体现模型市场收益分配的两阶段特征 (先进行主体间收益划分、后进行同类型主体内部分配细化).

2) 围绕现有研究焦点, 重点梳理现有在数据供方内部进行收益分配的研究, 包括融合数据贡献度、差异化指标的多维度收益分配因素, 揭示当前研究“从同质化分配到差异化补偿”的发展趋势.

3) 从数据供方、模型平台及开发方的主体角度, 概括在长期尺度下不同主体采取的策略性行为, 梳理现有的

应对技术和措施, 揭示当前研究“从短期静态策略到长期动态调整”的发展趋势.

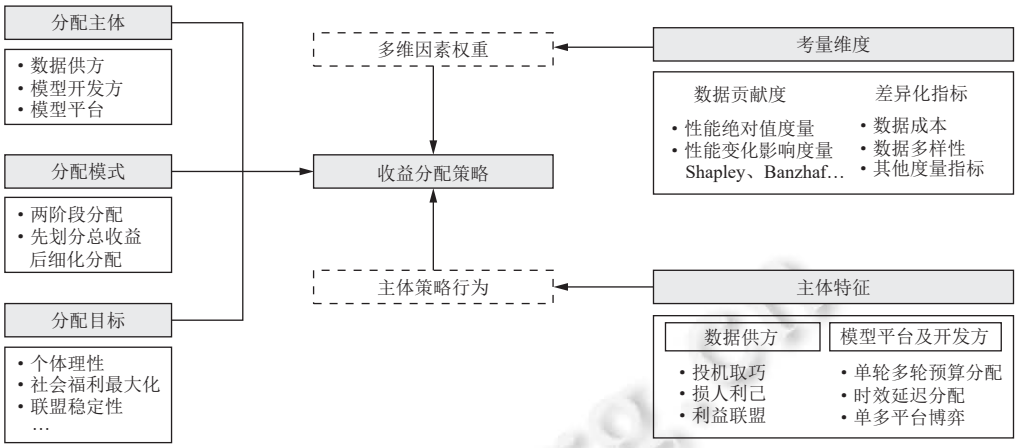


图 2 基于多维度因素和主体行为的收益分配策略框架

基于以上系统性梳理, 明确从综合多维差异化指标的混合策略、多主体长期策略性行为的动态策略两个角度优化收益分配策略的重要性, 为后续收益分配策略的设计与应用提供帮助和思路.

本文第 1 节梳理从数据市场到模型市场的演进历程以及现有收益分配的研究进展. 第 2 节介绍基于模型市场的多主体交互过程的分析, 构建包含分配主体、分配模式、分配目标在内的收益分配形式化框架. 第 3 节从数据贡献度和差异化指标这两个维度, 分析收益分配策略的关键因素, 揭示从同质化分配向差异化补偿的发展趋势. 第 4 节探讨在长期尺度下, 不同主体的策略性行为及相应的影响, 凸显从短期静态向长期动态的发展趋势. 第 5 节总结研究挑战并提出未来的研究方向.

1 研究现状

1.1 从数据市场到模型市场

数据市场这一概念的产生最早可追溯至 20 世纪 80 年代. Admati 等人[19]在 1986 年开创性地探讨了交易者从垄断卖方处购买信息的市场模式, 这些信息实质上是由个体代理所赋予或产生的数据. 2008 年成立的 Factual 和 2009 年成立的 Infochimps 公司被认为是早期数据市场的代表, 其允许企业和个人对数据集进行买卖. 其中, Factual 致力于出售与地理位置相关的数据集, Infochimps 则主要出售地理位置数据和社交网络数据. 2011 年, Balazinska 等人[20]提出的基于云的数据市场具有里程碑的意义, 其通过逻辑集中化的云平台重构了数据交易模式. 得益于云平台的集中化管理能力和灵活的服务架构, 数据交易从简单的数据买卖向多元化服务形态演进. 在此技术驱动下, 数据市场进入快速发展期, 其形态和交易模式持续迭代更新.

根据核心交易对象的不同, 目前的数据市场大致可以分为 3 大类: 第 1 类是将数据本身作为交易产品的数据市场, 卖方可以直接销售原始数据或提供用于特定任务的个性化数据集; 第 2 类是将数据相关服务作为交易产品的数据市场, 例如, 卖方可以出售订阅或查询服务, 买方定期访问更新的数据或按需查询多种数据资源[21−23]; 第 3 类是模型驱动的数据市场 (模型市场), 通过卖方提供的数据资源, 经过训练产生可用的机器学习模型, 并将其作为数据产品提供给买方[5,6]. 这类市场的典型代表是基于区块链的模型交换市场, 其通过智能合约实现自动化交易: 数据持有者发布训练任务, 开发者提交训练完成的模型即可获取相应的报酬[24].

模型市场作为数据要素流通和 AI 技术落地的关键基础设施, 正在深刻改变数据经济的价值分配模式. 不同于其他的数据市场, 模型市场通过交易机器学习模型等高阶数据衍生品, 实现了数据价值的深度挖掘和规模化应用[25]. 这类市场作为连接数据所有者、模型开发者和终端用户的关键枢纽, 不仅有效解决了传统数据共享中的“孤岛效应”问题, 更重要的是通过将原始离散的数据资源转化为可复用的 AI 能力, 显著提升了数据要素的流通效率.

研究表明, 理想的模型市场应当具有良好的可扩展性和灵活性, 这种特性不仅能够促进模型的开发和共享[8], 还能大幅度降低 AI 技术的应用门槛, 推动人工智能技术的民主化进程.

1.2 模型市场的收益分配研究

收益分配策略作为核心激励手段, 一直是学术界和产业界的研究重点. 近年来, 围绕收益分配的相关工作主要可以分为两类: 面向数据产品的理论机制研究和面向数字产品的场景化方案研究. 第 1 类研究聚焦于数据产品, 例如查询服务、聚合统计、机器学习模型等.

本文关注的模型市场的收益分配研究, 其并不直接提供具体场景的收益分配方案, 而是从理论层面探讨相关机制和影响因素. 基于研究动机可以细分为效率驱动型和公平驱动型. 对于效率驱动型的研究, 是通过定价机制优化收益分配效率, 主要基于以下理论框架: Stackelberg 博弈[26]、拍卖机制[27−29]和议价机制[30]. Stackelberg 博弈可用于建模数据卖家 (领导者) 和买家 (跟随者) 之间的交互关系. 通过优化设计卖家的定价策略来影响买家的购买决策, 从而间接调节收益分配. 拍卖机制研究包括 VCG (Vickrey-Clarke-Groves) 拍卖[27,28] (以买家的总报价最大化为目标) 和 Myerson 拍卖[29] (针对单参数环境的最优拍卖设计), 均通过定价规则来影响收益分配. 而议价机制则是通过买卖双方多次出价和还价来达到均衡价格.

虽然这些机制本身并不直接规定分配比例, 但通过影响最终交易价格间接影响了收益分配的效率. 对于公平驱动型的研究, 主要关注各种与收益分配公平性相关的影响因素[16]. 模型市场中经典的方法包括: 隐私补偿机制[6], 通过量化数据供方的隐私损失来确定补偿额度, 实现隐私成本与收益的正向关联; Shapley 值分配方案[31], 基于边际贡献的加权平均计算, 广泛应用于集中式机器学习和联邦学习等场景. 此外, 数据供方的声誉、资源等[32]因素也逐渐被纳入收益分配的讨论范畴, 以进一步优化收益分配的公平性.

相比之下, 第 2 类聚焦的数字产品 (如音乐、视频等) 的收益分配的研究已经形成较为成熟的实践方案. 以音乐流媒体平台的内容提供者收益分配为例, 目前已形成两种主流的方法: 按比例分配法和用户中心法[33]. 前者根据艺术家的总播放量按比例分配收益, 后者则将用户的订阅费按比例分配给该用户实际收听的艺术家. 这类方案具有指标明确、场景适配性强和可操作性高的特点. 进一步地, Jensen[34]系统性地归纳了 7 种音乐流收益分配依据. 除了上面的两种, 还包括: 基于收听时长的分配 (按照实际收听分钟数进行分配)、基于收听场景的分配 (区分用户选择主动收听与算法推荐被动收听)、基于历史行为的分配 (排除首次播放的影响)、基于艺术家成长模型的分配 (按照流行度阶梯分配) 以及基于最低播放量阈值的分配 (设置参与分配的最低标准). 这些数字产品的收益分配方案因其指标明确, 在实际应用中已经相对成熟. 反观数据产品特别是模型市场的收益分配研究, 由于涉及计算复杂度、隐私成本量化等特有挑战, 目前还没有形成统一的研究框架, 这也正是本文要重点探讨的方向.

总之, 当前关于收益分配的研究已经有了一定的基础, 学者们从不同的角度展开探索: 在理论层面, 基于博弈论、拍卖机制等构建定价模型, 并逐步拓展至隐私成本补偿、Shapley 值等公平性分配原则的探讨. 在实践层面, 聚焦场景化的分配方案设计; 这些研究都为理解模型市场的收益分配机制奠定了重要基础.

2 模型市场收益分配问题形式化定义

如前文图 1 所述, 在模型市场中, 数据供方、模型开发方、模型平台和模型需方围绕数据和数据产品形成交互. 从收益分配视角来看, 模型需方与模型平台就模型选择及定价的动态博弈影响参与收益分配的金额, 构成了收益分配策略的输入, 而不涉及金额在数据供方、模型开发方和模型平台之间如何进行分配. 本文中, 我们主要着眼于收益分配模式, 因此将模型需方与模型平台的交互视为给定, 从而将模型市场的收益分配问题细化为: 给定模型需方因购买模型而支付的费用, 应当如何将该收益合理地分配到各参与主体?

为此, 下文将结合模型市场的运作模式, 形式化定义收益分配过程中的分配主体、分配模式和分配目标, 为回答收益分配问题提供系统化框架.

2.1 分配主体

在研究模型市场的收益分配问题时, 首先需要明确参与分配的主体. 这是因为不同主体在数据价值创造过程

中的贡献方式、行为动机和利益诉求各不相同——数据供方提供数据支撑, 模型开发方负责模型的训练与优化, 模型平台则协调交易并制定分配规则. 只有清晰界定各主体的职能边界和互动关系, 才能准确评估其贡献, 识别潜在的策略行为, 并设计出公平高效的收益分配机制. 在一个典型的模型市场当中, 从其运作模式 (如图 1 所示) 可以看出, 参与分配的主体包括数据供方、模型开发方以及模型平台.

● 数据供方. 数据供方 i 提供数据 d_i, 其中 $D=\{d_i|1\leqslant i\leqslant n\}$ 表示所有数据.

● 模型开发方. 模型开发方 j 根据数据和需求进行模型开发, 最后产出模型产品 m_j. 需要注意的是, 部分模型产品可以在现有模型的基础上, 通过蒸馏、剪枝等操作并辅以数据进行微调得到. 因此, 可以从模型开发过程的角度对模型产品进行定义: $m_j=\{D_j,M_j\}$, $M_j=\{m_j|1\leqslant j\leqslant m\}$, 其中 $D_j\in D$ 代表模型 m_j 使用的数据, 而 $M_j\in M$ 则代表开发模型 m_j 时所使用的基准模型. 如果 D_j 为空, 意味着该模型不需要使用新的数据, 完全在现有基准模型的基础上通过变换操作产生. 如果 M_j 为空, 意味着该模型完全由新数据训练形成, 不依赖于现有的模型产品.

● 模型平台. 模型平台提供模型列表 M, 链接模型需方与模型开发方. 模型需方根据模型 m_j 的质量支付费用 R_j, 并成为模型的收益.

在本文中, 为了讨论的方便性, 我们假设每个数据仅由一个数据供方提供; 所有模型产品仅由一个模型开发方开发; 模型平台负责收益分配策略设计, 但本身不获得 R_j 的分配. 在实践过程中, 可能由多个模型开发方来共同开发模型产品, 一个数据供方可以提供多个数据. 模型平台可以通过分成的方式获得收益或者为平台提供补贴. 然而, 这些场景可以通过为不同数据设定相同供方、为不同模型产品设定不同模型开发方以及设定平台用于分配的费用系数等方式实现. 因此, 以上的定义具有一定的普适性, 能够体现实践中的灵活性和复杂性.

2.2 分配模式

围绕以上的相关分配主体, 收益分配模式包含两个阶段: 先主体间收益划分、后同类型主体内部分配细化. 其中主体间收益划分是指根据模型产品本身所需的开发资源、数据资源和依赖模型的相对贡献, 将总收益分配到不同类型的主体 (模型开发方、数据供方等); 同类型主体内部分配细化是指在每一类别的主体内, 进一步按一定比例分配收益. 这一模式既避免了单一阶段分配可能导致的粗粒度问题以及不同角色贡献缺乏可比性的挑战, 又能通过分阶段优化提升分配的透明度和合理性.

第 1 阶段: 主体间收益划分. 给定模型产品 $m_j=\{D_j,M_j\}$, 及其产生的收益 R_j, 选择收益分配模式 Π, 将收益 R_j 按照一定比例划分给模型产品本身 (代表的模型开发方) $V_j^{(\mathrm{dev})}$、所依赖的模型产品 $V_j^{(m)}$ 以及所使用的数据 (代表的数据供方) $V_j^{(d)}$:

$$\Pi\colon V_j^{(\mathrm{dev})}=\alpha\cdot R_j, V_j^{(m)}=\beta\cdot R_j, V_j^{(d)}=\gamma\cdot R_j,\ \alpha+\beta+\gamma=1 \tag{1}$$

如果 $V_j^{(d)}=R_j$, $\alpha=\beta=0$, $\gamma=1$, 意味着所有收益将全部分配给数据供方. 特别地, 如果 $V_j^{(d)}>R_j$, 即 $\gamma>1$, 意味着模型平台对数据供方进行转移支付, 提供超过模型产品本身收益的报酬, 以激励数据供方持续提供数据. 反之, 如果 $V_j^{(d)}=0$, 则意味着数据供方在这个过程中不能获得任何收益分配.

第 1 阶段的划分往往依赖于数据定价这个前期基础, 也就是确定 R_j 的过程中需要数据定价. 因为在数据交易时, 数据价值通常通过价格来进行体现. 现有的数据定价思路主要有基于任务的定价[22,23]、基于价值的定价[35–37]、基于经济学的定价[31,38]等. 这几类的划分不是简单的并行关系, 而是存在着相互交叉的内容. 其中经济学中的博弈论[39–43]作为重要理论, 近年来被广泛应用到数据定价中. 博弈论的核心在于探讨决策者之间存在策略互动时的行为选择以及这种行为决策的均衡问题.

在主体间的收益划分阶段, 确定分配比例 α、β、γ 是一个涉及多方利益平衡的复杂问题, 比例参数的设定并非静态, 而是受到市场结构、参与方关系、任务依赖等多重因素影响. 因此这一过程可以借鉴博弈论中的合作博弈框架, 把各方当作合作联盟, 将收益分配视为一个多方协商的动态过程, 形成基于合作博弈的均衡分配. 若在模型平台作为领导者、其他参与方作为追随者的场景下, 则可以通过基于 Stackelberg 博弈的层级决策来动态调整比例参数以增加市场活跃度. 若模型产品 m_j 同时服务于多个下游任务, 则可通过多任务贡献归因方法分解收益至各参与方. 另外, 通过纳什议价理论[44]来确保公平性, 当数据供方提供的数据都较为稀缺时, 可以适当提高 γ 以激

励供给; 当基准模型复用率比较高时, 适当提高β以激励共享. 除了利用博弈论框架来调整各主体间的收益比例, 还可以利用机器学习实现优化, 例如通过收集历史交易数据、模型复用情况以及市场供需关系等, 训练监督模型来预测最优的分配比例.

然而, 尽管第 1 阶段通过博弈论等方法可以实现宏观层面的合理分配, 但同类型主体内部的差异化贡献仍然需要通过第 2 阶段来进一步细化. 因为即使在同一个参与方类别内 (如多个数据供方), 各自的贡献程度也可能存在显著差异. 例如, 在多个数据供方中, 某些供方可能提供了更多高质量的训练数据, 而另一些供方的数据可能使用频率较低. 如果简单地采用统一的分配比例, 可能会导致贡献大的参与方得不到应有的回报, 长此以往将影响其继续参与的积极性. 因此, 第 2 阶段的细化分配不仅是对第 1 阶段的必要补充, 更是保障整个分配体系公平性和可持续性的关键环节.

第 2 阶段: 同类型主体内部分配细化. $V_j^{(d)}$将进一步分配给参与模型产品开发的相关数据供方:

$$\Pi_D\colon V_j^{(d_i)}=\pi_{j,i}\cdot V_j^{(d)},\sum_i\pi_{j,i}=1,\ 0\leqslant\pi_{j,i}\leqslant 1 \tag{2}$$

如果数据供方参与模型的开发, 按照该分配策略将从模型中获得直接收益; 另一方面, 数据供方如果参与基准模型的开发, 同样有可能从基准模型的收益当中获得相应的间接收益的分成. 如果基准模型还依赖于其他模型开发, 那么相应的数据供方也可能从中获得收益分成. 如果α、β、γ的值是不变的, 数据供方获得的总的收益可以表示为:

$$V_{j,i}^{(d)}=\pi_{j,i}\cdot\gamma\cdot R_j+\sum_k\left(\pi_{k,i}\cdot\gamma\cdot\beta^{l_k}\cdot R_k\right) \tag{3}$$

其中, 第 1 部分代表着数据供方从模型m_j获得的直接收益分成, 第 2 部分代表着数据供方通过参与基准模型获得的间接收益. l_k代表模型m_k变换到模型m_j的次数, l_k越大, 意味着离当前模型越远, 获得的收益分成就越少.

类似地, 模型开发方从模型m_j使用当中获得的收益可以定义为:

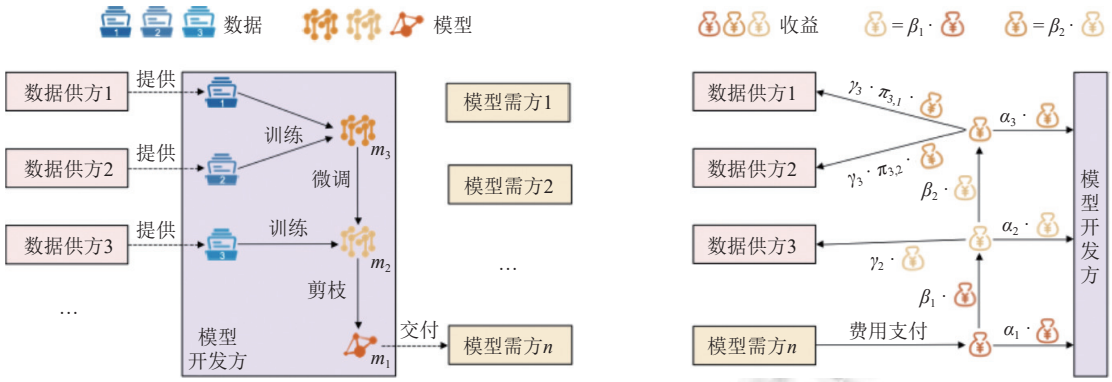
$$V_j^{(\mathrm{dev})}=\alpha\cdot R_j+\sum_k\left(\alpha\cdot\beta^{l_k}\cdot R_k\right) \tag{4}$$

其中, $\alpha\cdot R_j$代表模型开发方从模型m_j使用当中获得的直接收益, $\sum_k(\alpha\cdot\beta^{l_k}\cdot R_k)$代表模型开发方参与模型$m_j$开发链路中的多个相关基准模型开发所获得的间接收益. 显然, 如果$\beta=0$, 则意味着收益不穿透分配给所依赖的基准模型开发方及其对应的数据供方.

值得注意的是, α、β、γ的值可能随着不同的场景动态调整, 形成混合动态分配策略, 以更好地平衡各方利益分配. 如后文图 3 所示, 当一个模型m_1被交付给模型需方后, 其获得的收益将在模型开发方、数据供方及其所依赖的基准模型之间进行迭代分配. 首先, 由于交付的模型m_1是通过基准模型m_2直接剪枝得到的, 没有数据供方参与此过程, 因此m_1得到的收益只需分配给模型开发方和基准模型m_2, 收益分配的比例为α_1和β_1. 其次, 模型m_2是使用数据供方 3 提供的数据微调基准模型m_3得到的, 因此m_2从m_1得到的收益将被进一步分配到模型开发方、其依赖的基准模型m_3和数据供方 3, 比例为α_2、β_2、γ_2. 最后, 基准模型m_3是由数据供方 1 和 2 提供的数据训练得到的, 因此其从m_2得到的收益会按照比例α_3和γ_3分配给模型开发方和数据供方, 并且在数据供方 1 和 2 之间将按照一定的比例$(\pi_{3,1},\pi_{3,2})$进一步分配收益. 在这个例子中, 模型开发方除了从所交付模型本身可以获得比例收益以外, 还从该模型所依赖的基准模型中获取了部分比例收益. 对于模型开发方来说, 它是模型价值转换的核心主体, 它的贡献贯穿模型的全生命周期: 从需求解析、算法选择, 到数据预处理、模型训练, 再到模型部署适配、版本更新与维护等环节. 因此, 对于模型开发方的收益分配还需要结合过程指标与结果指标来进行综合评估.

2.3 分配目标

在现有研究中, 收益分配问题涉及管理学、经济学等多个领域. 在不同场景中, 收益分配策略的目标和约束条件也各不相同, 尚未形成统一的度量指标体系. 然而通过对现有文献的梳理发现, 当前研究中提出的分配目标主要围绕在个体理性和联盟稳定的基本约束条件下, 实现数据供方和模型开发方的整体社会福利最大化.



(a) 一个模型被交付给模型需方的示意流程 (b) 对应的收益分配流程

图 3 混合动态收益分配示意图

个体理性: 个体理性指的是每个数据供方在参与模型市场交易时, 其获得的收益至少不低于其不参与交易时的收益. 具体而言, 由于数据供方和模型开发方分别需要承担数据提供成本和模型训练成本, 因此在给定的上述分配策略下, 可以计算得到各个分配主体的效用:

$$u_i^{(d)} = \sum_j \left(v_{j,i}^{(d)} - c_{j,i}\right) \tag{5}$$

$$u^{(\mathrm{dev})} = \sum_j \left(v_j^{(\mathrm{dev})} - c_j\right) \tag{6}$$

其中, $c_{j,i}$ 代表数据供方 i 为开发模型 m_j 提供数据 d_i 时的成本, c_j 代表模型开发方训练 m_j 时需要投入的额外成本, 包括数据处理以及模型的蒸馏剪枝所需的资源等.

因此, 对于每个数据供方 $i \in n$, 假设其不参与市场时的效用为 $\tilde{u}_i^{(d)}$, 那么在模型市场中, 数据供方 i 的效用应满足:

$$u_i^{(d)} \geqslant \tilde{u}_i^{(d)} \tag{7}$$

这确保了在参与模型市场交易后, 其效用不会低于不参与时的效用, 从而激励其积极参与. 同理, 对于模型开发方也有:

$$u^{(\mathrm{dev})} \geqslant \tilde{u}^{(\mathrm{dev})} \tag{8}$$

其中, $\tilde{u}^{(\mathrm{dev})}$ 代表模型开发方不提供模型产品时的效用. 值得注意的是, $\tilde{u}_i^{(d)}$ 和 $\tilde{u}^{(\mathrm{dev})}$ 代表着机会成本, 因此不一定等于 0.

联盟稳定性: 联盟稳定性指的是在模型市场中, 没有任何一个数据供方子集能通过脱离当前联盟, 单独组建新联盟, 从而获得更高的收益. 如果所有供方都愿意留在现有联盟结构中, 说明这个联盟是稳定的, 称其为核心 (core):

$$\mathcal{V}(N) \geqslant \mathcal{V}(S) + \mathcal{V}(N \backslash S),\ \forall S \subseteq N \tag{9}$$

其中, $\mathcal{V}(N) = \sum_{j \in N}\left(v_{j,i}^{(d)} - c_{j,i}\right)$. 稳定性意味着联盟 N 的收益至少等于任何子联盟 S 与其余联盟 $N \backslash S$ 的收益之和, 从而保证没有参与联盟有动机分裂.

社会福利最大化: 社会福利最大化是考虑参与模型市场的所有参与分配方, 使其总体效用最大化.

$$\Pi^* = \max_{\Pi}\left\{\omega^{(\mathrm{dev})}\sum_j\left(v_j^{(\mathrm{dev})} - c_j\right) + \omega^{(d)}\sum_i\sum_j\left(v_{j,i}^{(d)} - c_{j,i}\right)\right\},\ \omega^{(\mathrm{dev})} + \omega^{(d)} = 1 \tag{10}$$

其中, $\omega^{(\mathrm{dev})}$ 和 $\omega^{(d)}$ 分别代表模型开发方和数据供方的效用权重, 第 1 部分代表模型开发方的效用, 第 2 部分代表所有数据供方的效用之和. 值得注意的是, 在不同的场景下, 平台可以给这两部分效用赋予不同的权重. 例如, $\omega^{(\mathrm{dev})} = 0$, $\omega^{(d)} = 1$, 意味着分配策略仅考虑数据供方的效用最大化, 以吸引数据供方提供数据. 反之, $\omega^{(\mathrm{dev})} = 1$, $\omega^{(d)} = 0$, 则意味着分配策略只以模型开发方的效用最大化为目标.

此外, 在模型市场的长期运行过程当中, 并非每时每刻都能够保证数据供方和模型开发方的效用大于 0, 因此也有部分工作研究在长期场景下, 放松对每一期个体理性的约束, 从遗憾值[42] (量化分配主体因成本与回报之间的暂时不匹配而产生的不满) 最小化的角度来讨论收益分配策略:

$$\Pi^* = \min_{\Pi}\left(\omega^{(\text{dev})}\sum_j\left(\max\left[c_j - v_j^{(\text{dev})},0\right]\right)+\omega^{(d)}\sum_i\sum_j\left(\max\left[c_{j,i} - v_{j,i}^{(d)},0\right]\right)\right) \tag{11}$$

因此, 从模型市场的运作模式出发, 我们可以得出, 收益分配策略问题可以形式化定义为一个参数优化问题. 即在给定成本设定 $c_{j,i}$、c_j、$\tilde{u}_i^{(d)}$、$\tilde{u}^{(\text{dev})}$, 如何通过设置 α、β、γ、$\omega^{(\text{dev})}$、$\omega^{(d)}$、$\pi_{j,i}$, 在满足个体理性和联盟稳定性的约束下实现社会福利最大化.

3 多维收益分配权重的量化与整合

基于以上的定义, 模型市场的收益分配是一个两阶段分配的问题, 包含两个核心步骤: 第 1 步是确定收益在模型开发方、依赖的基准模型以及数据供方之间进行分配的比例, 即定义公式 (1) 中的参数 α、β、γ. 第 2 步则是在各类主体中进行细化分配, 如多个数据供方之间的权重, 即定义公式 (2) 中的 $\pi_{j,i}$.

如表 1 所示, 第 1 阶段的收益划分往往跟模型市场的商业模式紧密相关. 现有研究和实践过程往往可以通过设置 α、β、γ 之间不同的参数组合, 从而形成不同的主体间收益划分策略, 以满足不同商业模式的需求. 例如, 只考虑数据供方之间的收益分配 ($\alpha=\beta=0$, $\gamma=1$), 而忽略模型开发方参与成本的数据成本法.

表 1 主体间收益划分模式

分配机制	参数设置	分配规则	特点说明	典型场景
数据成本法	$\alpha=\beta=0$, $\gamma=1$	$V_j^{(d)}=R_j$	完全按数据成本比例分配, 无模型开发方等分成	医疗影像模型市场
模型开发方独占	$\alpha=1$, $\beta=\gamma=0$	$V_j^{(m)}=R_j$	收益归开发方, 无数据供方分成	基于公开数据进行模型的微调
转移支付激励	$\gamma>1$	$V_j^{(d)}=\gamma\cdot R_j$	平台额外补贴数据供方, 需外部支持	新兴 AI 市场, 如高质量数据集的补贴
模型穿透分配	$\beta>0$	$V_j^{(m)}=\beta^l\cdot R_j$	支持多级模型依赖分配, l 为依赖层级	智能体市场, 如基于 Agent 的模型复用
收益固定比例法	$\alpha+\beta+\gamma=1$ α,β,γ 固定	$V_j^{(\text{dev})}=\alpha\cdot R_j$ $V_j^{(m)}=\beta\cdot R_j$ $V_j^{(d)}=\gamma\cdot R_j$	按预设的固定比例分配	数据空间协商确定, 一般在早期
混合动态分配	$\alpha,\beta,\gamma\in(0,1)$ α,β,γ 不固定	直接收益按比例分配, 间接收益穿透分配	平衡各方利益的动态分配方式	数据空间协商确定, 一般在后期

在实际应用中, 对于数据驱动型市场 ($\alpha=\beta=0$, $\gamma=1$): 适用于数据供方主导, 数据稀缺、模型开发门槛低的场景, 典型案例包括医疗影像模型市场, 数据供方 (医院) 提供高质量的标注数据, 模型开发方仅进行轻量微调. 对于模型开发方独占的场景 ($\alpha=1$, $\beta=\gamma=0$): 一般是基于公开数据对模型进行微调, 而无需考虑数据成本, 常见于开源模型微调. 对于平台补贴型市场 ($\gamma>1$): 适用于市场培育期, 平台通过补贴吸引数据供给, 典型场景为新兴 AI 市场的发展, 例如当前对高质量数据集构建的补贴. 对于模型复用型市场 ($\beta>0$): 适用于基准模型成熟的场景, 例如智能体市场, 通过 token 的方式给使用的不同基准模型或所调用的其他智能体进行分配. 最后的收益固定比例法和混合动态分配, 一般是在数据空间内协商确定的, 属于商定的范围. 早期因缺乏历史数据, 常依赖协商确定固定比例以快速启动合作, 后期基于运行中积累的数据和经验, 转向采用混合动态分配.

对于第 2 阶段的分配细化, 当前研究重点关注数据供方之间的分配策略, 以激发数据供方的积极性为目标. 即现有研究往往聚焦在给定 α、β、γ, 并设定 $\omega^{(\text{dev})}=0$, $\omega^{(d)}=1$ 的情况下, 如何有效设置权重 $\pi_{j,i}$ 以实现数据供方的社会福利最大化. 因此, 下文将在简述数据供方分配策略的基础上, 介绍分配权重 $\pi_{j,i}$ 的计算方法及其影响.

3.1 基于数据贡献度的数据供方收益分配策略

3.1.1 策略简述

在理想化假设条件下, 即当数据供方均为善意参与者时, 收益分配可基于严格的数据贡献度评估体系实现: 即按照各数据供方所提供数据对模型性能的贡献比例进行分配. 首先, 模型平台会根据模型需方的需求设定测试场景, 并定义对应的贡献函数 $C: D \to \mathbb{R}^+$, 用于衡量第 i 个数据供方提供的数据 d_i 在测试场景中对模型或任务目标的贡献程度, 以此作为收益分配的依据. 其中, $C(d_i)$ 表示数据供方 i 提供的数据 d_i 在贡献函数 C 下的贡献值. 所有数据供方提供的数据在贡献函数 C 上的总贡献表示为:

$$C_{\text{total}} = \sum_i C(d_i) \tag{12}$$

平台将数据供方提供的数据的贡献比例作为权重, 分配相应的收益:

$$\pi_{j,i} = \frac{C(d_i)}{C_{\text{total}}} \tag{13}$$

需要注意的是, 对于不同的场景, 权重的具体计算公式会发生一定的改变, 例如将误差作为贡献的负指标, 那么计算时应该考虑“误差越大则贡献越小”的原则. 衡量数据贡献的方法主要通过量化数据供方在合作过程中对模型性能、预测精度或者整体目标达成的影响. 以下是现有经典的衡量数据贡献的方法.

3.1.2 基于性能绝对值的数据贡献度量方法

该方法一般要求在数据估值阶段, 构造一个完整无偏的小规模测试集, 从而使得对于任一数据供方提供的数据, 将其用于模型训练后, 可以基于该模型在给定测试集上的性能来量化贡献. 其中, 常见的用作数据贡献度的度量方法包括分类任务中的分类准确率 (公式 (14)) 以及回归任务中的均方误差 (公式 (15)).

$$C(d_i) = \frac{1}{|y_{\text{out}}|} \sum\nolimits_{y_k \in y_{\text{out}}} \mathbb{I}[\hat{y}_k = y_k] \tag{14}$$

$$C(d_i) = -\frac{1}{|y_{\text{out}}|} \sum\nolimits_{y_k \in y_{\text{out}}} (\hat{y}_k - y_k)^2 \tag{15}$$

其中, y_{out} 代表测试集, $\hat{y}_k$ 代表利用数据训练模型后得到的预测结果, y_k 代表真实值, $\mathbb{I}[\cdot]$ 是示性函数, 用来表示预测结果是否与真实值一致. 针对分类任务, 数据 d_i 的贡献度与其训练所得模型的分类准确率呈正相关, 该指标虽然直观但存在两个固有的局限: (1) 对各类别错误一视同仁, 无法体现样本级预测差异; (2) 难以反映困难样本的分类难度, 不利于细化分类任务的具体误差评估. 为了赋予更难被正确分类的数据以更大的估值权重, 可以采用交叉熵来作为细粒度评估指标, 使数据价值度量更加细致[45]. 对于回归任务来说, 真实值和预测值的均方误差只是其中一种度量标准, 不同场景和任务下的误差度量标准不同, 例如还有均方根误差、加权误差 (如平均绝对误差通过时间权重加权) 等. 不论使用哪种度量标准, 当误差相反数作为数据的贡献度时, 误差越小, 贡献越大, 对应数据供方分配到的收益越大.

3.1.3 基于性能变化影响的数据贡献度量方法

此类方法的核心思想在于通过量化数据提供方所贡献数据对模型性能的边际提升效应来评估其价值, 其通过评估数据点在不同组合场景下对模型预测能力的增益幅度, 实现公平的价值量化. 其代表性方法很多都来源于合作博弈论的相关概念, 如 Shapley 值、Banzhaf 值以及 Owen 值.

● 基于 Shapley 值的方法. 当个体通过合作的方式创造总收益时, Shapley 值通过对所有可能的个体排列顺序下的边际贡献求平均, 以衡量每一个体对整体收益的边际贡献. Shapley 值满足零贡献 (无贡献者不分配收益)、合理性 (个体收益不低于单干)、对称性 (贡献相同者收益相同)、可加性 (线性) 等理论性质. 在模型市场中, 对于模型产品 $m_j = \{D_j, M_j\}$, 针对 m_j 使用的任一供方的数据 $d_i \in D_j$, 其 Shapley 值 $C(d_i)$ 的计算公式为:

$$C(d_i) = \sum_{D_0 \subseteq D_j \setminus \{d_i\}} \frac{|D_0|!\left(\left|D_j\right| - |D_0| - 1\right)!}{\left|D_j\right|!} [v(D_0 \cup \{d_i\}) - v(\{d_i\})] \tag{16}$$

其中, $v(D_0)$ 表示 D_j 中的一个数据子集 D_0 在合作中创造的收益, $v(D_0 \cup \{d_i\})$ 表示在子集 D_0 中加入数据 d_i 后所能

创造的收益, 因此两者的差值即可视为加入 d_i 后对数据子集 D_0 带来的额外收益. $\frac{|D_j|!\left(|D_j|-|D_0|-1\right)!}{|D_j|!}$ 是加权因子, 表示 d_i 恰好在子集 D_0 之后加入 D_j 的概率.

Shapley 值作为模型市场中衡量数据贡献度的常用方法, 具有两个显著优势: (1) 在一定程度上支持动态定价, 当一个新数据加入时, 可实时计算其对现有模型的边际提升, 调整收益分配[46]; (2) 具有理论完备性、可解释性, 数据供方能清晰地理解自身数据的实际价值[47].

然而 Shapley 值计算复杂度高, 特别是当数据供方数量较多时, 计算开销大. 因此, 当以 Shapley 值作为衡量数据贡献度的方法时, 现有大部分研究的目标转移到了其计算效率问题上, 以减少计算消耗. 例如, van Campen 等人[48]通过分层蒙特卡洛采样的方法加速 Shapley 值的计算, 该方法基于分层采样的方法, 从原始的 $|D_0|!$ 个排列中采样出均匀而具有代表性的少量排列, 用于估算 Shapley 值; Wang 等人[49]对 K 近邻分类器的场景中进行数据估值时, 利用阈值和 K 近邻的局部性质, 通过距离对数据点进行排序, 并递归地计算每个数据点的 Shapley 值, 从而将其计算从指数级时间复杂度降低为多项式级, 有效地提升了其在大规模数据估值任务中的可用性. 表 2 对比了近年来由 Shapley 值衍生而来的各种方法. 另外, 结合 Shapley 的内容, 王勇等人[50]在联邦学习场景下, 对参与方贡献评估的工作也进行了归纳和对比.

表 2 Shapley 值衍生方法对比

方法分类	基本技术原则	方法特点	代表方法	参考文献
精确计算	精确计算每个参与者的公平贡献	排列组合数量爆炸式增长, 计算量大	传统 Shapley	[51]
近似计算	减少特征组合, 降低计算成本	结构化排列采样	SMC-Shapley	[48]
		分组检验计算	GTB-Shapley	[52]
		截断训练剪枝	TMC-Shapley	[53]
		阈值K近邻剪枝	Threshold KNN-Shapley	[49]
		引导阶段梯度, 加速收敛	GTG-Shapley	[54]
		样本抽样方法	CCN (Neyman approach)、CCB (empirical Bernstein bound)	[55]
		基于最小二乘回归的通用估计	GELS-Shapley	[56]
		与边际贡献概念分离的近似估计	SVARM	[57]
		利用Shapley值之间的差异矩阵	Diff、S-Diff	[58]
动态计算	复用原有的Shapley值	适用数据动态增加或减少的场景	Dynamic Shapley	[46]
分布建模	分布参数化	引入Beta分布和概率分布建模参与方分布建模	Beta-Shapley D-Shapley	[59] [60]

除了对 Shapley 值的计算方式进行改进, 学术界还从多个维度拓展了数据贡献度估计的研究. 首先, 传统及近似 Shapley 值计算都是基于数据源具有完整特征或样本空间的假设, 而针对仅包含部分特征和样本的碎片化数据场景, Liu 等人[61]创新性地提出了二维 Shapley (2D-Shapley) 理论框架, 有效量化了碎片化数据片段对于特定机器学习的任务的价值. 其次, 传统方法因忽略模型结构的差异, 导致同一数据在不同模型中的价值分配可能出现偏差. Wang 等人[62]提出了一种与训练过程耦合的模型相关方法 (如动态更新效用函数), 通过实时跟踪 Shapley 值, 减少了模型差异无关假设带来的归因偏差, 使估值结果更加贴合特定模型的行为.

除了归纳 Shapley 值的评估体系, 我们进一步分析其在方法特点之外的实际应用局限性, 并建立与机器学习模型市场真实场景的映射关系. 在实际应用中, Shapley 值存在场景适配局限、隐私与数据安全冲突、收益分配粒度矛盾等不足. 具体如下.

● 对于场景适配问题: Shapley 值的“全排列边际贡献计算”假设数据供方贡献独立且可完全枚举, 但若模型市场中存在数据互补性强 (如多源异构数据融合) 或数据依赖关系 (如时序数据前后关联) 的场景, 其边际贡献评估易出现偏差, 无法准确反映数据协同价值.

● 对于隐私问题: 计算 Shapley 值需获取各数据供方的原始数据或详细特征信息, 在模型市场强调数据隐私保

护的情况下, 直接计算可能违反数据脱敏要求, 增加隐私泄露风险.

● 对于收益分配粒度问题: Shapley 值倾向于“精细化公平分配”, 但模型市场中部分中小数据供方可能因数据规模小、贡献分散, 其 Shapley 值计算结果极低, 导致收益分配不足, 难以激励其持续参与.

在集中式模型交易场景中, 数据供方较少、数据隐私要求较低, Shapley 值的“公平性优势”可充分发挥, 适用于高价值数据的收益分配, 例如区块链驱动的小众高质量数据模型交易平台. 在联邦学习型模型市场场景中, 分布式协作、数据不出本地, Shapley 值的“隐私冲突”和“计算复杂度”问题凸显, 需结合近似计算方法与隐私计算技术, 适配联邦学习场景下的多主体收益分配, 例如跨机构医疗 AI 模型联合训练后的收益划分. 在动态迭代的场景中, 模型需持续迭代优化, Shapley 值可与动态计算方法结合, 适配数据供方动态更新、模型版本迭代的场景, 确保多轮交易中收益分配的连贯性, 例如 AI 模型 API 服务平台的按次调用收益回溯分配.

此外, 学者们还探索了替代 Shapley 值的新范式: Kwon 等人[63] 提出了 Data-OOB, 一种基于袋外 (out-of-bag) 估计的估值方法, 该方法通过复用已训练的弱学习器实现高效计算, 即利用集成模型, 避免了传统 Shapley 值方法需要多次训练模型的复杂性. Just 等人[64] 提出了一种基于类间 Wasserstein 距离的数据估值框架, 该方法在计算距离时从现有的优化求解器中直接解析数据价值, 而不用额外计算. Wu 等人[65] 提出了 Davina 算法, 通过函数映射替代训练过程, 实现了无训练环境下的高效估值.

● 基于 Banzhaf 值的方法. 此方法通过计算所有可能的数据子集中某一数据供方的加入所带来的边际贡献的平均值:

$$C(d_i)=\frac{1}{2^{|D_j|-1}}\sum_{D_0\subseteq D_j\setminus\{d_i\}}[v(D_0\cup\{d_i\})-v(\{d_i\})] \tag{17}$$

可以看到, 与 Shapley 值不同, Banzhaf 值对所有子集情况赋予相同的权重, 而不是根据子集大小进行加权. 有研究表明, 这种方法在处理模型训练中的随机性和噪声时表现出更高的鲁棒性. 例如, Wang 等人[10] 提出的 Data Banzhaf 框架, 展示了在多种机器学习任务中, Banzhaf 值在数据质量评估和噪声检测方面优于其他方法. 而针对更复杂的噪声结构, 研究者又提出了加权 Banzhaf 值, 通过对不同子集大小赋予非均匀权重, 从而进一步提升估计精度与抗噪性能[66].

● 基于 Owen 值的方法[67]. Owen 值是 Shapley 值的扩展, 适用于存在先验联盟结构的合作博弈, 因此这种方法特别适用于数据供方之间存在合作的场景. 这一方法考虑数据供方之间的联盟关系, 首先在联盟之间分配总收益, 然后再在联盟内部通过例如基于 Shapley 值的方法进行分配, 从而更好地确保跨机构合作的公平性.

上述方法都是从添加数据获得的收益这一角度来衡量数据价值. 相反地, 通过分析移除部分数据对模型性能带来的影响, 也可以度量数据价值. 典型方法包括留一法和影响函数法.

● 留一法 (leave-one-out)[68]. 在机器学习中, 通过移除单个样本来观察模型性能的变化, 可以评估该样本对整体模型的重要性或价值. 假设有一个大小为 N 的数据集 $\{(x_i,y_i)\}_{i=1}^{N}$, 留一法的步骤是: (1) 对于每个 $i\in\{1,2,\ldots,N\}$, 使用去掉第 i 个数据 (x_i,y_i) 的剩余 $N-1$ 个数据来训练模型, 得到模型 f_{-i}; (2) 用模型 f_{-i} 在第 i 个数据点 (x_i,y_i) 上进行预测, 得到预测值 $\hat{y}_i$, 并计算第 i 个数据点的误差 $e_i=L(y_i,\hat{y}_i)$, 其中 $L(\cdot,\cdot)$ 是误差函数. 那么留一法对应的贡献函数可定义为:

$$C(d_i)=\Delta V_i=L(y_i,\hat{y}_i)-\left(y_i,\hat{y}_i^{\text{full}}\right) \tag{18}$$

其中, $\hat{y}_i^{\text{full}}$ 为完整数据集训练的模型在样本 (x_i,y_i) 上的预测值. 当 $\Delta V_i>0$, 为正向贡献, 表明移除样本 d_i 导致其在原位置的预测误差增大, 说明该样本对模型的局部拟合能力具有增强作用; 若 $\Delta V_i<0$, 为负向贡献, 说明该样本 d_i 可能为噪声或冲突样本. 假设平台仅对正向贡献进行收益分配, 可通过负值截断来确保权重非负, 即总贡献表示为:

$$C_{\text{total}}=\sum_i\max(0,\Delta V_i) \tag{19}$$

收益分配比例为:

$$\pi_{j,i}=\frac{\max(0,\Delta V_i)}{C_{\text{total}}} \tag{20}$$

通过留一法, 可以得到每个样本在模型中的预测误差或损失. 这个误差或损失可以看作是样本对模型性能的影响程度, 从而反映了样本的价值. 具体来说, 如果一个样本在移除后导致模型性能显著下降 (即预测误差增大), 那么这个样本就被认为是具有较高价值的, 从而可以被分配到一个较高的价值权重. 但是该方法依赖于样本自身误差估计, 可能高估噪声点的贡献.

● 影响函数 (influence function) 法[69]. 影响函数是统计学习中广泛应用的重要工具, 它通过量化训练样本对模型参数的边际影响来评估样本重要性. 该方法的核心优势是能够在不重新训练模型的情况下, 通过计算 Hessian 矩阵的逆与损失函数梯度的乘积, 估计每个训练样本对模型预测的贡献程度. 与需要重新定义数据集或损失函数的传统方法相比, 影响函数法的计算效率优势使其特别适用于大规模数据集和复杂模型的解释性分析. 具体地, 影响函数 $IF(x_i)$ 衡量了移除训练样本 x_i 对模型参数 θ 的影响. 根据一阶近似, 影响函数可以表示为:

$$IF(x_i) = -H^{-1}\nabla_\theta L(\theta, x_i) \tag{21}$$

其中, $\nabla_\theta L(\theta, x_i)$ 是损失函数在样本 x_i 上对参数 θ 的梯度. 由于影响函数的结果可能是一个向量 (即参数变化的方向), 因此可以用 L2 范数进行转换得到一个标量度量来表示相应的贡献函数值:

$$C(d_i) = \|IF(x_i)\|_2 \tag{22}$$

那么对应的收益分配比例为:

$$\pi_{j,i} = \frac{-H^{-1}\nabla_\theta L(\theta, x_i)}{\sum_k -H^{-1}\nabla_\theta L(\theta, x_k)} \tag{23}$$

在机器学习中, 影响函数近似留一法. 或者更一般地, 用两个模型的平均性能差异来进行近似: 一个是在训练过程中包含感兴趣的数据点的模型性能, 另一个是不包含该数据点的模型性能.

小结: 在基于数据贡献度的数据供方收益分配策略中, 基于性能绝对值的数据贡献度量方法能够反映具体实际的应用场景, 但是它依赖于完备的测试集, 并且容易出现由于多个数据供方提供同质数据而导致冗余的后果. 基于性能变化影响的数据贡献度量方法, 其中相关合作博弈论的一系列方法满足零贡献等理论优势, 但是在实际应用中往往受到大规模数据的限制, 其复杂度和计算成本都比较高. 另外, 留一法可以直接量化数据点对模型预测的局部影响, 但是计算成本昂贵, 难以扩展到大数据集. 同时, 对于拥有相似的高价值数据的供方来说, 移除其中相似的数据, 可能对模型影响不大, 从而导致这些数据供方的数据被错误地评估为低价值, 造成不公平现象; 而对于稀缺性数据, 留一法可以较好地识别出它的高价值. 影响函数法基于一阶梯度与 Hessian 矩阵逆, 近似估计样本对模型参数的影响强度, 从而间接量化其贡献. 虽然它避免了重复训练, 但由于利用近似算法, 其估计结果可能存在理论偏差.

3.2 基于差异化指标的数据供方收益分配策略

3.2.1 策略简述

除了基于数据贡献度的分配策略外, 另一类策略是从数据自身的差异化指标出发, 结合数据成本、数据多样性以及其他统计指标进行收益分配. 基于数据贡献度表现的分配策略依赖于特定模型的训练和评估, 其适用范围受到模型类型、测试集选择和计算环境等多方面限制. 而基于差异化指标的分配策略, 侧重于数据本身, 无需依赖具体模型或测试集.

3.2.2 常见的数据差异化指标

(1) 数据成本. 数据成本影响到数据供方的盈利能力、市场竞争力等方面, 主要涵盖时间成本、计算成本、管理成本和隐私成本. 这些成本不仅反映了数据供方提供数据所付出的直接和间接代价, 也为分配策略的公平性提供了重要依据.

● 时间成本. 它反映了数据供方在数据处理和共享过程中消耗的时间资源, 这个花销往往处于交易前期. 对于原始的数据, 数据供方往往需要进行数据整理[70]. 有研究表明, 很多大数据分析任务, 80% 以上的工作都是花费在数据整理上[71]. 例如, 对非结构化数据进行结构化处理, 对数据中的缺失、不一致、异常进行数据清洗等. 另外, 在分布式系统或联邦学习中, 数据供方还需要花费时间上传数据 (参数)、下载更新后的模型并完成同步操作.

● 计算成本. 它指数据供方在数据处理和协作过程中消耗的计算资源, 包括硬件资源、能耗和存储需求. 特别地, 在联邦学习中数据供方需要进行本地模型训练, 这对硬件能力有限的边缘设备或中小企业构成显著负担. 同时, 在长期存储的场景下, 大规模数据集的存储需求也会显著增加硬件设备的压力[72].

● 管理成本. 它是数据供方为满足数据合规和管理要求所付出的资源投入. 例如, 合规性成本涉及访问控制等操作, 以满足 GDPR[73,74]等法规要求. 数据质量认证则需通过第三方审计或认证, 以证明数据的可靠性和完整性[75]. 管理成本体现在对数据的生命周期管理, 如更新、存档和删除等.

● 隐私成本. 它主要涉及数据供方为保护数据隐私而投入的技术及潜在隐私泄露的风险代价. 在这过程中, 数据供方还可能因隐私泄露的潜在风险而承担心理成本, 特别是敏感数据 (例如医疗记录、财务信息) 的共享场景[76], 这种风险感知可能直接影响供方的参与意愿.

在上述成本中, 隐私成本是目前最为广泛研究的, 而不同的隐私保护机制会带来不同的隐私成本. 差分隐私能够通过数学公式明确定义隐私泄露的边界, 并将隐私成本转化为可计算的指标, 因此下面以差分隐私为例进行介绍.

差分隐私 (differential privacy, DP)[77]是一种用于保护数据隐私的数学框架. 它通过引入随机性确保单个数据样本的加入或移除对算法输出的影响是可忽略的, 从而使攻击者无法通过输出推断出个体信息. 隐私损失的随机性源于算法中注入的噪声, 常见的噪声机制包括: 拉普拉斯机制[78]、高斯机制[79]、指数机制[80]等, 这使得即便知道数据 D 或 D', 也很难区分出真实的输入数据集. 一个经典的差分隐私方法是 ϵ-DP[81]. 随机机制 A 满足 ϵ-DP, 当且仅当对于任意两个仅相差一个数据点的数据集 D 和 D', 以及任何可能的输出 o, 有:

$$\Pr[A(D)=o] \leqslant \mathrm{e}^{\epsilon} \cdot \Pr[A(D')=o] \tag{24}$$

其中, ϵ 为隐私损失参数, 用于衡量隐私保护的强度. 当 ϵ 越接近于 0 时, 表明随机机制 A 使得数据集 D 和 D' 上输出相同结果的概率分布也越相近, 随机机制 A 所泄露的信息也就越少, 保护程度就越高. 但传统的 ϵ-DP 在多次顺序查询的场景中会产生隐私损失累积的问题. 目前有研究提出 ρ-zCDP[82], 使用集中差分隐私概念, 通过隐私损失随机变量 Y 的分布 (如高斯分布) 来精确量化隐私泄露, 避免了单一参数 ϵ 的粗糙估计.

对于 ρ-zCDP 来说, Ghosh 等人[83]认为隐私成本主要包括隐私预算 ρ_n 和隐私偏好 a_n. 隐私预算 ρ_n 是数据供方愿意承受的最大隐私损失, 它由数据供方根据自身隐私需求设定, 是隐私保护的量化指标. 隐私偏好 a_n 体现了数据供方对隐私泄露的敏感程度. 因此对于数据供方的隐私成本可以量化为:

$$c_n = \alpha_n \cdot c_p(\rho_n) \tag{25}$$

其中, $c_p(\cdot)$ 是隐私成本函数, 描述在特定隐私预算下的隐私代价. 当隐私成本函数是线性时[84], 如 $c_p(\rho_n)=\rho_n$, 此时 $c_n=\alpha_n \cdot \rho_n$, 隐私成本和隐私预算成正比; 当隐私成本函数为非线性时, 如 $c_p(\rho_n)=\log(\rho_n+1)$, 隐私成本的增长随隐私预算的增长呈现递减趋势. 对于模型平台的分配策略来说, 隐私预算越高的数据供方, 其参与的隐私成本也越高, 也需要考虑给予其更多的收益补偿.

(2) 数据多样性. 数据多样性是衡量数据价值的重要指标之一. 一般来说, 数据多样性反映了数据在特征空间的覆盖范围和广度[85]. 多样性高的数据能够更全面地反映现实世界的复杂性, 具有更好的代表性, 同时可以减少偏见和提高公平性[86,87]. 在没有恶意干扰的理想情况下, 数据分布的多样性能够反映数据的价值, 即分布多样性越高, 数据价值越大. 数据多样性的常见衡量方法如下.

● 格拉姆行列式法. 假设数据集包含 k 个特征, 每一条数据可以表示为 k 维欧氏空间中的一个向量. 数据向量在欧氏空间中的分布越分散, 其多样性越高. 格拉姆行列式 (Gram determinant) 是一种用于量化欧氏空间内数据分布分散程度的有效工具:

$$Gram(D)=\det\left(X_D^{\mathrm{T}}X_D\right) \tag{26}$$

其中, X_D 是一个大小为 $|D|\times k$ 的数据矩阵, $|D|$ 是样本数, k 是特征数. 格拉姆行列式的值可以量化这些数据向量形成的平行多面体的体积, 体积越大表明数据分布越发散, 多样性越高, 从而数据价值越大[88].

● 网格划分法与衰减系数调整法. 格拉姆行列式法在实际应用中存在一个重要缺陷: 数据供方可能通过复制

自身数据来人为增加数据分布的体积, 从而不公平地提高其多样性价值. 为应对这一问题, 可以采用网格划分法将 k 维欧氏空间划分为若干子空间, 用每个子空间的数据样本均值代表该子空间, 再基于这些均值计算数据分布的多样性. 此外, 引入衰减系数作为调节参数, 当同一子空间内的数据量增加时, 衰减系数同步增大, 以抑制数据复制带来的虚假贡献[89]. 近期的一项研究通过数据矩阵的体积形式化多样性, 提出一种稳健且无需验证的基于体积的数据评估方法来应对“复制相同的数据点增加数据的多样性”的漏洞[90].

● 基于熵的多样性度量. 该方法用于评估数据分布的离散程度和均匀性, 其核心是通过香农熵 (Shannon entropy)[91]来量化信息的不确定性和多样性:

$$H(X) = -\sum\nolimits_{\mathrm{i}} P(x_i)\log P(x_i) \tag{27}$$

其中, 熵值与数据多样性呈正相关, 熵值越高表明数据分布越均匀、信息越分散, 从而数据的多样性越高. 这种方法适用于多种数据场景, 例如在文本分析中可用于评估词汇分布的丰富程度, 在联邦学习中衡量参与方数据对全局模型的潜在贡献. 熵度量的优势在于对数据简单复制的鲁棒性和跨领域适用性[92]. 然而, 基于熵的多样性度量在实际应用中也存在一些缺陷. 首先, 其对概率分布估计的准确性依赖较高, 尤其是在小样本场景下容易受到样本不足或离散化处理的影响[89]. 其次, 熵值无法反映数据点或类别的重要性, 可能忽略语义信息和类别的实际贡献, 从而低估某些稀有的高价值数据的作用. 另外, 熵对噪声数据也较为敏感. 尽管熵对简单数据复制具有一定鲁棒性, 但在局部复制或伪造类别的情况下, 仍可能被操控, 导致分配结果的不公平.

● 任务相关的多样性度量. 前述的多样性指标大多从纯统计或者分布角度出发, 然而数据的多样性价值很多时候依赖于下游任务需求. 例如, 在分类任务中, 可引入类别平衡度或类别覆盖度作为多样性度量, 即数据集中各类别样本分布的均匀程度或覆盖的类别数量, 这对于多类别不平衡场景下的模型泛化至关重要. 又例如在回归或者结构化预测任务中, 特征空间的覆盖密度或预测不确定性区域的样本分布可作为多样性指标, 反映数据在关键决策区域的代表性.

(3) 其他度量指标. 由于模型的准确率与训练样本的大小有关[93–95], 许多研究采用数据规模作为分配依据, 即数据量越大, 对应的数据价值越高, 模型平台分配给数据供方的收益也相应越多. 除了数据规模, 数据质量是另一个度量指标. 目前广泛采用的数据质量框架包括 TDQM (total data quality management)、DQAF (data quality assessment framework) 等, 里面包括了准确性、完整性、一致性、及时性、可靠性等数据质量维度[96]. 以金融数据为例, 实时更新的股票交易数据相比滞后一天的数据具有显著的时效性 (及时性) 优势, 因而在价值评估中可获得更高权重. 此外, 数据来源的可靠性和权属清晰度 (数据确权) 也会影响价值评估, 明确的数据归属和可靠的数据来源能够有效降低数据使用风险, 提升数据的可信度与附加价值, 因此相关数据供方往往可以获得额外的价值补偿.

小结: 在基于差异化的分配策略中, 数据成本中的时间、计算、管理成本指标简单直观, 但是相较于隐私成本来说, 目前缺少深入的理论研究, 难以进行明确的计算. 数据多样性属于一种直观且易于计算量化的数据统计类指标, 但是容易被恶意行为攻击. 对于数据规模、数据质量等其他度量指标, 它们同样具有简单直观, 不依赖于具体的模型和任务的优点, 但也存在不能完全反映数据的有效性、存在估计偏差等问题.

3.3 基于数据贡献度和差异化指标的混合分配策略

如表 3 所示, 基于数据贡献度和基于差异化指标的数据供方分配策略各有其独特的优势和局限性. 前者通过模型在具体任务中的表现评估数据对模型性能的贡献, 具有较高的透明性、可解释性和长期激励效果. 然而, 高昂的计算成本和实施复杂性限制了其在数据市场中的广泛应用. 相比之下, 后者通过数据成本、数据多样性等差异性指标进行评估, 显著降低了计算成本和实施难度, 同时在使用领域和数据类型方面更具灵活性, 不仅能适用于未建模阶段 (即未构建目标模型框架), 也更容易纳入新指标并进行扩展. 然而, 其评估结果可能不够全面, 且存在数据供方提供虚假信息的风险, 长期激励效果较弱.

为此, 近年来, 部分研究开始考虑融合两种分配方法的混合策略. 一个直观案例是欧洲主要足球联赛的奖金分配, 它们普遍采用混合分配机制[97], 即将总收入划分为基础部分 (均等分配)、绩效部分 (按竞技表现比例分配) 和衍生部分 (基于社会绩效指标分配, 如观众指标). 在数据市场领域, Dawande 等人[12]以数据合作社为背景, 在分配

收益时将隐私成本 (上文提到的差异性指标中的一种) 纳入考量, 即成员主体在向数据合作社提供数据时所需承担的隐私风险与损失. 并将隐私成本进一步细分为同质隐私成本与异质隐私成本两类: 同质隐私成本意味着所有成员承担的隐私成本相同, 不存在差异; 而异质隐私成本则表明不同成员因隐私敏感度等不同而面临不同的隐私成本, 有的成员面临高隐私成本, 有的则面临低隐私成本. 基于此, Dawande 等人[12]设计了 3 种混合分配策略, 其对比情况如表 4 所示.

表 3　基于数据贡献度与基于差异化指标的分配策略差异

维度	基于数据贡献度的分配策略	基于差异化指标的分配策略
评估原理	依据数据对模型性能的实际提升效果	自定义的数据特征差异化指标
主要优势	结果可验证, 直接反映数据价值 促进高质量数据供给 分配客观性强	实验简便灵活 适合多种场景 计算效率高
主要局限	计算复杂度高 存在过拟合风险 实施门槛较高	评估标准主观性较强 数据真实性难验证 评估维度较单一
公平性	较高 (直接关联数据贡献)	较低 (易受虚假信息影响)
激励导向	激励供方提供高质量数据	可能激励供方夸大数量或多样性等
验证机制	同意模型表现直接验证数据价值	缺乏对数据真实价值的有效验证机制
可扩展性	较低 (受限于模型训练和测试的计算资源)	较高 (可自定义, 易于扩展到更大规模与更多场景)

表 4　基于数据合作社的收益分配方案对比

方案	个体理性	联盟稳定性	社会福利最大化
比例分配方案	√	—	—
混合分配方案	—	—	√
罗宾汉分配方案	√	√	√

● 比例分配方案 PROP (proportional allocation scheme). 其核心在于确保每个成员主体所获得的价值与其个人数据贡献量严格成正比. 对于同质隐私成本, 完全按照贡献比例分配, 不进行其他的调整; 对于异质隐私成本, 同样按照贡献比例分配, 但对高隐私成本成员可能存在不公平, 导致后续不愿继续参与.

● 混合分配方案 HYB (hybrid allocation scheme). 通过为隐私成本高的成员主体提供额外的收益补偿, 以平衡各主体间的激励结构. 然而, 这种补偿机制的实施依赖于低隐私成本主体对额外负担的接受程度. 当补偿调整幅度过大时, 低隐私成本主体可能因为收益成本失衡而退出合作, 进而影响合作联盟的稳定性.

● 罗宾汉分配方案 RHOOD (RobinHood allocation scheme). 它是在比例分配与混合分配基础上的一种创新性扩展. 该方案不仅考虑了成员的个体贡献和隐私成本, 还引入了一种“再分配”机制, 即从高隐私成本成员处获取一部分收益, 再分配给低隐私成本成员, 以实现更公平的分配. 具体而言, 首先, 通过适度调整低隐私成本成员的部分收益, 向高隐私成本成员进行补偿性分配, 确保收益分配与隐私成本相匹配; 其次, 为低隐私成本成员设置基础收益保障 (防止低隐私成本成员因利益受损而退出数据合作社), 在隐私成本补偿之外提供固定收益分配, 从而在一定程度上维持合作参与方的稳定性.

小结: 在模型市场的收益分配中, 同质化分配与差异化补偿代表了两种不同的价值评估与分配逻辑. 基于数据贡献度的分配策略体现了同质化分配的特点, 即所有数据供方的价值均通过统一的模型性能影响来量化, 强调公平性和直接激励, 但其计算复杂性和高成本限制了适用性. 而基于差异化指标的分配策略则采用差异化补偿的思路, 通过数据成本、多样性等灵活指标进行价值评估, 降低了计算负担并增强了可扩展性, 但可能因评估标准的主观性而影响公平性. 随着模型市场的发展, 单一的衡量维度难以满足多样化需求, 分配策略应兼顾同质化分配的公平性与差异化补偿的灵活性, 混合分配策略的重要性日益凸显, 需要引起业界的重点关注.

与此同时, 需要特别注意的是, 收益分配的两个阶段在实践过程中需要紧密结合, 形成端到端的完整的收益分

配策略. 现有研究往往相对割裂地看待两个阶段的分配过程, 对端到端的收益分配策略的研究尚处于萌芽阶段. 这也是值得业界重点关注的方向.

4 主体策略性行为对收益分配的影响与应对

现有研究大多基于静态假设考虑模型市场的收益分配策略, 即无论是同质化分配还是差异化补偿, 往往将参与主体和模型环境视为固定不变的参数, 在此基础上形成固定的收益分配策略. 然而, 现实中的模型市场呈现出明显的动态演化特征: 一方面, 数据供方会根据先前的收益反馈不断调整其数据供给策略, 表现出动态策略性行为; 另一方面, 模型性能会随数据分布的变化而波动, 市场需求和竞争关系也会随时间推移而演变, 因此模型平台及其开发方会持续优化算法并调整数据需求偏好. 这种长期尺度下的双重动态性使得静态分配策略难以持续保证公平与效率, 因此需要突破传统短期静态分析的局限, 建立能够适应时变特征的动态调整机制. 下面分别介绍数据供方、模型平台及模型开发方在长期视角下表现出的策略性行为特征, 进而梳理现有研究对这些动态行为的应对机制.

4.1 数据供方的策略性行为

在模型市场中, 数据供方的策略性行为可能对市场运行效率、分配公平性以及市场的长期稳定发展产生显著影响. 这些行为往往源于数据供方对自身利益最大化的追求. 具体来说, 数据供方的策略性行为主要包括“投机取巧”“损人利己”和“利益联盟”这 3 种类型. 这些策略性行为不仅直接影响市场运行效率, 更与模型市场收益分配的核心目标: 个体理性 (确保各方参与收益不低于其机会成本)、联盟稳定性 (维持合作关系的长期存续与动态平衡) 与社会福利最大化 (提升整体模型性能与市场总收益) 密切相关. 不同性质的行为对上述目标的影响不同:“投机取巧”类行为主要通过侵蚀贡献公平性, 动摇个体理性的长期基础并削弱联盟凝聚力; “损人利己”类行为则直接破坏模型性能与数据可信度, 损害社会福利并危及联盟存续; 而“利益联盟”类行为则具有双重性, 正当性协作可同时促进 3 个目标, 而对抗性合谋则会以牺牲部分主体利益为代价, 破坏公平与稳定. 下文将具体分析 3 类典型策略性行为, 并阐释其与分配目标的冲突关系.

4.1.1 “投机取巧”性质的策略性行为

对于数据供方来说, 投机取巧的策略性行为通常表现为在不做出实质性贡献的情况下获取合作带来的收益. 最经典的就是“搭便车”行为[98]: 某些数据供方在未积极参与数据提供或模型训练的情况下, 借助其他供方的高质量贡献所提升的整体模型性能, 从而间接获得收益. 一个典型的例子是, 在模型市场中, 部分数据供方在早期投入了大量高质量数据, 显著提高了模型性能, 但是其他数据供方则可能利用市场收益的平均分配或基于参与数量的激励机制, 故意延迟数据提交, 从而在借助已有模型成果获利时减少早期投入的成本和风险. 因此, 为了防止出现“搭便车”的情况, 数据供方可能选择延迟自己的贡献.

然而, 这种方式会导致整体效率下降, 甚至在延迟足够大时, 可能影响联盟的平衡与稳定, 甚至导致市场的崩溃. 由于完全搭便车的行为容易被发现, 因而有些数据供方开始采取“半搭便车”的行为. 具体来说, 数据供方在合作中虽然做出了一定程度的贡献, 但其贡献水平低于原有的平均贡献水平, 主要分为以下两类情况: (1) 低质量参与: 虽然形式上参与了合作, 但提供的数据规模或数据质量明显不足; (2) 选择性参与: 只在数据需求明确且收益较高时提供数据, 而在需要长期投入或收益不确定时选择不参与. 相较于完全搭便车的行为, “半搭便车”通常更加隐蔽, 数据供方可能在多次合作中, 随机或者间隔选择几次进行投机取巧的行为, 以减少被发现的风险.

Fershtman 等人[13]最早系统性研究了长期合作中的稳定性问题, 尤其关注“搭便车”行为及历史贡献对于未来决策的影响. 联邦学习可以被视作一个模型市场的应用场景, 其中客户端 (数据供方) 通过本地数据或者模型更新参与协作训练, 以换取模型性能的提升或经济收益, 而中央服务器 (模型平台) 负责组织训练流程 (由平台中的模型开发方进行)、验证贡献真实性并分配收益.

针对数据供方的“搭便车”行为, 现有联邦学习研究主要通过动态奖惩机制和贡献验证框架来进行应对解决. 例如, Zhang 等人[99]从长期跨孤岛联邦学习角度研究了客户端的自私行为, 提出了一种带有惩罚方案的合作策略, 从而减少“搭便车”现象.

Bi 等人[100]建立了长期囚徒困境博弈模型, 揭示了客户端形成合作关系的决策条件, 并提出了一种基于惩罚机制的契约框架以激励客户端的持续贡献. 艾秋媛等人[101]为预防“搭便车”行为, 设计了一种新型的中央服务器, 不再统一全部更新参数, 而是根据客户端是否在当次贡献数据来差异化地反馈参数. 这一设定使得那些企图通过“搭便车”的客户端无法获得全部最新的模型参数.

Lu 等人[102]提出了两阶段客户端选择方案和奖励机制, 在候选阶段根据客户端的数据规模和模型质量计算奖励, 奖励高的客户端更有可能被选中, 在训练选择阶段根据拍卖机制选择成本最低的客户端. 最终通过奖励和拍卖, 激励客户端提供更多数据并降低训练成本, 来解决“搭便车”的问题. 目前的大多策略主要集中在利用技术手段来检测和消除搭便车者, 然而, 这些方法往往忽略了搭便车者自发转变为诚实参与者的可能性.

基于此, Xu 等人[103]提出了一个基于三方演化博弈论的动态激励模型 (TEG-DI) , 将模型市场的规则设计、价值创造、贡献验证分离成 3 种角色协同进行, 并通过实验证明了该模型能够有效地激励搭便车者转变为诚实参与者.

针对“投机取巧”性质的策略性行为, 部分研究在收益分配策略中引入长期激励机制, 对长期保持高参与频率的数据供方给予额外奖励; 同时考虑声誉因素, 赋予高声誉数据供方更高的收益分配权重. 声誉是一种可以基于过去的行为来衡量数据供方在当前活动中的可靠性或信誉度的常用指标[104,105]. 史专[106]提出了基于贡献和质量检测结果的声誉值方法, 通过分析模型版本的声誉值及其局部与全局梯度的分布, 构造了 FedAIM 奖励模块, 实现奖励公平的目的. Li 等人[107]设计的声誉更新规则遵循一次性偏差原则, 根据历史贡献动态调整声誉, 并将未来的收益分配与声誉值直接挂钩, 从而激励数据供方不断提供高质量的数据.

此类行为直接破坏个体理性的长期实现: 投机方虽然短期获利, 但其行为若被检测并惩罚 (如降低声誉或收益权重), 将导致其长期效用下降. 同时, 搭便车的行为削弱了联盟稳定性, 持续的投机行为可能导致高质量供方退出, 联盟瓦解, 使得联盟总收益下降. 因此, 基于声誉的长期激励机制本质上是在动态环境中对个体理性与联盟稳定性的再平衡.

4.1.2 “损人利己”性质的策略性行为

数据供方的“损人利己”性质的策略性行为主要包括两类: 数据复制攻击[108,109]和投毒攻击[110–112].

数据复制攻击是指恶意数据供方试图通过低成本复制数据以获取更大的收入. Han 等人[108]研究了恶意数据供方通过复制其数据并以多个虚拟身份参与市场, 从而在交易中获取超额收益的情形. 具体来说, 在模型市场交易中, 恶意数据供方 i 通过复制其私有数据 d_i, 并生成 k 个副本, 从而以 $k+1$ 个虚拟身份 $\{i_0,i_1,\ldots,i_k\}$ 参与市场, 这种策略使得供方 i 分配的总收益增加 $\{v_1,v_2,\ldots,v_k\}$, 而其他善意供方的收益则可能被不公平地稀释. 在上文提到的影响收益分配的因素中, 基于数据贡献度的分配策略受其影响较小, 因为简单的数据复制并不会给模型性能带来明显的提升, 从而不会获得额外的收益. 然而, 对于差异化指标的分配策略来说, 数据复制可以很明显地扩大数据规模. 特别是当采用数据规模作为核心指标时, 数据供方可凭借低成本复制虚增数据量, 从而获得较大收益, 进而扰乱市场公平, 影响联盟稳定性.

投毒攻击是另一种常见的“损人利己”性质的策略性行为, 其目的是通过操纵数据或模型来破坏数据质量或模型性能. Tolpegin 等人[110]将投毒攻击进一步分为两类: 数据投毒和模型投毒. 数据投毒是指数据供方通过恶意注入虚假数据或篡改现有数据来降低数据质量.

例如, 在联邦学习中, 恶意供方通过提供错误标签的数据来毒害全局模型[111,112], 从而显著降低最终模型的预测准确性. 模型投毒是指在联邦学习等分布式场景中, 恶意供方通过修改本地模型的参数更新, 向中央服务器发送对抗性参数, 从而破坏全局模型的性能[113–115]. 在模型市场中, 攻击行为对不同类型的收益分配策略会产生差异化影响. 例如, 攻击者通过操纵训练数据可以使全局模型的参数更新方向偏离正常轨迹, 从而导致模型错误率大幅上升. 这种攻击方式会严重影响基于数据贡献度的分配策略, 特别是当采用模型在测试数据上的表现作为核心评估指标时, 其破坏效果更为显著. 相比之下, 基于差异化指标的分配策略受攻击的影响相对有限. 因为这类策略主要依据数据本身的固有属性进行收益分配, 而非数据与模型表现的直接关联, 使得恶意供方难以通过操纵模型性能来获取不当收益.

针对数据复制攻击的防范, 可通过数据指纹技术或哈希算法来检测重复数据, 即通过特定算法处理原始数据生成具有唯一标识性的固定长度数字指纹, 同时结合基于区块链[116]的身份验证机制来识别重复的虚拟身份. 此外, 为了减少数据投毒攻击的机会, Gafni 等人[117]提出限制通信主权的方式, 采用服务端 (模型平台) 主导的周期性聚合 (如固定轮次同步), 而非客户端 (数据供方) 自由推送数据, 从而减少供方持续频繁注入有毒数据的风险. 对于模型投毒攻击, 可利用联邦学习中的鲁棒聚合方法[118]进行抵御. 针对“损人利己”性质的策略性行为, 可通过降低攻击者的声誉信用得分[106]、优化分配因素权重 (减少数据规模因素、增加数据质量与模型贡献度的权重) 等措施进行应对, 并将检测攻击的结果记录于区块链上以确保信用惩戒的公开透明.

数据复制攻击通过虚增数据规模稀释真实贡献者的收益, 导致核心供方效用下降, 违背了个体理性; 投毒攻击则是直接破坏模型性能, 使得市场的总收益下降, 进而影响所有参与方的长期收益. 因此, 防范机制本质上是在维护联盟的稳定性与整体社会福利的最大化.

4.1.3 “利益联盟”性质的策略性行为

在模型市场中, 数据异构性对模型性能具有显著的负面影响: 不同数据间的分布差异 (如类别不均衡、特征偏移或样本量悬殊) 往往会导致全局模型的局部泛化能力下降. 为应对这一问题, 模型平台通常会建立激励机制, 优先筛选并奖励那些提供更均衡数据分布的供方. 同时, 平台自身也会采取多种技术手段来缓解数据异构性的问题, 包括选择共享部分全局数据[119]、利用 Zero-shot 方法来生成伪数据混合训练[120]和利用 GAN 生成新样本来进行数据增强[121]等方式. 基于上述原因, 数据供方往往会自发组建“利益联盟” (例如具有数据互补性的协作团体), 通过成员间的数据共享或联合训练, 将联盟内部的数据分布调整为“局部异构但全局平衡”的状态, 从而通过提升整体模型性能以获得更大的总收益, 并借助合理的分配策略使每个成员获得的收益超过单独参与时的水平.

从数据供方角度来看, 现有研究主要采用聚类和联盟构建的思想来应对数据异构性问题. Wang 等人[122]根据数据分布相似性对数据供方进行聚类, 并在每个聚类中选择最优供方形成组合来参与模型训练, 从而提升异构数据环境下的模型性能. Long 等人[123]则提出了一种多中心聚合机制, 不再训练单一全局模型, 而是将客户端聚类为多个组 (中心), 每组训练一个专属的全局模型, 客户端能够根据自身数据特点选择最匹配的组, 实现“分而治之”的效果, 从而获得较高的收益. Arisdakessian 等人[124]在联邦学习中提出一种信任支持的联盟博弈, 建立端到端的信任机制, 使得客户端能够自主形成利益目标一致的联盟. Zhang 等人[125]在上述基础上进一步提出了联盟式联邦学习框架, 通过让数据供方参与联盟形成博弈, 使其能够相互协作并组建联盟, 既降低了数据的不平衡性, 又通过选择最优联盟子集来提高模型的精度, 从而提升整体收益.

值得注意的是, 除了上述数据供方通过形成团体联盟来获取正当的额外收益的行为 (正当性协作联盟) 之外, 近期研究也开始关注数据供方之间通过复杂的互动来获取不正当的收益的现象. 例如, 参与者 A 通过向参与者 B 支付“租金”[103], 以换取不参与训练, 但同时仍然受益于全局模型. 这种利益联盟的行为具有明显的“搭便车”特征, 属于对抗性合谋联盟的范畴.

正当性协作联盟与对抗性合谋联盟对分配目标的影响呈现显著差异: 正当联盟通过数据互补提升模型性能, 既增加了联盟总收益 (促进社会福利最大化), 又使联盟内各成员收益高于单独参与时的水平 (满足个体理性), 同时因利益共享强化了联盟稳定性; 而对抗性合谋联盟 (如“租金交换式”的搭便车合谋) 则通过侵占非联盟成员的收益, 导致部分主体效用低于机会成本 (违背个体理性), 同时破坏收益分配的公平性, 引发联盟分裂风险 (冲击联盟稳定性). 因此, 联盟构建机制需在激励协作与防范合谋之间取得平衡, 以同时满足个体理性、联盟稳定与社会福利目标.

4.2 模型平台及模型开发方策略性行为

在模型市场的收益分配中, 现有研究主要聚焦于单轮的激励建模, 通过反复迭代单周期的优化来实现全局的激励效果[126–128].

然而, 这种研究范式忽视了长期尺度下多轮协作之间的相互影响. 当采用人为拆解的单轮激励框架时, 模型平台不仅面临多周期预算分配的动态规划难题, 同时还需考虑不同预算分配方法对模型迭代质量的传导效应. 为此,

部分研究开始考虑在单一预算限制的场景下如何实现多轮的激励机制设计. 例如, Ren 等人[129]提出了一种 BCQM 框架, 即用一次单预算来设计多轮联邦学习的激励机制. 他们将多轮的激励机制建模为一个预算约束下的累积质量最大化问题 (BCQM), 并提出了一种名为 RABORN 的反向拍卖机制, 实现了在单预算下高效选择参与者并分配奖励, 同时确保个体理性、真实性和计算效率, 从而优化了模型性能和预算利用率. 部分研究则着眼于在激励机制中引入声誉等代表长期特性的指标, Li 等人[107]设计了一个可持续的联邦激励机制 (RATE), 通过引入声誉来衡量数据供方的长期贡献, 从而最大化 RATE 的社会福利.

在长期尺度下, 模型需方支付费用可能存在滞后性, 导致模型平台无法及时获得预期的收益. 这种支付延迟会直接影响参与方的参与积极性——参与方因贡献和回报分配延迟而退出或者消极参与, 从而损害联盟的长期稳定性. 为此, 部分研究开始将支付及时性引入到收益分配机制当中. 例如, Yu 等人[130]针对贡献与回报之间的时间差问题设计了一种动态调整预算分配的奖励机制 (FLI), 基于支付时效性定义遗憾值, 并根据遗憾值对参与方未及时收到收益的情况进行补偿, 以实现遗憾值最小化. 值得注意的是, 时效性优化需要平台与数据供方的双向协同: 平台需解决支付延迟问题, 而数据供方则需提升任务执行效率. 对此, Jiang 等人[131]提出了基于时效性的选择激励机制 (TsIFedCrowd), 激励数据供方以更高的质量和效率完成数据支持的任务.

另一方面, 针对长期场景下多任务并发场景的普遍性, 部分研究视角从传统的单一任务发布者 (单一模型平台和单一模型开发方) 假设向多任务发布者场景拓展. 例如, Xuan 等人[132]提出了多任务发布者场景下基于契约理论的奖励机制, 赋予工作者节点自主选择权, 使其不再受限于唯一任务的限制, 而是有权主动选择和比较收益. 同样地, 在模型市场中, 当平台采用多任务开放的策略时, 多个模型开发方的参与不仅扩展了数据供方的选择空间, 还通过竞争机制促进了模型产品的多样化.

此外, 正如在第 2.1 节中提到的“模型平台负责收益分配策略设计, 但本身不获得 R_j 的分配”, 现有研究大多假设模型平台是中立的分配者, 但现实中平台作为独立利益主体, 存在强烈的自利动机, 其策略性行为将会打破原有假设的公平. 具体来说, 平台可能对分配规则进行操纵, 通过调整分配参数 (如公式 (1) 中的 α、β、γ) 权重、扭曲贡献度的评估标准, 从而最大化自身利益, 导致数据供方与模型开发方的合理收益被侵犯; 平台也可能滥用自身的信息优势, 通过信息不对称进行策略性定价 (如向数据供方隐瞒模型实际收益、向模型需方虚报成本); 平台也可能通过垄断性策略, 依靠自身的数据市场的主导地位限制数据供方与模型开发方的选择权, 强化自身议价能力, 长此以往导致市场竞争力不足, 创新活力下降.

总之, 当前大部分研究对长期尺度下模型平台和模型开发方的策略性行为的研究还处于初步阶段, 对相应的收益分配策略的研究同样方兴未艾.

5 总结与展望

本文聚焦模型市场这一关键的数据市场分支, 对其中的收益分配策略进行了全面的总结与系统化的分析. 本文从模型市场的运作模式出发, 梳理了收益分配的两阶段模式, 在此基础上形式化定义了涵盖分配主体、分配模式和分配目标的收益分配策略问题, 为进一步的研究提供了一个体系化框架. 基于这一框架, 本文从数据贡献度和差异化指标角度出发, 系统探讨影响收益分配权重计算的关键因素及其研究演进; 并且进一步分析相关主体的策略性行为及其对收益分配策略设计的影响.

通过对现有文献的系统性梳理, 可以看出, 收益分配策略的研究大多聚焦于数据供方之间如何进行收益分配的细化, 而对于主体间的分配策略研究仍缺乏系统性探讨. 这一阶段的分配策略涉及更复杂的利益主体相互关系以及风险分担机制设计. 本文的框架与分析也存在两个方面的主要不足, 这些不足反映了从理论向实践转化过程中的难点.

(1) 分配框架的假设简化与现实复杂性存在差距. 本文第 2 节所提出的收益分配形式化框架, 虽然在理论分析上提供了清晰的建模基础, 但其建立在若干理想化假设之上, 尤其是假设各参与方的成本与效用函数为已知且可量化. 但在现实中, 这一假设往往难以满足: 数据供方的隐私成本、数据整理与计算成本具有高度主观性和异质性;

模型开发方的训练成本随算法、硬件与数据规模动态变化; 模型需方的支付意愿也受市场环境、模型性能及替代品影响而波动. 此外, 信息不对称、模型性能波动以及市场风险等因素, 进一步增加了收益分配策略设计的复杂性. 因此, 未来研究需进一步放宽完全信息假设, 探索在不完全信息、风险与不确定性环境下的收益分配机制, 以增强分配策略在真实场景中的适用性与稳健性.

(2) 市场主体的职能重叠与利益冲突增加了分配策略的复杂性. 我们在本文的探讨中没有考虑模型需方的参与. 事实上, 模型需方本身有可能参与到数据供给, 具备一定的数据供方的职能. 那么模型需方的场景化数据与数据供方的通用数据价值如何区分? 例如, 银行提供的场景化信贷数据, 和数据公司提供的通用用户行为数据, 都用在了风控模型训练中, 但场景化数据决定了模型能不能适配需方需求, 通用数据决定了模型的基础性能, 二者对最终模型的价值不一样, 因此不能简单地按数据量计算, 也不能按同一套贡献标准计算. 另外, 模型需方也可能参与到模型开发过程, 具备模型开发方的职能, 例如参与模型的迭代反馈. 那么需方的迭代反馈与模型开发方的技术优化贡献如何权衡? 这些职能重叠带来的参与式价值共创, 可能影响到数据模型的定价, 从而对端到端的收益分配策略设计带来了新的复杂性.

最直接的, 模型需方与数据供方、模型开发方、平台等其他主体之间存在天然的利益冲突. 需方的核心诉求是成本最小化和价值最大化: 既希望以更低的费用获取高精度、高适配性的模型产品, 又可能通过降低开发报酬等方式控制总投入, 甚至在反馈优化等价值共创环节, 倾向于弱化自身投入以减少成本分摊. 而数据供方的核心目标是收益最大化, 模型开发方则是希望通过技术贡献、研发投入等获得合理报酬, 平台作为中间枢纽, 既要平衡各方利益以维持市场稳定, 也存在通过规则设计获取自身利益的诉求. 这种“需方成本控制”与“供方/开发方/平台利益追求”的核心诉求对立, 极容易发生分配博弈: 如需方刻意压低模型采购报价, 导致供方与开发方收益不足以覆盖成本; 或供方联合抬高数据价格、开发方夸大技术贡献, 推高需方使用成本. 此外, 利益冲突还会延伸至贡献量化、规则制定等环节, 最终破坏联盟稳定性与市场可持续性. 因此, 如何进一步厘清多主体之间的关系, 形成端到端的收益分配策略, 是未来研究中的一个关键问题.

此外, 本文总结了模型市场中的收益分配研究中 3 个明显的发展趋势: 一是收益分配依据正由同质化分配向差异化补偿演进, 二是策略设计正由短期静态向长期动态拓展, 三是分配逻辑从单一主体利益考量向多主体复杂协同深化. 然而, 这 3 大趋势尚处于初步探索阶段, 在构建高效且合理的收益分配策略方面, 仍面临以下 3 项关键挑战.

(1) 如何综合衡量和整合多维差异化指标以设计混合收益分配策略? 相较于如模型预测准确率等可直接反映数据贡献的传统指标, 多维差异化指标主要用于表征数据特质, 在实际应用中面临主观性强、场景依赖明显及影响滞后的挑战, 具体体现在以下 3 个方面: 首先是主观性问题. 以数据规模为例, 通常人们认为数据规模越大, 其价值越高, 但如何为其赋予合理的数值或权重, 缺乏统一评价基准. 其次是场景依赖性. 同一指标在不同应用场景中权重差异可能很大, 例如在医疗影像任务中, 数据的像素多样性至关重要, 而在工业质检中, 异常样本的稀缺性更具价值. 最后是时滞效应. 部分指标的影响需要通过长期协作才能显现. 例如数据新鲜度对模型性能下降的影响具有延迟性, 难以在单次实验中被观察. 因此, 如何在收益分配策略设计中合理衡量与有效整合这些差异化指标, 构建具有可解释性与泛化良好的混合收益分配策略, 是一个亟待解决的关键挑战.

(2) 如何理解长期尺度下模型市场主体的动态性行为以设计动态策略? 在长期尺度下, 模型市场的稳定性和可持续性是收益分配策略的核心目标. 由于现实中的数据产品交易往往涉及多轮次交互, 因此收益分配策略需要从静态的单次交易场景扩展到动态的多次交易场景; 与此同时, 模型市场中的数据供方、模型需方等主体的策略性行为 (如供方选择性提供低价值数据、需方通过策略性报价压低价格), 也要求分配策略具有反操纵的动态设计. 此外, 由于市场环境本身也随着时间在不断变化, 分配策略本身也应具备环境适应性, 能够根据市场的发展阶段和外部条件变化进行动态调整和优化. 这意味着, 有效的收益分配策略不是静态的, 而是与市场演进相互协调的. 因此, 如何在长期尺度下实现收益分配策略的自适应演化, 同样是模型市场中构建收益分配策略的关键挑战之一.

(3) 如何适配模型需方等深度参与的共创模式以设计协同分配机制? 模型需方的参与形式呈现多元化特征 (需求定义、反馈优化、联合开发等), 不同参与形式的价值贡献难以用统一标准量化, 容易导致分配公平性争议.

更关键的是, 需方的各类参与行为并非孤立存在, 而是与供方、开发方等主体形成深度交织的协同关系. 因此, 如何建立多维度、可解释、场景化的需方贡献量化体系, 平衡需方与供方等主体之间的利益关系, 设计出既满足个体理性又保障联盟稳定性的协同分配机制, 成为当前收益分配策略设计的关键挑战.

针对上述挑战, 未来研究可以从以下角度实现突破: 一是构建融合领域知识的动态量化模型, 通过联邦学习、博弈论等方法实现差异化指标的综合评估; 二是设计具有环境适应能力的智能分配策略, 利用强化学习等技术实现策略的自主优化. 在此基础上, 探索如何将差异化指标评估与策略动态优化机制有机地嵌入到收益分配的全流程中, 构建公平性好、自适应性强、可解释性高的端到端收益分配策略. 三是探索模型需方参与下的协同分配机制, 平衡数据供方、模型平台、模型开发方、模型需方的利益关系, 确保分配结果满足个体理性与联盟稳定性的约束; 并建立相应的信息对称机制, 例如通过区块链存证、第三方评估等技术, 解决主体间的信息不对称问题, 防范各主体的策略性行为所带来的分配失衡风险. 这些角度的突破, 将有助于优化模型市场中收益分配策略的设计, 支撑模型市场的长期稳定和可持续发展.

References

[1] Yang M, Feng HL, Wang X, Huo JD, Jiao XG, Zhang H. Survey of data factor market: Value, pricing, and trading. Chinese Journal of Network and Information Security, 2024, 10(3): 1–19 (in Chinese with English abstract). [doi: 10.11959/j.issn.2096-109x.2024036]

[2] Huang KM, Du XY. Value chain model of data governance and its application on data governance regulation analysis. Big Data Research, 2022, 8(4): 3–16 (in Chinese with English abstract). [doi: 10.11959/j.issn.2096-0271.2022062]

[3] Jiang RH. The risks, causes and practical strategies of embedding artificial intelligence into social governance. Journal of Sichuan University of Science & Engineering (Social Sciences Edition), 2024, 39(6): 30–38 (in Chinese with English abstract). [doi: 10.11965/xbew20240604]

[4] Shang XX, Han HT, Zhu ZZ. Mechanism design of right to earnings of data utilization based on evolutionary game model. Computer Science, 2021, 48(3): 144–150 (in Chinese with English abstract). [doi: 10.11896/jsjkx.201100056]

[5] Chen LJ, Koutris P, Kumar A. Towards model-based pricing for machine learning in a data marketplace. In: Proc. of the 2019 Int'l Conf. on Management of Data. Amsterdam: ACM, 2019. 1535–1552. [doi: 10.1145/3299869.3300078]

[6] Liu JF, Lou J, Liu JX, Xiong L, Pei J, Sun JM. Dealer: An end-to-end model marketplace with differential privacy. Proc. of the VLDB Endowment, 2021, 14(6): 957–969. [doi: 10.14778/3447689.3447700]

[7] Pei J. A survey on data pricing: From economics to data science. IEEE Trans. on Knowledge and Data Engineering, 2022, 34(10): 4586–4608. [doi: 10.1109/TKDE.2020.3045927]

[8] Qian M, Musa AA, Biswas M, Guo YF, Liao WX, Yu W. Survey of artificial intelligence model marketplace. Future Internet, 2025, 17(1): 35. [doi: 10.3390/fi17010035]

[9] Jia RX, Dao D, Wang BX, Hubis FA, Gurel NM, Li B, Zhang C, Spanos C, Song D. Efficient task-specific data valuation for nearest neighbor algorithms. Proc. of the VLDB Endowment, 2019, 12(11): 1610–1623. [doi: 10.14778/3342263.3342637]

[10] Wang JT, Jia RX. Data Banzhaf: A robust data valuation framework for machine learning. In: Proc. of the 26th Int'l Conf. on Artificial Intelligence and Statistics. Valencia: PMLR, 2023. 6388–6421.

[11] Yan T, Procaccia AD. If you like Shapley then you'll love the core. In: Proc. of the 35th AAAI Conf. on Artificial Intelligence. Vancouve: AAAI Press, 2021. 5751–5759. [doi: 10.1609/aaai.v35i6.16721]

[12] Dawande M, Mehta S, Mu LY. Robin hood to the rescue: Sustainable revenue-allocation schemes for data cooperatives. Production and Operations Management, 2023, 32(8): 2560–2577. [doi: 10.1111/poms.13995]

[13] Fershtman C, Nitzan S. Dynamic voluntary provision of public goods. European Economic Review, 1991, 35(5): 1057–1067. [doi: 10.1016/0014-2921(91)90004-3]

[14] Liu N, Hao XJ, Chen YH. A review and comparative analysis of domestic and foreign research on big data pricing methods. Big Data Research, 2021, 7(6): 89–102 (in Chinese with English abstract). [doi: 10.11959/j.issn.2096-0271.2021063]

[15] Liang F, Yu W, An D, Yang QY, Fu XW, Zhao W. A survey on big data market: Pricing, trading and protection. IEEE Access, 2018, 6: 15132–15154. [doi: 10.1109/ACCESS.2018.2806881]

[16] Zhang JY, Bi YR, Cheng MY, Liu JF, Ren K, Sun QH, Wu YH, Cao Y, Fernandez RC, Xu HF, Jia RX, Kwon Y, Pei J, Wang JT, Xia HC, Xiong L, Yu XH, Zou J. A survey on data markets. arXiv:2411.07267, 2024.

[17] Cai L, Huang ZH, Liang Y, Zhu YY. Survey of data pricing. Journal of Frontiers of Computer Science and Technology, 2021, 15(9): 1595–1606 (in Chinese with English abstract). [doi: 10.3778/j.issn.1673-9418.2103069]

[18] Jiang D, Yuan Y, Zhang XW, Wang GR. Survey on data pricing and trading research. Ruan Jian Xue Bao/Journal of Software, 2023, 34(3): 1396–1424 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6751.htm [doi: 10.13328/j.cnki.jos.006751]

[19] Admati AR, Pfleiderer P. A monopolistic market for information. Journal of Economic Theory, 1986, 39(2): 400–438. [doi: 10.1016/0022-0531(86)90052-9]

[20] Balazinska M, Howe B, Suciu D. Data markets in the cloud: An opportunity for the database community. Proc. of the VLDB Endowment, 2011, 4(12): 1482–1485. [doi: 10.14778/3402755.3402801]

[21] Chawla S, Deep S, Koutrisw P, Teng YF. Revenue maximization for query pricing. Proc. of the VLDB Endowment, 2019, 13(1): 1–14. [doi: 10.14778/3357377.3357378]

[22] Koutris P, Upadhyaya P, Balazinska M, Howe B, Suciu D. Query-based data pricing. In: Proc. of the 31st ACM SIGMOD-SIGACT-SIGAI Symp. on Principles of Database Systems. Scottsdale: ACM, 2012. 167–178. [doi: 10.1145/2213556.2213582]

[23] Koutris P, Upadhyaya P, Balazinska M, Howe B, Suciu D. Toward practical query pricing with QueryMarket. In: Proc. of the 2013 Int'l Conf. on Management of Data. New York: ACM, 2013. 613–624. [doi: 10.1145/2463676.2465335]

[24] Kurtulmus AB, Daniel K. Trustless machine learning contracts; Evaluating and exchanging machine learning models on the Ethereum blockchain. In: Proc. of the 2018 Artificial Intelligence Conf. San Francisco: 2018. [doi: 10.48550/arXiv.1802.10185]

[25] Pei J, Fernandez RC, Yu XH. Data and AI model markets: Opportunities for data and model sharing, discovery, and integration. Proc. of the VLDB Endowment, 2023, 16(12): 3872–3873. [doi: 10.14778/3611540.3611573]

[26] von Stackelberg H. Market Structure and Equilibrium. Heidelberg: Springer, 2011. [doi: 10.1007/978-3-642-12586-7]

[27] Friedman D. The double auction market institution: A survey. In: Friedman D, ed. The Double Auction Market. New York: Routledge, 2018. 3–26.

[28] Cai H, Zhu YM, Li J, Yu JD. Double auction for a data trading market with preferences and conflicts of interest. The Computer Journal, 2019, 62(10): 1490–1504. [doi: 10.1093/comjnl/bxz025]

[29] Myerson RB. Optimal auction design. Mathematics of Operations Research, 1981, 6(1): 58–73. [doi: 10.1287/moor.6.1.58]

[30] Sutton J. Non-cooperative bargaining theory: An introduction. The Review of Economic Studies, 1986, 53(5): 709–724. [doi: 10.2307/2297715]

[31] Agarwal A, Dahleh M, Sarkar T. A marketplace for data: An algorithmic solution. In: Proc. of the 2019 ACM Conf. on Economics and Computation. Phoenix: ACM, 2019. 701–726. [doi: 10.1145/3328526.3329589]

[32] Zhan YF, Zhang J, Hong ZC, Wu LJ, Li P, Guo S. A survey of incentive mechanism design for federated learning. IEEE Trans. on Emerging Topics in Computing, 2022, 10(2): 1035–1044. [doi: 10.1109/TETC.2021.3063517]

[33] Bergantiños G, Moreno-Ternero JD. Revenue sharing at music streaming platforms. Management Science, 2025, 71(10): 8319–8335. [doi: 10.1287/mnsc.2023.03830]

[34] Jensen FJ. Rethinking royalties: Alternative payment systems on music streaming platforms. Journal of Cultural Economics, 2024, 48(3): 439–462. [doi: 10.1007/s10824-024-09507-z]

[35] Stahl F, Vossen G. Data quality scores for pricing on data marketplaces. In: Proc. of the 8th Asian Conf. on Intelligent Information and Database Systems. Da Nang: Springer, 2016. 215–224. [doi: 10.1007/978-3-662-49381-6_21]

[36] Yu HF, Zhang MX. Data pricing strategy based on data quality. Computers & Industrial Engineering, 2017, 112: 1–10. [doi: 10.1016/j.cie.2017.08.008]

[37] Yang J, Zhao CC, Xing CX. Big data market optimization pricing model based on data quality. Complexity, 2019, 2019: 5964068. [doi: 10.1155/2019/5964068]

[38] Jiao YT, Wang P, Niyato D, Abu Alsheikh M, Feng SH. Profit maximization auction and data management in big data markets. In: Proc. of the 2017 IEEE Wireless Communications and Networking Conf. San Francisco: IEEE, 2017. 1–6. [doi: 10.1109/WCNC.2017.7925760]

[39] Mei LJ, Li W, Nie K. Pricing decision analysis for information services of the Internet of Things based on Stackelberg game. In: Zhang ZJ, Zhang RT, Zhang JL, eds. Proc. of the 2nd Int'l Conf. on Logistics, Informatics and Service Science. Heidelberg: Springer, 2013, 1097–1104. [doi: 10.1007/978-3-642-32054-5_155]

[40] Bi YR, Liu JF, Zhao C, Zhao JY, Ren K, Xiong L. Share: Stackelberg-Nash based data markets. In: Proc. of the 40th IEEE Int'l Conf. on Data Engineering (ICDE). Utrecht: IEEE, 2024. 3573–3586. [doi: 10.1109/ICDE60146.2024.00275]

[41] Liu K, Qiu XY, Chen WH, Chen X, Zheng ZB. Optimal pricing mechanism for data market in blockchain-enhanced Internet of Things. IEEE Internet of Things Journal, 2019, 6(6): 9748–9761. [doi: 10.1109/JIOT.2019.2931370]

[42] Xu CZ, Zhu K, Yi CY, Wang R. Data pricing for blockchain-based car sharing: A Stackelberg game approach. In: Proc. of the 2020 IEEE Global Communications Conf. Taipei: IEEE, 2020. 1–5. [doi: 10.1109/GLOBECOM42002.2020.9322221]

[43] Bi YR, Wu YH, Liu JF, Ren K, Xiong L. When data pricing meets non-cooperative game theory. In: Proc. of the 40th IEEE Int'l Conf. on Data Engineering (ICDE). Utrecht: IEEE, 2024. 5548–5559. [doi: 10.1109/ICDE60146.2024.00443]

[44] Roth AE. Laboratory Experimentation in Economics: Six Points of View. Cambridge: Cambridge University Press, 1987.

[45] Chen YQ, Yang XD, Qin X, Yu H, Chan P, Shen ZQ. Dealing with label quality disparity in federated learning. In: Yang Q, Fan LX, Yu H, eds. Federated Learning: Privacy and Incentive. Cham: Springer, 2020: 108–121. [doi: 10.1007/978-3-030-63076-8_8]

[46] Zhang JY, Xia HC, Sun QH, Liu JF, Xiong L, Pei J, Ren K. Dynamic Shapley value computation. In: Proc. of the 39th Int'l Conf. on Data Engineering. Anaheim: IEEE, 2023. 639–652. [doi: 10.1109/ICDE55515.2023.00055]

[47] Sundararajan M, Najmi A. The many Shapley values for model explanation. In: Proc. of the 37th Int'l Conf. on Machine Learning. JMLR.org, 2020. 859. [doi: 10.5555/3524938.3525797]

[48] van Campen T, Hamers H, Husslage B, Lindelauf R. A new approximation method for the Shapley value applied to the WTC 9/11 terrorist attack. Social Network Analysis and Mining, 2018, 8(1): 3. [doi: 10.1007/s13278-017-0480-z]

[49] Wang JT, Zhu YQ, Wang YX, Jia RX, Mittal P. A privacy-friendly approach to data valuation. In: Proc. of the 37th Conf. on Neural Information Processing Systems. New Orleans: OpenReview.net, 2023.

[50] Wang Y, Li GL, Li KY. Survey on contribution evaluation for federated learning. Ruan Jian Xue Bao/Journal of Software, 2023, 34(3): 1168–1192 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6786.htm [doi: 10.13328/j.cnki.jos.006786]

[51] Roth AE. Introduction to the Shapley value. In: Roth AE, ed. The Shapley Value. Cambridge: Cambridge University Press, 1988. 1–28. [doi: 10.1017/CBO9780511528446.002]

[52] Jia RX, Dao D, Wang BX, Hubis FA, Hynes N, Gürel NM, Li B, Zhang C, Song D, Spanos CJ. Towards efficient data valuation based on the Shapley value. In: Proc. of the 22nd Int'l Conf. on Artificial Intelligence and Statistics. Naha: PMLR, 2019. 1167–1176.

[53] Ghorbani A, Zou J. Data Shapley: Equitable valuation of data for machine learning. In: Proc. of the 36th Int'l Conf. on Machine Learning. Long Beach: PMLR, 2019. 2242–2251.

[54] Liu ZL, Chen YY, Yu H, Liu Y, Cui LZ. GTG-Shapley: Efficient and accurate participant contribution evaluation in federated learning. ACM Trans. on Intelligent Systems and Technology, 2022, 13(4): 60. [doi: 10.1145/3501811]

[55] Zhang JY, Sun QH, Liu JF, Xiong L, Pei J, Ren K. Efficient sampling approaches to Shapley value approximation. Proc. of the ACM on Management of Data, 2023, 1(1): 48. [doi: 10.1145/3588728]

[56] Li WD, Yu YL. Faster approximation of probabilistic and distributional values via least squares. In: Proc. of the 12th Int'l Conf. on Learning Representations. Vienna: OpenReview.net, 2024.

[57] Kolpaczki P, Bengs V, Muschalik M, Hüllermeier E. Approximating the Shapley value without marginal contributions. In: Proc. of the 38th AAAI Conf. on Artificial Intelligence. Vancouver: AAAI Press, 2024. 13246–13255. [doi: 10.1609/aaai.v38i12.29225]

[58] Pang JY, Pei J, Xia HC, Li X, Liu JF. Shapley value estimation based on differential matrix. Proc. of the ACM on Management of Data, 2025, 3(1): 75. [doi: 10.1145/3709725]

[59] Kwon Y, Zou J. Beta Shapley: A unified and noise-reduced data valuation framework for machine learning. In: Proc. of the 25th Int'l Conf. on Artificial Intelligence and Statistics. Valencia: PMLR, 2022. 8780–8802.

[60] Ghorbani A, Kim MP, Zou J. A distributional framework for data valuation. In: Proc. of the 37th Int'l Conf. on Machine Learning. Vienna: JMLR.org, 2020. 331.

[61] Liu ZH, Just HA, Chang XY, Chen X, Jia RX. 2D-Shapley: A framework for fragmented data valuation. In: Proc. of the 40th Int'l Conf. on Machine Learning. Honolulu: JMLR.org, 2023. 899.

[62] Wang JT, Mittal P, Song D, Jia RX. Data Shapley in one training run. In: Proc. of the 13th Int'l Conf. on Learning Representations. Singapore: OpenReview.net, 2025.

[63] Kwon Y, Zou J. Data-OOB: Out-of-bag estimate as a simple and efficient data value. In: Proc. of the 40th Int'l Conf. on Machine Learning. Honolulu: JMLR.org, 2023. 749.

[64] Just HA, Kang FY, Wang TH, Zeng Y, Ko M, Jin M, Jia RX. LAVA: Data valuation without pre-specified learning algorithms. In: Proc. of the 11th Int'l Conf. on Learning Representations. Kigali: OpenReview.net, 2023.

[65] Wu ZX, Shu Y, Low BKH. DAVINZ: Data valuation using deep neural networks at initialization. In: Proc. of the 39th Int'l Conf. on Machine Learning. Baltimore: PMLR, 2022. 24150–24176.

[66] Li WD, Yu YL. Robust data valuation with weighted Banzhaf values. In: Proc. of the 37th Int'l Conf. on Neural Information Processing Systems. New Orleans: Curran Associates Inc., 2023. 2637.

[67] Owen G. Values of games with a priori unions. In: Henn R, Moeschlin O, eds. Mathematical Economics and Game Theory. Berlin: Springer, 1977. 76–88. [doi: 10.1007/978-3-642-45494-3_7]

[68] Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and

Computing, 2017, 27(5): 1413–1432. [doi: 10.1007/s11222-016-9696-4]

[69] Koh PW, Liang P. Understanding black-box predictions via influence functions. In: Proc. of the 34th Int'l Conf. on Machine Learning. Sydney: JMLR.org, 2017. 1885–1894.

[70] Du XY, Chen YG, Fan J, Lu W. Data wrangling: A key technique of data governance. Big Data Research, 2019, 5(3): 13–22 (in Chinese with English abstract). [doi: 10.11959/j.issn.2096-0271.2019020]

[71] Hellerstein J M, Heer J, Kandel S. Self-service data preparation: Research to practice. IEEE Data Engineering Bulletin, 2018, 41(2): 23–34.

[72] Meng XF, Ci X. Big data management: Concepts, techniques and challenges. Journal of Computer Research and Development, 2013, 50(1): 146–169 (in Chinese with English abstract). [doi: 10.3969/j.issn.1672-5719.2017.19.112]

[73] Albrecht JP. How the GDPR will change the world. European Data Protection Law Review, 2016, 2(3): 287–289. [doi: 10.21552/edpl/2016/3/4]

[74] Li H, Yu L, He W. The impact of GDPR on global technology development. Journal of Global Information Technology Management, 2019, 22(1): 1–6. [doi: 10.1080/1097198X.2019.1569186]

[75] Yuan M, Liu F, Zeng C, Xie L. Review of data quality dimension and framework. Journal of Jilin University (Information Science Edition), 2018, 36(4): 444–451 (in Chinese with English abstract). [doi: 10.3969/j.issn.1671-5896.2018.04.014]

[76] Yi YZ, Zhu NF, He JS, Jurcut AD, Ma XJ, Luo YH. A privacy-dependent condition-based privacy-preserving information sharing scheme in online social networks. Computer Communications, 2023, 200: 149–160. [doi: 10.1016/j.comcom.2023.01.010]

[77] Wei K, Li J, Ding M, Ma C, Yang HH, Farokhi F, Jin S, Quek TQS, Poor HV. Federated learning with differential privacy: Algorithms and performance analysis. IEEE Trans. on Information Forensics and Security, 2020, 15: 3454–3469. [doi: 10.1109/TIFS.2020.2988575]

[78] Dwork C, McSherry F, Nissim K, Smith A. Calibrating noise to sensitivity in private data analysis. In: Proc. of the 3rd Theory of Cryptography. New York: Springer, 2006. 265–284. [doi: 10.1007/11681878_14]

[79] Nikolov A, Talwar K, Zhang L. The geometry of differential privacy: The sparse and approximate cases. In: Proc. of the 45th Annual ACM Symp. on Theory of Computing. Palo Alto: ACM, 2013. 351–360. [doi: 10.1145/2488608.2488652]

[80] McSherry F, Talwar K. Mechanism design via differential privacy. In: Proc. of the 48th Annual IEEE Symp. on Foundations of Computer Science. Providence: IEEE, 2007. 94–103. [doi: 10.1109/FOCS.2007.66]

[81] Li FH, Li H, Jia Y, Yu NH, Weng J. Privacy computing: Concept, connotation and its research trend. Journal on Communications, 2016, 37(4): 1–11 (in Chinese with English abstract). [doi: 10.11959/j.issn.1000-436x.2016078]

[82] Wang LX, Gu QQ. Differentially private iterative gradient hard thresholding for sparse learning. In: Proc. of the 28th Int'l Joint Conf. on Artificial Intelligence. Macao: AAAI Press, 2019. 3740–3747.

[83] Ghosh A, Roth A. Selling privacy at auction. In: Proc. of the 12th ACM Conf. on Electronic Commerce. San Jose: ACM, 2011. 199–208. [doi: 10.1145/1993574.1993605]

[84] Sun P, Chen X, Liao GC, Huang JW. A profit-maximizing model marketplace with differentially private federated learning. In: Proc. of the 2022 IEEE Conf. on Computer Communications. London: IEEE, 2022. 1439–1448. [doi: 10.1109/INFOCOM48880.2022.9796833]

[85] Domingos P. A few useful things to know about machine learning. Communications of the ACM, 2012, 55(10): 78–87. [doi: 10.1145/2347736.2347755]

[86] Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. ACM Computing Surveys, 2022, 54(6): 115. [doi: 10.1145/3457607]

[87] Barocas S, Hardt M, Narayanan A. Fairness and Machine Learning: Limitations and Opportunities. Cambridge: MIT Press, 2023.

[88] Bishop CM. Pattern Recognition and Machine Learning. New York: Springer, 2006.

[89] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed., New York: Springer, 2009. [doi: 10.1007/978-0-387-84858-7]

[90] Xu XY, Wu ZX, Foo CS, Low BKH. Validation free and replication robust volume-based data valuation. In: Proc. of the 35th Int'l Conf. on Neural Information Processing Systems. Curran Associates Inc., 2021. 829.

[91] Shannon CE. A mathematical theory of communication. The Bell System Technical Journal, 1948, 27(3): 379–423. [doi: 10.1002/j.1538-7305.1948.tb01338.x]

[92] Li T, Sahu AK, Talwalkar A, Smith V. Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine, 2020, 37(3): 50–60. [doi: 10.1109/MSP.2020.2975749]

[93] Zhan YF, Li P, Wang K, Guo S, Xia YQ. Big data analytics by crowdlearning: Architecture and mechanism design. IEEE Network, 2020, 34(3): 143–147. [doi: 10.1109/MNET.001.1900286]

[94] Li T, Li J, Chen XF, Liu ZL, Lou WJ, Hou YT. NPMML: A framework for non-interactive privacy-preserving multi-party machine

learning. IEEE Trans. on Dependable and Secure Computing, 2021, 18(6): 2969–2982. [doi: 10.1109/TDSC.2020.2971598]

[95] Feng SH, Niyato D, Wang P, Kim DI, Liang YC. Joint service pricing and cooperative relay communication for federated learning. In: Proc. of the 2019 Int'l Conf. on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). Atlanta: IEEE, 2019. 815–820. [doi: 10.1109/iThings/GreenCom/CPSCom/SmartData.2019.00148]

[96] Wang RY, Reddy MP, Kon HB. Toward quality data: An attribute-based approach. Decision Support Systems, 1995, 13(3–4): 349–372. [doi: 10.1016/0167-9236(93)E0050-N]

[97] Bergantiños G, Moreno-Ternero JD. Sharing the revenues from broadcasting sport events. Management Science, 2020, 66(6): 2417–2431. [doi: 10.1287/mnsc.2019.3313]

[98] Kessing SG. Strategic complementarity in the dynamic private provision of a discrete public good. Journal of Public Economic Theory, 2007, 9(4): 699–710. [doi: 10.1111/j.1467-9779.2007.00326.x]

[99] Zhang N, Ma Q, Chen X. Enabling long-term cooperation in cross-silo federated learning: A repeated game perspective. IEEE Trans. on Mobile Computing, 2023, 22(7): 3910–3924. [doi: 10.1109/TMC.2022.3148263]

[100] Bi X, Gupta A, Yang MC. Understanding partnership formation and repeated contributions in federated learning: An analytical investigation. Management Science, 2024, 70(8): 4974–4994. [doi: 10.1287/mnsc.2023.00611]

[101] Ai QY, Zhan ZJ, Wang C, Song J. Incentive mechanism design for sustainable federated learning based on reinforcement learning. Quarterly Journal of Economics and Management, 2024, 3(1): 115–144 (in Chinese with English abstract). [doi: 10.20180/j.qjem.2024.01.004]

[102] Lu RH, Zhang WZ, He H, Li Q, Zhong XX, Yang HW, Wang DS, Shi L, Guo YL, Wang ZJ. Two-stage client selection for federated learning against free-riding attack: A multiarmed bandits and auction-based approach. IEEE Internet of Things Journal, 2024, 11(20): 33773–33787. [doi: 10.1109/JIOT.2024.3431555]

[103] Xu JY, Zhao YZ, Li XW, Zhou L, Zhu KS, Xu XR, Duan Q, Zhang RR. TEG-DI: Dynamic incentive model for federated learning based on tripartite evolutionary game. Neurocomputing, 2025, 621: 129259. [doi: 10.1016/j.neucom.2024.129259]

[104] Kang JW, Xiong ZH, Niyato D, Ye DD, Kim DI, Zhao J. Toward secure blockchain-enabled Internet of vehicles: Optimizing consensus management using reputation and contract theory. IEEE Trans. on Vehicular Technology, 2019, 68(3): 2906–2920. [doi: 10.1109/TVT.2019.2894944]

[105] Liu YN, Li KQ, Jin YW, Zhang Y, Qu WY. A novel reputation computation model based on subjective logic for mobile ad hoc networks. Future Generation Computer Systems, 2011, 27(5): 547–554. [doi: 10.1016/j.future.2010.03.006]

[106] Shi Z. Incentive mechanism design for supply side in data trading [Ph.D. Thesis]. Hefei: University of Science and Technology of China, 2023 (in Chinese with English abstract). [doi: 10.27517/d.cnki.gzkju.2023.000224]

[107] Li B, Lu JF, Cao SQ, Hu LJ, Dai Q, Yang SS, Ye ZW. RATE: Game-theoretic design of sustainable incentive mechanism for federated learning. IEEE Internet of Things Journal, 2025, 12(1): 81–96. [doi: 10.1109/JIOT.2024.3460349]

[108] Han DG, Wooldridge M, Rogers A, Ohrimenko O, Tschiatschek S. Replication robust payoff allocation in submodular cooperative games. IEEE Trans. on Artificial Intelligence, 2023, 4(5): 1114–1128. [doi: 10.1109/TAI.2022.3195686]

[109] Engstrom L, Ilyas A, Santurkar S, Tsipras D, Steinhardt J, Madry A. Identifying statistical bias in dataset replication. In: Proc. of the 37th Int'l Conf. on Machine Learning. JMLR.org, 2020. 274.

[110] Tolpegin V, Truex S, Gursoy ME, Liu L. Data poisoning attacks against federated learning systems. In: Proc. of the 25th European Symp. on Research in Computer Security. Guildford: Springer, 2020. 480–501. [doi: 10.1007/978-3-030-58951-6_24]

[111] Hitaj B, Ateniese G, Perez-Cruz F. Deep models under the GAN: Information leakage from collaborative deep learning. In: Proc. of the 2017 ACM SIGSAC Conf. on Computer and Communications Security. Dallas: ACM, 2017. 603–618. [doi: 10.1145/3133956.3134012]

[112] Sun H, Xiao MJ, Xu Y, Gao GJ, Zhang S. Privacy-preserving stable crowdsensing data trading for unknown market. In: Proc. of the 2023 IEEE Conf. on Computer Communications. New York: IEEE, 2023. 1–10. [doi: 10.1109/INFOCOM53939.2023.10228966]

[113] Tian ZY, Cui L, Liang J, Yu S. A comprehensive survey on poisoning attacks and countermeasures in machine learning. ACM Computing Surveys, 2023, 55(8): 166. [doi: 10.1145/3551636]

[114] Yazdinejad A, Dehghantanha A, Karimipour H, Srivastava G, Parizi RM. A robust privacy-preserving federated learning model against model poisoning attacks. IEEE Trans. on Information Forensics and Security, 2024, 19: 6693–6708. [doi: 10.1109/TIFS.2024.3420126]

[115] Shejwalkar V, Houmansadr A, Kairouz P, Ramage D. Back to the drawing board: A critical evaluation of poisoning attacks on production federated learning. In: Proc. of the 2022 IEEE Symp. on Security and Privacy (SP). San Francisco: IEEE, 2022. 1354–1371. [doi: 10.1109/SP46214.2022.9833647]

[116] Zeng SQ, Huo R, Huang T, Liu J, Wang S, Feng W. Survey of blockchain: Principle, progress and application. Journal on

Communications, 2020, 41(1): 134–151 (in Chinese with English abstract). [doi: 10.11959/j.issn.1000-436x.2020027]

[117] Gafni Y, Tennenholtz M. Long-term data sharing under exclusivity attacks. In: Proc. of the 23rd ACM Conf. on Economics and Computation. Boulder: ACM, 2022. 739–759. [doi: 10.1145/3490486.3538311]

[118] Gao Y, Chen XF, Zhang YY, Wang W, Deng HH, Duan P, Chen PX. A survey of attack and defense techniques for federated learning systems. Chinese Journal of Computers, 2023, 46(9): 1781–1805 (in Chinese with English abstract). [doi: 10.11897/SP.J.1016.2023.01781]

[119] Zhao Y, Li M, Lai LZ, Suda N, Civin D, Chandra V. Federated learning with non-IID data. arXiv:1806.00582, 2018.

[120] Hao WT, El-Khamy M, Lee J, Zhang JY, Liang KJ, Chen CY, Carin L. Towards fair federated learning with zero-shot data augmentation. In: Proc. of the 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops. Nashville: IEEE, 2021. 3305–3314. [doi: 10.1109/CVPRW53098.2021.00369]

[121] Jeong E, Oh S, Kim H, Park J, Bennis M, Kim SL. Communication-efficient on-device machine learning: Federated distillation and augmentation under non-IID private data. arXiv:1811.11479, 2018.

[122] Wang YL, Wolfrath J, Sreekumar N, Kumar D, Chandra A. Accelerated training via device similarity in federated learning. In: Proc. of the 4th Int'l Workshop on Edge Systems, Analytics and Networking. Edinburgh: ACM, 2021. 31–36. [doi: 10.1145/3434770.3459734]

[123] Long GD, Xie M, Shen T, Zhou TY, Wang XZ, Jiang J. Multi-center federated learning: Clients clustering for better personalization. World Wide Web, 2023, 26(1): 481–500. [doi: 10.1007/s11280-022-01046-x]

[124] Arisdakessian S, Wahab OA, Mourad A, Otrok H. Coalitional federated learning: Improving communication and training on non-IID data with selfish clients. IEEE Trans. on Services Computing, 2023, 16(4): 2462–2476. [doi: 10.1109/TSC.2023.3246988]

[125] Zhang N, Ma Q, Mao WX, Chen X. Coalitional FL: Coalition formation and selection in federated learning with heterogeneous data. IEEE Trans. on Mobile Computing, 2024, 23(11): 10494–10508. [doi: 10.1109/TMC.2024.3375325]

[126] Zhan YF, Li P, Qu ZH, Zeng DZ, Guo S. A learning-based incentive mechanism for federated learning. IEEE Internet of Things Journal, 2020, 7(7): 6360–6368. [doi: 10.1109/JIOT.2020.2967772]

[127] Zhang JW, Wu YZ, Pan R. Incentive mechanism for horizontal federated learning based on reputation and reverse auction. In: Proc. of the 2021 Web Conf. Ljubljana: ACM, 2021. 947–956. [doi: 10.1145/3442381.3449888]

[128] Deng YH, Lyu F, Ren J, Chen YC, Yang P, Zhou YZ, Zhang YX. FAIR: Quality-aware federated learning with precise user incentive and model aggregation. In: Proc. of the 2021 IEEE Conf. on Computer Communications. Vancouver: IEEE, 2021. 1–10. [doi: 10.1109/INFOCOM42981.2021.9488743]

[129] Ren ZH, Zhang XL, Ng WWY, Zhang JN. Incentive mechanism design for multi-round federated learning with a single budget. IEEE Trans. on Network Science and Engineering, 2025, 12(1): 198–209. [doi: 10.1109/TNSE.2024.3488719]

[130] Yu H, Liu ZL, Liu Y, Chen TJ, Cong MS, Weng X, Niyato D, Yang Q. A sustainable incentive scheme for federated learning. IEEE Intelligent Systems, 2020, 35(4): 58–69. [doi: 10.1109/MIS.2020.2987774]

[131] Jiang XQ, Diao HY, Zhou CQ, Zhang J. Timeliness-selective incentive federated crowdsourcing. In: Proc. of the 2024 IEEE Int'l Conf. on Web Services. Shenzhen: IEEE, 2024. 632–642. [doi: 10.1109/ICWS62655.2024.00083]

[132] Xuan SC, Wang MD, Zhang JY, Wang W, Man DP, Yang W. An incentive mechanism design for federated learning with multiple task publishers by contract theory approach. Information Sciences, 2024, 664: 120330. [doi: 10.1016/j.ins.2024.120330]

附中文参考文献

[1] 杨明, 冯宏霖, 王鑫, 霍吉东, 焦绪国, 张恒. 数据要素市场研究综述: 价值、定价与交易. 网络与信息安全学报, 2024, 10(3): 1–19. [doi: 10.11959/j.issn.2096-109x.2024036]

[2] 黄科满, 杜小勇. 数据治理价值链模型与数据基础制度分析. 大数据, 2022, 8(4): 3–16. [doi: 10.11959/j.issn.2096-0271.2022062]

[3] 蒋蕊韩. 人工智能嵌入社会治理的风险、诱因及实践策略. 四川轻化工大学学报 (社会科学版), 2024, 39(6): 30–38. [doi: 10.11965/xbew20240604]

[4] 商希雪, 韩海庭, 朱郑州. 基于演化博弈的数据收益权分配机制设计. 计算机科学, 2021, 48(3): 144–150. [doi: 10.11896/jsjkx.201100056]

[14] 刘枘, 郝雪镜, 陈俞宏. 大数据定价方法的国内外研究综述及对比分析. 大数据, 2021, 7(6): 89–102. [doi: 10.11959/j.issn.2096-0271.2021063]

[17] 蔡莉, 黄振弘, 梁宇, 朱扬勇. 数据定价研究综述. 计算机科学与探索, 2021, 15(9): 1595–1606. [doi: 10.3778/j.issn.1673-9418.2103069]

[18] 江东, 袁野, 张小伟, 王国仁. 数据定价与交易研究综述. 软件学报, 2023, 34(3): 1396–1424. http://www.jos.org.cn/1000-9825/6751.

htm [doi: 10.13328/j.cnki.jos.006751]
[50] 王勇, 李国良, 李开宇. 联邦学习贡献评估综述. 软件学报, 2023, 34(3): 1168–1192. http://www.jos.org.cn/1000-9825/6786.htm [doi: 10.13328/j.cnki.jos.006786]
[70] 杜小勇, 陈跃国, 范举, 卢卫. 数据整理——大数据治理的关键技术. 大数据, 2019, 5(3): 13–22. [doi: 10.11959/j.issn.2096-0271.2019020]
[72] 孟小峰, 慈祥. 大数据管理: 概念、技术与挑战. 计算机研究与发展, 2013, 50(1): 146–169. [doi: 10.3969/j.issn.1672-5719.2017.19.112]
[75] 袁满, 刘峰, 曾超, 谢兰. 数据质量维度与框架研究综述. 吉林大学学报 (信息科学版), 2018, 36(4): 444–451. [doi: 10.3969/j.issn.1671-5896.2018.04.014]
[81] 李凤华, 李晖, 贾焰, 俞能海, 翁健. 隐私计算研究范畴及发展趋势. 通信学报, 2016, 37(4): 1–11. [doi: 10.11959/j.issn.1000-436x.2016078]
[101] 艾秋媛, 詹志坚, 王聪, 宋洁. 基于强化学习的可持续联邦学习激励机制设计. 经济管理学刊, 2024, 3(1): 115–144. [doi: 10.20180/j.qjem.2024.01.004]
[106] 史专. 面向数据交易供给侧的激励机制研究 [博士学位论文]. 合肥: 中国科学技术大学, 2023. [doi: 10.27517/d.cnki.gzkju.2023.000224]
[116] 曾诗钦, 霍如, 黄韬, 刘江, 汪硕, 冯伟. 区块链技术研究综述: 原理、进展与应用. 通信学报, 2020, 41(1): 134–151. [doi: 10.11959/j.issn.1000-436x.2020027]
[118] 高莹, 陈晓峰, 张一余, 王玮, 邓煌昊, 段培, 陈培炫. 联邦学习系统攻击与防御技术研究综述. 计算机学报, 2023, 46(9): 1781–1805. [doi: 10.11897/SP.J.1016.2023.01781]

作者简介

何佳妮, 博士生, 主要研究领域为数据市场, 数据治理, 数据价值评估.
黄科满, 博士, 副教授, CCF 专业会员, 主要研究领域为人智交互, 数据治理, 数字安全行为、政策和策略.
刘金飞, 博士, 研究员, 博士生导师, CCF 专业会员, 主要研究领域为数据要素交易, 数据安全与合规.
卢卫, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为数据治理, 数据库基础理论, 大数据系统研制, 云数据库系统和应用.
范举, 博士, 教授, 博士生导师, CCF 专业会员, 主要研究领域为跨域数据治理, 智能数据库系统, 大数据分析.
杜小勇, 博士, 教授, 博士生导师, CCF 会士, 主要研究领域为跨域数据治理, 高性能数据库, 智能信息检索, 非结构化数据管理.