



北京邮电大学

硕士研究生学位论文开题报告

学 号： 2023111430

姓 名： 王雪钦

学 院： 经济管理学院

专业(领域)： 管理科学与工程

研究方向： 信息系统与信息管理

导师姓名： 曹建彤

攻读学位： 工学硕士

2024 年 11 月 25 日

论文题目	基于多维度因素驱动的数据价值评估与定价研究		
选题来源	非立项	论文类型	综合研究
开题日期	2024-12-02	开题地点	经管楼 202
课程名称	学分	成绩	班级排名
自然辩证法概论	1	79	共 (109) 人, 排第 (15) 名
新时代中国特色社会主义理论与实践	2	87	共 (108) 人, 排第 (98) 名
研究生英语学术阅读与写作	2	90	共 (38) 人, 排第 (5) 名
服务运营管理	2	87	共 (4) 人, 排第 (3) 名
数据挖掘与商务智能	2	88	共 (72) 人, 排第 (26) 名
高级计量经济学	3	81	共 (34) 人, 排第 (29) 名
大数据与人工智能	2	94	共 (58) 人, 排第 (50) 名
管理科学研究方法	3	92	共 (88) 人, 排第 (3) 名
研究生学术与职业素养讲座 (MOOC)	2	92	共 (857) 人, 排第 (126) 名
有效沟通技巧 (MOOC)	1	98	共 (354) 人, 排第 (132) 名
信息素养——学术研究的必修课 (MOOC)	2	99	共 (847) 人, 排第 (103) 名
如何写好科研论文 (MOOC)	2	98	共 (2273) 人, 排第 (907) 名
英文科技论文写作与学术报告 (MOOC)	2	99	共 (272) 人, 排第 (181) 名
科研诚信与学术规范 (MOOC)	1	98	共 (2984) 人, 排第 (1701) 名
<p>已修课程是否满足培养方案要求： 是</p> <p>修课计划：</p>			

一、立题依据（包括研究目的、意义、国内外研究现状和发展趋势，需结合科学研究发展趋势来论述科学意义；或结合国民经济和社会发展中迫切需要解决的关键科技问题来论述其应用前景。附主要参考文献目录）（不少于 800 字）

1 研究背景

在当今数字经济的推动下，数据已逐渐成为与传统生产要素同等重要的经济资源。随着信息技术、人工智能等领域的创新发展，数据在各行业中发挥着越来越关键的作用，使得全球数据量迅速增长，国际数据公司(International Data Corporation)的报告指出，预计到 2025 年，全球数据总量将达到 175ZB^[1]，其中我国的数据总量约占 1/4，将成为世界规模最大的数据圈。这种增长不仅加速了数据市场化的需求，还推动了数据在生产、消费、管理等各环节中的价值重塑。数据已不仅是用于存储和传输的信息载体，而是一种具备战略性、可量化和可交易的经济资产^[2]。

1.1 数据要素市场化进程的加速

随着经济数据的发展，数据要素化和市场化逐渐成为中国特色社会主义市场经济发展的必然要求，国家在政策层面高度重视数据要素市场的构建与发展。2020 年，中共中央、国务院颁布的《关于构建更加完善的要素市场化配置体制机制的意见》首次将数据明确纳入要素市场体系，标志着我国在数据市场化发展上迈出了关键一步^[3]。2022 年发布的《关于构建数据基础制度更好发挥数据要素作用的意见》（即《数据二十条》）明确了数据在经济中的战略价值，提出要构建完善的数据市场体系，以支持数据要素在各领域的流通与应用^[4]。2024 年 1 月，国家数据局等 17 部门近日联合印发《“数据要素×”三年行动计划（2024—2026 年）》，进一步强调数据作为关键生产要素的价值日益凸显，应发挥我国超大规模市场、海量数据资源、丰富应用场景等多重优势，促进数据要素价值倍增^[5]。

在我国数据要素市场发展迅速的背景下，各地陆续建立了多个数据交易平台和机构。截至 2024 年 3 月底，全国共计成立 49 家数据交易所，为数据的跨行业流通和产业链的资源整合提供了基础设施，为数据产品的生成和流通提供了平台，推动了数字经济的蓬勃发展。数据交易平台作为连接数据需求方与供应方的重要枢纽，推动了数据在有交易意愿的双方之间流通。通过数据交易的实现，不仅打破了“数据孤岛”现象，还有效激活了海量数据资源，进一步挖掘了各种数据的潜在价值^[6]。数据估值与定价机制作为数据要素市场交易的核心基础，在这一过程中扮演着至关重要的角色。数据价值反映了数据产品在市场中的实际或潜在贡献，对其市场吸引力起到直接作用^[7]。数据定价在充分考虑数据价值的前提下，通过科学合理的定价机制，确保了数据产品的市场定位和竞争优势。基于此，数据交易通过买卖双方的协商实现数据产品的流通和交换。数据价值、数据定价与数据交易三者之间的协调互动，为数据要素市场的稳健发展和高效运行提供了关键支持。

1.2 数据估值和定价机制的复杂性与挑战

数据的价值并非单一维度所能涵盖。数据作为数字资产，其特性决定了传统资产评估方法难以直接应用到数据市场中。首先，数据具有复制成本低、存储形式复杂等特点，导致其在市场中的价值波动较大^[9]。其次，数据的价值受多种因素影响，包括数

据的质量、完整性、容量、时效性、隐私保护等，而这些因素在不同应用场景中可能产生不同的价值体现^[10]。同时，数据的买卖双方对数据价值的衡量存在“双边不确定性”^[11]，即买方和卖方对于数据的具体价值无法达成共识，导致交易过程中信息不对称。数据价值评估不仅要考虑数据的量化属性，还需要评估其潜在的创新价值和社会价值^[12]，通过客观、全面的评估，才能为数据的交易提供可靠依据。因此，如何对数据的多维度特性进行综合评估，准确把握其价值，成为当前数据交易中亟待解决的问题。

数据的定价应当反映数据价值^[13]，需要综合数据的特性、价值属性、市场需求等多个维度，遵循真实性、公平性、利益最大化等基本原则^[14]。此外，定价机制还应具有多版本定价的能力，即在满足不同消费者需求的前提下，提供多种不同质量或属性的数据版本供选择。这种多版本定价模式不仅能满足市场上不同层次和需求的消费者，还能通过优化价格差异化，增加市场渗透率，提升数据资源的总体交易量^[15]。因此，数据定价不仅需要保障交易各方的基本权益，还要通过科学的定价方式，合理反映数据价值，实现多版本、公平、高效定价，进一步促进数据市场的高效运行和可持续发展，并使数据要素的市场价格更加接近其真实经济价值，从而提升数据资源的整体效用。

然而，当前我国数据交易市场尚未建立权威的、普适的数据估值和定价方法，现有平台普遍采用按次计费或协商议价等简单定价方式。这两种方法存在严重的信息不对称，导致交易双方容易产生误导，进而影响数据要素市场的正常运作^[16]。因此，研究并建立一套通用的、科学的数据估值与定价体系，对推动数据要素市场的健康发展具有重要意义。

1.3 研究数据价值评估与定价策略的必要性

随着我国数据要素市场的快速发展和政策的持续推动，建立科学合理的数据价值评估和定价机制显得尤为迫切。为了促进数据市场的健康发展，需要针对数据的特性，制定更加完善的定价机制，以提高数据交易的透明度和市场效率。

当前许多数据价值评估方法仍然过于单一，如数据的规模或数据质量的某一维度，缺乏对数据多维度属性的全面考虑，导致现有数据估值方法无法全面反映数据的真实经济价值，也未能有效推动市场机制的成熟与完善^[17]。此外，单一因素的定价模式往往无法适应数据市场复杂多变的需求。在不同的应用场景中，消费者对数据的需求和价值认知各不相同，面对消费者的差异化需求，数据市场需要引入多版本定价策略，满足不同层次、不同类型的市场需求。基于此，数据供应商可以优化价格差异化，增加市场渗透率，提升数据交易的整体效益，保障数据消费者的选择权和公平性^[18]。同时，现有的定价方法应当充分考虑通过合理定价实现市场利益的最大化。定价策略要考虑到市场需求、数据质量、竞争环境等多方面因素的综合影响，具备动态调整能力^[19]，以应对市场的变化。通过灵活的定价策略，数据交易可以最大化各方的利益，推动数据市场的繁荣与创新。

综上所述，构建科学合理的数据价值评估体系和定价机制，是推动数据要素市场化、促进数字经济可持续发展的必然要求。当前数据价值评估方法未能全面考虑数据的多维价值特征，且缺乏有效的定价机制来调动市场潜力。因此，如何设计一个科学的数据价值评估体系，并在此基础上提出有效的定价模型，已成为推动数据要素市场化和规范化发展的关键。

2 研究目的与研究意义

2.1 研究目的

基于上述研究现状，本研究旨在提出一种基于多维度因素的数据价值评估与定价方法，为数据要素的交易流通提供更为科学的价值评估体系。该方法从多维度因素出发，综合考虑数据容量、数据质量、隐私保护、预测准确性等多个维度量化数据价值，结合消费者支付意愿构建科学的多版本定价模型，以填补，游效解决定价单一、需求匹配不足等问题，从而提升数据提供者的利润空间，优化数据交易的公平性和效率。目标可以细分为解决以下三个核心问题：

(1) 多维度数据价值评估方法：

针对当前数据价值评估方法的单一性问题，本研究将基于多维度因素量化数据价值，综合考虑数据容量、数据质量、隐私保护和预测准确性等多个维度。通过引入多因素加权计算，建立多维数据效用函数模型，以全面反映数据的真实价值，从而为数据定价研究提供一个系统的评估基础。这种方法不仅可以为数据交易中的定价决策提供可靠依据，还将为数据提供者和消费者之间的价值评估提供科学依据。

(2) 多版本数据定价策略：

在实际的数据交易市场中，不同的数据消费者对数据的需求差异明显。因此，本研究将针对不同需求，提出多版本数据定价策略，数据供应商可以通过降低低版本数据的价格，吸引更多消费者，而高版本数据则可以定价较高，获得更多的利润。该策略将通过不同版本的数据效用差异实现定价的灵活性，以满足消费者多样化需求，从而提高数据市场的交易效率和数据提供者的利润空间。

(3) 基于消费者支付意愿的最优定价策略：

在数据交易过程中，消费者的支付意愿通常随着数据价值的变化而呈现非线性波动。为了模拟这一现实情况，本研究将消费者支付意愿作为关键因素引入到数据定价模型中。通过构建与消费者支付意愿相关的多维数据价值函数，模型能够动态地调整数据价格，旨在探究在何种情况下，能够使得数据提供者获取的利润最大化。此外，本研究将使用遗传算法进行求解，以应对多约束条件下的非线性问题，确保在复杂的市场环境中实现全局最优的定价策略。

2.2 研究意义

本研究从数据提供者与消费者两大市场主体的实际需求出发，围绕数据价值评估与科学定价策略展开深入探讨，以下为具体的理论和现实意义。

(1) 理论意义层面：

本研究对数据定价策略的研究进行了多方面拓展。①通过引入多维度因素，如数据容量、质量、隐私保护和预测准确性，丰富了现有的数据价值量化模型。这种多维评估方法为数据价值量化提供了一个全面而系统的理论框架，突破了传统单一维度评估的局限，为研究数据的多层次价值提供了更科学的理论依据。②多版本数据定价策略从差异化定价的角度填补了现有理论中对不同需求市场适应性分析的空白。该策略结合消费者支付意愿的动态变化，通过建立非线性定价模型，使数据定价在应对市场需求时更加灵活，进一步拓展了数据经济学的定价策略理论。③本研究在模型求解中应用了遗传算法，以应对数据交易中多因素、非线性及多约束等复杂性。这一优化方法的应用，不仅丰富了遗传算法在数据经济学中的应用场景，也为处理类似复杂经济问

题提供了新的方法论支持。

(2) 现实意义层面：

本研究对数据交易市场的健康发展具有实际推动作用。①研究提出的多维数据价值量化方法为数据交易平台提供了科学合理的定价依据，能够有效减少因数据价值低估或高估导致的市场定价偏差，提升交易过程的透明度和合理性，从而促进数据市场的公正与规范发展。②多版本数据定价策略在满足不同消费者需求的基础上，推动了数据服务的差异化和市场覆盖率的提高。通过低价位的低版本数据吸引更多消费者的同时，数据提供者可通过高质量版本数据获得更高的收益，实现利润最大化与市场覆盖的双重目标，在优化利润的同时保障消费者的满意度，进而增强市场竞争力。③本研究成果对政府和监管机构在数据市场中的政策制定也具有参考意义，能够为构建科学、公正的数据交易规范提供理论支持，推动数据要素市场化的规范化发展，从而促进数据资源的高效流通与配置，为数字经济的可持续发展提供强有力的支撑。

3 文献综述

随着数据流动和共享日益普及，数据交易的复杂性也逐渐显现。目前，包括贵阳大数据交易所在内的国内主要数据交易平台、数据资源企业及互联网企业，基本上都采取协商定价或订阅模式进行数据交易^[20]。在这个领域，现有的研究大多集中在数据权属的确认上，然而当前主流的数据显示交易方式仍缺乏统一的价值评估与定价标准，这一问题仍处于探索阶段。国内外学者从多角度对数据估值方法和定价策略进行了深入探讨，提出了适应不同交易需求的数据定价方法，为数据估值与定价的理论基础和实践奠定了重要基础。

3.1 传统会计学中的数据资产价值评估

部分国内研究借鉴了传统的无形资产评估方法来对数据进行估值。中国资产评估协会发布的《资产评估专家指引第 9 号——数据资产评估》将数据资产的价值评估方法概括为市场法、成本法和收益法三种基本方法，以及其衍生方法^[21]。李永红等人结合市场法，利用层次分析法和灰色关联分析法确定了估值因素的权重与可比资产，构建了数据资产价值评估模型^[22]。邹贵林等人基于两阶段修正成本法，通过比较成本价格与市场价格，界定了数字资产的价格区间，并据此建立了数据资产的定价方法^[23]。陈芳等人在多期超额收益法的基础上，综合考虑了数据资产的权属不明确、数据安全等特殊风险因素，提出了基于剩余法的多期超额收益模型，用以评估数字化转型企业的数据资产价值^[24]。这些研究方法能够较为客观地反映市场的基本情况，但由于缺乏成熟的数据交易市场，并未充分考虑到数据未来可能的增值、市场竞争以及消费者需求等外部因素。

3.2 基于数据要素本身的价值评估

为了更加准确地评估数据的价值，一些学者从数据本身的特性出发，尝试对其内在价值进行衡量。当前，数据质量、隐私保护与数据效用等方面的因素被广泛认定为数据价值的核心指标。在这些因素中，数据质量被视为数据的内在属性，是决定其价值的关键要素。Heckman 等人提出了一系列与数据价值相关的因素，包括数据的成本、年份、周期性、准确性和数量等，并通过为这些因素赋予不同权重，构建了一个线性估值模型^[25]。Pipino 则提出了三种衡量数据质量的函数形式：比率、最大或最小算子

以及加权平均,并设计了一种结合主客观评价的综合方法来评估数据质量^[26]。Wang 和 Strong 对数据的各个维度进行了分类和调查,构建了一个层次化框架,确定了 15 个与数据质量相关的维度^[27]。通过衡量数据集的信息量,并结合定价函数来将数据的信息量转换为合理的价格,尽管这一领域已有了许多有益的探索,但关于数据质量维度的选取仍未有统一的标准,因此在实际定价时,仍需要根据具体数据的特点做出合理判断。

数据隐私主要涉及到个人数据的保护,可以通过信息熵来进行衡量。Shen 等人基于数据的信息熵提出了无套利定价模型,专门应用于个人数据^[28]。Li 等人进一步提出了三种基于信息熵的定价函数,充分证明了信息熵在数据定价中的合理性^[29]。彭慧波等人则结合信息熵,提出了一种分级的隐私度量方法,进一步构建了基于信息熵的数据定价模型^[30]。信息熵的引入标志着数据定价从策略性定价向数学模型定价的转变,推动了定价方法的科学性与合理性。然而,基于信息熵的定价模型更加关注数据的数量与分布,尚未有效体现数据的质量和其实际价值。

除了信息熵外,一些研究开始使用效用理论对数据的评估因素进行量化。刘枏等人提出了基于效用的定价方法,综合考虑了数据容量和数据质量两个方面的价值^[31]。林娟娟等人首次提出了一个综合得分的概念,涵盖了数据质量、数据数量以及它们之间的相互作用,并基于此构建了一个效用函数^[32]。Yang 等人则基于消费者的效用敏感度的异质性,首次构建了一个多维度的定价模型,并将消费者分为三类,适应了不同消费者的需求^[33]。效用理论有助于综合考虑多重评估因素并量化数据的价值,但需要充分考虑市场的多样性,效用函数的选择也会因此而有所不同,使用单一因素很难实现客观、全面的评估。

3.3 数据定价方法

数据的流通和利用依赖于有效的数据定价机制。数据价值的评估为数据所有者提供了一个参考价格,而在此基础上,必须建立一个科学合理的定价体系,才能将数据的战略、经济和市场价值转化为具体、可量化的指标。目前,数据定价方法的研究可大致分为以下几类:基于查询的定价、拍卖定价、元组定价、博弈论定价和机器学习定价。

查询操作作为数据库的基本功能之一,具备计算和条件检索的能力,用户可以通过设定查询条件从视图中选择需要的数据。Koutris 等人提出了一种基于查询的定价模型,旨在实现无套利和无折扣原则,同时改进了买家只能从一组特定视图选择数据的情况^[34]。Chawla 等人进一步研究了在查询定价环境下如何最大化收益,并将这一问题转化为单一买家与无限供应情况下的捆绑定价问题^[35]。然而,查询技术主要作为一种底层技术存在,衍生出的定价模型通常需要在查询的基础上结合其他技术。

为了应对数据价值的不确定性,拍卖机制可使交易价格在买卖双方之间更具公平性。An 等人提出了一种多轮假名证明拍卖方案(MFPA),其中数据交易平台根据数据卖方与买方的出价来操控市场,从而实现双方效用的最大化^[36]。郭鑫鑫等人设计了一种迭代双边拍卖方法,用于健康数据交易的分配、定价及奖励规则,并鼓励买方提出合理的价格^[37]。Xiong 首次提出了基于智能合约的反串谋拍卖机制,使得不信任的买卖双方能够在没有可信第三方的情况下安全地参与数据拍卖,确保交易的公平性和真实性^[38]。虽然拍卖机制能够促进公平竞争,但由于数据并非稀缺资源,其在实际应用中仍较为有限。

元组定价技术能够精确量化数据库中的每一条信息,更好地识别数据的价值并为

其匹配合理的价格。Balazinska 等人提出了一种基于元组的细粒度定价方案，能够客观计算导出数据的价值，消费者通过发出 SQL 查询来读取数据并为查询结果付费^[39]。Tang 等人通过数据溯源技术，找出了与查询结果对应的最小元组集合，并将其作为数据定价的依据^[40]。Shen 等人构建了一个基于元组粒度的个人大数据正向分级与反向定价模型，该模型能够动态调整^[28]。这种基于元组的定价方法相较于信息熵的复杂计算方式，能够更有效地利用数据资源，并适合于结构化数据的定价。

在数据定价中，博弈论模型可用于分析卖方和买方之间的策略互动，从而设计出最优定价策略。Shen 等人假设数据交易市场是一个完全信息的定价市场，其中每个参与者都拥有所有市场信息，并基于此建立了一个两阶段 Stackelberg 博弈模型，以模拟数据提供商、服务提供商与买方之间的互动^[41]。Chen 等人提出了一种包括数据量、可用性和完整性在内的多维数据效用评估方法，构建了一个基于数据效用的三阶段 Stackelberg 博弈模型，并提出最优数据定价方案^[42]。张小伟等人基于这一理论框架，提出了一个三阶段博弈模型并结合 Shapley 值来评估各方效用^[43]。基于博弈论的数据定价模型主要适用于市场类型明确，供给需求规模较大的数据交易市场，该方法更偏向于交易分析，还需结合其他方法明确数据价值。

随着机器学习的不断发展，越来越多的学者尝试将其应用于数据定价优化领域。Tsai 等人将机器学习应用于时间相关智能数据定价（TDP），通过预测模型优化定价策略，从而提高数据定价的效率和准确性^[44]。Chen 等人则基于机器学习算法设计了数据定价模型，提出了模型的准确性直接影响定价结果的结论^[45]。Fan 等人则借助联邦学习提出了一种新的数据评估方法，结合联邦 Shapley 值保护数据隐私，并激励数据交易^[46]。虽然机器学习方法在理论上具有较高的潜力，但由于数据交易市场尚未完全成熟，且缺乏足够的历史数据用于模型训练，因此其定价模型在现实市场中仍存在较大的不确定性和挑战。

参考文献

- [1] IDC. (2021). The Data Age 2025: The Evolution of Data to Life-Critical. International Data Corporation.
- [2] Wen, F., & Wang, Y. (2021). The Development and Future of Data Element Market in China: Status, Challenges, and Prospects. *Journal of Digital Economy*, 1(2), 45-60.
- [3] 中共中央国务院. 关于构建更加完善的要素市场化配置体制机制的意见[EB/OL]. (2020-04-09). https://www.gov.cn/zhengce/2020-04/09/content_5500622.htm.
- [4] 中共中央国务院. 关于构建数据基础制度更好发挥数据要素作用的意见[EB/OL]. (2022-12-02). https://www.gov.cn/gongbao/content/2023/content_5736707.htm.
- [5] 中共中央国务院. “数据要素×”三年行动计划（2024—2026 年）[EB/OL]. (2024-01-04). https://www.gov.cn/lianbo/bumen/202401/content_6924380.htm.
- [6] 蔡继明, 刘媛, 高宏 & 陈臣. (2022). 数据要素参与价值创造的途径——基于广义价值论的一般均衡分析. *管理世界* (07), 108-121. doi:10.19744/j.cnki.11-1235/f.2022.0096.
- [7] Chen, H., Chiang, R. H., & Storey, V. C. (2012). *Business Intelligence and Analytics: From Big Data to Big Impact*. *MIS Quarterly*, 36(4), 1165-1188.
- [8] Varian, H. R. (2009). *Computer Mediated Transactions*. *American Economic Review*, 99(2), 1-10.
- [9] Varian, H. R. (2009). *Computer Mediated Transactions*. *American Economic Review*, 99(2), 1-10.
- [10] Kwon, O., Lee, N., & Shin, B. (2014). *Data quality management, data usage experience and acquisition intention of big data analytics*. *International Journal of Information Management*, 34(3), 387-394.

- [11] 刘朝阳. (2016). 大数据定价问题分析. *图书情报知识*, (1), 57-64.
- [12] Darwish, T. S. J., Abu Bakar, K., Kaiwartya, O., & others. (2020). TRADING: Traffic aware data offloading for big data enabled intelligent transportation system. *IEEE Transactions on Vehicular Technology*, 69(7), 6869-6879. <https://doi.org/10.1109/TVT.2020.2970042>
- [13] Balazinska, M., Howe, B., Koutris, P., & others. (2013). A discussion on pricing relational data. In *Search of Elegance in the Theory and Practice of Computation: Essays Dedicated to Peter Buneman* (pp. 167-173). Springer. https://doi.org/10.1007/978-3-642-32477-0_13
- [14] 王标, 刘兴洋, 许卡, & 等. (2020). 政府开放数据的国家安全风险评估模型研究. *网络与信息安全学报*, 6(6), 80-87.
- [15] Chen, L., & Liu, D. (2017). *On the Optimal Versioning Strategy for Digital Goods under Competition*. *Information Systems Research*, 28(3), 673-691.
- [16] 蔡莉, 黄振弘, 梁宇 & 朱扬勇. (2021). 数据定价研究综述. *计算机科学与探索*(09), 1595-16
- [17] 国家信息中心. (2023). 数据要素市场化配置的政策演进与市场机制研究. *数字经济研究报告*.
- [18] Jiang, H., & Luo, X. (2022). Market demand-oriented data pricing models and application. *Data Market and Digital Transformation*, 27(1), 33-52.
- [19] Gopal, R. D., & Gao, G. (2019). Pricing of data products and monetizing data markets. *Journal of Management Information Systems*, 36(2), 625-654.
- [20] 邓鑫, 冯育宁, 黎贤昶 & 张世奇. (2024). 数据要素定价方法: 研究评述与展望. *中国经济学*(03), 284-320+338-339.
- [21] 中国资产评估协会印发《资产评估专家指引第9号——数据资产评估》[EB/OL]. [2022-10-13]. <http://sdicpa.org.cn/news/2020110/n27342725.html>.
- [22] 李永红 & 张淑雯. (2018). 数据资产价值评估模型构建. *财会月刊*(09), 30-35. doi:10.19641/j.cnki.42-1290/f.2018.09.005.
- [23] 邹贵林, 陈雯, 吴良峥, 张继钢, 梁燕妮 & 余泽远. (2022). 电网数据资产定价方法研究——基于两阶段修正成本法的分析. *价格理论与实践*(03), 89-93+204. doi:10.19851/j.cnki.cn11-1010/f.2022.03.048.
- [24] 陈芳 & 余谦. (2021). 数据资产价值评估模型构建——基于多期超额收益法. *财会月刊*(23), 21-27. doi:10.19641/j.cnki.42-1290/f.2021.23.003.
- [25] HECKMAN J R, BOEHMER E L, PETERS E H, et al. A pricing model for data markets [C] //Proceedings of the iConference, 2015.
- [26] Pipino L L, Lee Y W, Wang R Y. Data quality assessment[J]. *Communications of the ACM*, 2002, 45(4): 211-218.
- [27] Wang R Y, Strong D M. Beyond accuracy: What data quality means to data consumers [J]. *Journal of management information systems*, 1996, 12(4): 5-33.
- [28] Shen Y.C., Guo B., Shen Y., et al. 2016. "A Pricing Model for Big Personal Data." *Tsinghua Science and Technology* 21(5):482-490.
- [29] Li Y., Huang J., Qin S., et al. 2017. "Big Data Model of Security Sharing Based on Blockchain." 2017 3rd International Conference on Big Data Computing and Communications
- [30] 彭慧波、周亚建, 2019, 《基于隐私度量的数据定价模型》, 《软件》第1期
- [31] 刘枏, 徐程程, 陈俞宏. 基于效用理论的数据定价方法研究[J]. *价格理论与实践*, 2022, (11):164-167+211. DOI:10.19851/j.cnki.CN11-1010/F.2022.11.276.
- [32] 林娟娟、黄志刚、唐勇, 2023, 《数据质量、数量与数据资产定价: 基于消费者异质性视角》, 《中国管理科学》, 网络首发文献, 2023年3月10日。
- [33] Yang J., Zhao C., Xing C. 2019. "Big Data Market Optimization Pricing Model Based on Data Quality." *Complexity* 10:1-5964068.

- [34] Koutris P., Upadhyaya P., Balazinska M., et al. 2015. "Query-based Data Pricing." *Journal of the ACM* 62 (5):1-44.
- [35] Chawla S., Deep S., Koutris P., et al. 2019. "Revenue Maximization for Query Pricing." *arXiv preprint arXiv:1909.00845*.
- [36] An D., Yang Q. Y., Yu W., et al. 2017. "Towards Truthful Auction for Big Data Trading." *IEEE 36th International Performance Computing and Communications Conference*.
- [37] 郭鑫鑫、王海燕、孔楠, 2022, 《信息不对称下个人健康数据交易双边定价策略研究》, 《管理工程学报》第4期
- [38] Xiong W., Xiong L. 2020. "Anti-collusion Data Auction Mechanism Based on Smart Contract." *Information Science* 555: 386-409.
- [39] Balazinska M., Howe B., Dan S. 2011. "Data Markets in the Cloud: An Opportunity for the Database Community." *Proceedings of the VLDB Endowment* 4:1482-1485.
- [40] Tang R., Shao D., Bressan S., et al. 2013b. "What You Pay for is What You Get." *Database on Expert Systems Applications: 24th International Conference*.
- [41] Shen B., Shen Y., Ji W. 2019. "Profit Optimization in Service-Oriented Data Market: A Stackelberg Game Approach." *Future Generation Computer Systems* 95:17-25.
- [42] Chen Y., Bai R., Wu Y. T., et al. 2023. "A Multidimensional Data Utility Evaluation and Pricing Scheme in the Big Data Market." *Wireless Communications and Mobile Computing*.
- [43] 张小伟、江东、袁野等, 2023, 《Mass: 基于单位数据贡献的模型定价框架》, 《计算机科学与探索》第9期
- [44] Tsai Y. C., Cheng Y. D., Wu C. W., et al. 2017. "Time-dependent Smart Data Pricing Based on Machine Learning." *Canadian Conference on Artificial Intelligence*.
- [45] Chen L., Koutris P., Kumar A. 2019. "Towards Model-based Pricing for Machine Learning in a Data Marketplace." *The 2019 International Conference*.
- [46] Fan Z. N., Fang H., Zhou Z., et al. 2022. "Improving Fairness for Data Valuation in Horizontal Federated Learning." *IEEE 38th International Conference on Data Engineering*.

二、研究内容和目标（说明课题的具体研究内容，研究目标和效果，以及拟解决的关键科学问题。此部分为重点阐述内容）（不少于 2500 字）

1 研究内容

基于上述研究目的，本研究将从数据价值评估和数据定价方法两方面出发，针对现有数据估值方法考虑因素单一、定价策略无法满足市场多样化等问题，提出一种基于数据质量、数据容量、数据隐私、机器学习预测效果等多维度因素评估数据价值的方法，同时考虑数据提供方的多版本数据产品策略和买方的非线性消费意愿，构建相应的定价模型，实现数据交易利润的最大化。研究内容将从以下三个部分展开：数据价值的衡量与量化、数据定价模型的构建、模型求解与数据验证。

1.1 数据价值的衡量与量化

数据作为一种重要的资产，其价值并非简单的价格标签能够覆盖，数据的价值评估不仅是对其内在特性的衡量，更需要结合其应用场景和对特定任务的贡献度。本研究希望数据价格可以体现数据的本身属性，让价格与数据集的价值形成密切的联系，同时能够从多角度全面反映数据的真实价值，从而为数据定价研究提供一个系统的评估基础。针对当前数据价值评估方法的单一性问题，本研究将基于多维度因素量化数据价值，综合考虑以下几个维度对数据进行综合评估：

（1）数据容量（Volume）

数据容量通常指的是数据的存储大小，通常用字节（如 GB、TB、PB）来衡量。数据容量的大小可以反映数据集的覆盖范围和丰富程度。数据量的增长与数据的潜在价值有着密切关系，尤其是在机器学习和人工智能技术驱动下，数据量的增加往往能够提升模型的效果。大规模的数据集可能会包含更多的信息，意味着它有更高的潜在价值，能够提供更精确的分析、更全面的见解以及更具预测性的能力，是衡量数据价值的基础指标之一。

（2）数据质量（Quality）

数据质量涉及数据的准确性、完整性、一致性、可靠性等多个维度，高质量的数据能够提供更精准的分析结果和预测，对于数据消费者来说，其市场价值远高于低质量数据。结合相关学者对数据质量的研究，本研究将选取最具代表性的三项维度，准确性、完整性以及冗余性，对机器学习中的数据集进行数据质量的衡量。其中准确性是数据的基础特性，其根据数据源的域和数据的信息类型，表示数据源中具有正确值的单元格比例。完整性表示数据集中完整单元格的比例，即单元格不为空，并且单元格中的值有意义。冗余性是指数据中重复记录的数据比例。任何的数据都会在这三项维度上形成差异，因此，这三项维度也是最为通用的，适用于衡量大多数的数据质量。

（3）数据隐私（Privacy）

随着隐私保护法规的逐渐完善（如 GDPR 等），数据隐私性成为数据定价的重要考量因素之一。隐私性较高的数据通常具有较高的市场价值，尤其在涉及敏感信息的领域（如医疗、金融等）更为显著。隐私性评估不仅关注数据的匿名化、加密等技术手段，还需要考虑数据在交易过程中的合规性和数据消费者的隐私保护需求。本研究将基于现有研究中普遍使用的信息熵和隐私等级度量数据的隐私。其中，信息熵可以度量信息，表示数据信息的不确定程度，数据集集中的隐私含量随熵值的增大而呈单调递

增趋势,。但是隐私的价值仅靠 信息熵是远远不充分的,在本文研究方法中,利用隐私分级理论来修正信息熵计算所得结果,并将其作为信息主体隐私含量的一部分。

(4) 机器学习训练效果 (ML Effectiveness)

机器学习的训练效果是评估数据价值的重要因素之一,高效的机器学习模型能够从价值更高的数据中提取深层次的模式,生成更加精准的预测和分析结果,并直接关系到数据能否被有效利用,从而提供实际的商业或决策价值,因此数据的价值往往与其能够产生的预测效果紧密相关。数据本身通常处于未经加工的原始状态,难以直接产生有价值的信息,而其本身的价值并非一成不变,它通常取决于其能否在实际应用中提供有效的信息和做出可靠的预测。机器学习通过算法模型的训练,通过训练分类模型、回归模型等,机器学习能够从大量原始数据中识别出有意义的模式和趋势,使得这些数据具备更高的预测价值,能够揭示数据中潜藏的非显性信息。当一个数据集被用于训练机器学习模型时,模型的预测能力决定了这些数据能否为决策提供足够的信息支持,即模型的准确度可以体现数据的贡献程度,如果模型的预测结果不准确,那么这些数据的实际效用就大打折扣,数据的价值也会显得较低,而贡献度越高的数据可能会显著提高模型的预测准确性,从而使得这些数据的市场价值提升。为量化这一维度,本研究将通过训练模型来评估数据集对模型准确性的提升贡献,并根据数据对特定应用任务的效果来评估其在机器学习中的实际价值。

为了全面评估数据的价值,单一因素的衡量往往不足以反映数据的整体效益。前述的评估因素,如数据容量、数据质量、数据隐私和机器学习预测效果,虽各自揭示了数据的不同特性,但缺乏一个统一的综合框架来全面体现数据的多维价值。因此,本研究引入数据效用 (Data Utility) 理论,将上述多个因素进行量化和整合,旨在通过综合考量各个维度,提供一个更加全面、准确的数据价值评估模型,帮助深入理解数据的潜在价值,为数据的定价与交易提供依据。

效用在经济学中,可以通俗描述为商品满足人们需求的能力。效用是一种基于偏好的价值度量方式,可以量化商品之间的价值关系,使得各类商品在某种情景下可以在一个统一的尺度中进行比较。效用可以通过基数效用论和序数效用论进行度量。基数效用论认为效用是可以量化的,其采用边际效用分析法,含义为消费者对于某种商品,每增加一单位商品所增加的满足程度。而边际效用分析法认为商品随着消费量的变化,其对消费者造成的边际效用是递减的。当商品数量无限时,即使商品具有边际效用递减规律,消费者为了其效用最大化仍然可以选择无限消费商品。但现实社会中商品资源往往是有限的,在不确定因素以及风险下,为了效用最大化,消费者一定是通过将有限的资源在需求中进行合理的分配。而序数效用论是让消费者对商品的效用进行排序,主要考虑消费者对于不同商品的偏好程度比较,不确定效用的准确值。序数效用论的成立需要满足以下条件:①完备性:即消费者需要对相同价格的不同商品偏好度有排序;②可传递性:消费者对商品的偏好有传递性,比如消费者对同价格的三种商品 A、B、C 的偏好为 $A > B$, $B > C$,那么消费者商品 A 的偏好 $>$ 商品 C;③不充分满足性:消费者对于商品数量的需求无上限。当数据作为一种商品时,数据对数据买家的有用程度即称之为数据的效用。由于数据的价值需要量化,同时不存在进行偏好比较的情况,根据对两种不同的度量方法的比较,本研究将采用基数效用论量化数据的效用。

Niyato 等 (2016) 首次提出基于机器学习的数据库效用函数,以数据库效用函数衡量数据库价值,在其构建的模型中,用于训练的数据库数量越大、质量越好,模型的准确度越高,

数据的效用就越高。在本研究中，数据效用将作为数据价值的综合衡量标准，不只局限于体现单一的价值影响因素，将基于前述的多个维度（数据容量、数据质量、数据隐私、机器学习训练效果）进行量化与整合，以形成一个综合的数据效用评估框架。

1.2 数据定价模型的构建

前述对于数据价值的量化，为数据定价提供了合理依据，而在数据市场中，数据的定价不仅仅是对其价值的直接反映，还需要考虑市场需求、消费者的支付意愿以及不同版本数据的定价策略。为更好地反映数据市场的供需关系，本研究提出一种基于消费者效用函数和数据多版本策略的定价模型。

（1）数据要素市场结构

为构建更加合理的数据定价模型，首先需要明确数据要素市场结构。在竞争激烈的市场环境中，定价策略相对简单，数据提供者往往仅需通过定价接近于成本来实现市场份额的扩大和销售的最大化。而在垄断市场环境下，由于竞争对手较少，数据提供者有较大的定价空间，可以通过更精细的定价策略实现利润的最大化，并提高市场覆盖率。因此，本研究将假设数据要素市场为垄断市场，在此环境下讨论定价策略。

数据要素市场中的交易主体主要包括数据提供者（卖方）和数据消费者（买方）。数据提供者通常是数据的持有者或产生者，是数据市场中数据的来源，提供者将拥有的数据提供给数据交易平台，按照规定定价策略完成定价后出售给数据消费者。而数据消费者则根据需求找到合适的数据，通过支付相应的金额获取数据的所有权。提供者希望通过定价策略获得尽可能高的利润，而数据消费者则希望根据数据的效用获得最佳的性价比。

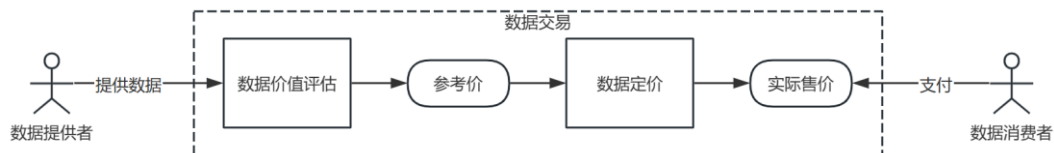


图 2-1 数据市场结构

在这一市场环境下，数据提供者面临的核心问题是如何通过合理的定价策略实现利润最大化，并扩大市场份额。这要求数据提供者在考虑数据的实际售价时，要综合评估数据的市场需求、消费者支付意愿以及数据本身的效用特征。与此同时，数据消费者则希望能够根据数据的效用，支付合理的价格以获取最大价值。因此，数据的定价不仅仅是对其市场价值的简单反映，还需要考虑消费者的支付意愿、数据版本化策略等多方面的因素，以确保定价策略能够在满足市场需求的同时，最大化数据提供者的收益。

（2）非线性消费者支付意愿函数

消费者行为理论在研究消费者购买时的决策过程中，强调消费者对产品质量的感知和偏好，消费者对不同质量水平的产品有不同的支付意愿。因此本研究设定消费者为对数据效用敏感的群体，且每位消费者对数据的敏感程度不同，对于同一份数据，购买的意愿不相同，在多个数据效用不同的数据商品存在的条件下，消费者会选择消费意愿最大的数据产品。而在实际市场中，消费者的支付意愿并不是简单的线性函数。预期效用理论认为，消费者在决策时会基于对不同结果的效用预期进行权衡。当产品的实际价值未达到消费者的预期时，消费者会更倾向于选择价值分数较高的版本

以期望获得更高的效用。然而，一旦产品的价值分数显著超出消费者的预期，消费者的支付意愿增速可能会减缓，因为他们认为额外的价值所带来的效用已经不再显著，他们最终可能会选择价格更低的产品质量版本。因此，消费者的支付意愿应随着数据效用的提升呈现非线性变化。

消费者选择满足支付意愿的数字产品，其满意程度同样可以采用前述中的基数效用理论进行量化。消费者效用可以理解为消费者在面对不同商品或服务时，根据其对产品效用的感知和偏好，所获得的主观满足度。在数据市场中，消费者的效用为消费者的支付意愿与数字产品售价之间的差值，效用不同的数据产品对于消费者的效用也会不同，消费者会倾向于选择对自己效用最大的产品。基于这种理解，本研究将消费者支付意愿与数据效用的关系量化为非线性函数，并将消费者效用作为数据定价模型中的核心变量之一，建立一个更加符合市场实际情况的定价策略。

(3) 多版本数据定价策略

由于市场需求的多样化，生产者通常提供不同质量水平的产品版本，满足不同消费者的需求。市场细分理论强调，市场由具有不同需求和偏好的消费者群体组成。生产者可以通过将产品细分为不同的质量水平，满足各个细分市场的需求。这种理论认为，消费者的需求是异质的，因此提供不同质量和价格的产品可以最大化市场覆盖率，提升企业再市场中的竞争力。从消费者的角度来讲，高质量产品通常能够吸引追求品质的消费者，而低质量或中等质量产品则能满足对价格敏感的消费。除此之外，生产者在设计以及生产最高质量版本的产品时，会产生固定的成本，由于信息产品的特殊性，在生产出高质量的产品之后，开发低质量版本的技术以及资源投入相对来说会减少，因为低版本的开发过程中并不存在挑战性的技术任务，因此，从成本控制的方面来讲，多版本策略也是合理的一种选择。

在数据市场交易中，数据提供者通过提供不同质量的版本满足不同消费者的需求，而每个版本的定价需基于数据的效用及市场需求来决定。本研究考虑通过数据效用来区分数据产品，由于消费者对数据效用有着不同的敏感度，对同一份数据进行处理，形成效用不同的多个版本。

最终的数据定价模型将同时考虑数据的多版本策略以及消费者效用，在消费者根据支付意愿进行选择的情况下，针对多版本数据采取合理的定价策略。数据提供者需要决定具体的定价策略，提供具体的版本数量，以及每个版本所对应的价格，以此来达到利润的最大化。而消费者的自由选择直接影响到数据产品的销售情况，从而决定了数据市场拥有者的利润。通过数据提供者和消费者之间的相互影响，以利润的大小反应当前定价策略的优劣性，从而对定价策略进行调整以得到该情况下的最优解。

1.3 模型求解与数据验证

本研究提出的定价模型涉及多个变量和复杂的市场行为，在模型构建完成后，将采用最优化算法中的遗传算法对模型进行求解。遗传算法（GA）作为一种基于自然选择和遗传机制的优化方法，能够高效地搜索到最优解，因此非常适合用于复杂的定价模型求解。在数据定价问题中，遗传算法可以帮助寻找最优的定价策略，确保在满足不同消费者的需求，最终目的是实现数据提供方的利润最大化。

为了进一步证明模型的可行性，本研究还将选择实际的数据集进行数据验证。此过程中，将采用数据集作为数据产品模拟市场中的交易情况，并根据模型结果进行定价决策。通过数据验证，能够确认模型在实际应用中的表现，并评估其在不同数据版本、不同消费者需求下的适用性。此外，为了实现数据交易的实际应用，本研究将设

计并构建一个小型的数据交易仿真系统。该系统能够模拟一个完整的数据定价过程，包括数据的上传、数据价值评估，最优数据定价策略。通过仿真系统，验证了本研究提出的数据定价模型的实操性，确保其能够在真实的市场环境中落地应用，从而证明本研究的定价策略具备较强的实用价值和可操作性。

2 研究目标和效果

2.1 研究目标

(1) 理论目标：

①构建更加准确的数据价值评估函数：本研究的首要目标是构建一个全面、精确的数据价值评估函数。综合考虑多维数据价值评估因素（如数据容量、数据质量、隐私保护和机器学习预测效果），建立一个系统的、多维度的数据效用函数。该函数能够准确反映数据的多维特性及其在不同场景下的实际应用价值，为科学合理的数据定价提供理论基础。

②提出基于消费者支付意愿的定价模型：消费者的支付意愿是数据定价中的关键因素，本研究目标之一是基于消费者的支付意愿构建一个非线性的定价模型，模拟不同数据版本和数据效用对消费者决策的影响。在该模型中，消费者的支付意愿不再是简单的线性关系，而是随着数据效用的变化而发生非线性变化。这一模型考虑了消费者在面临不同质量数据时的选择偏好，能够为数据提供者提供更加精确的定价方案。

③优化多版本定价策略：数据市场的多样性要求数据提供者能够针对不同消费者的需求推出不同版本的数据产品。本研究旨在构建能够适应不同消费者需求的多版本定价策略，为数据提供者提供优化的定价方案。通过分析消费者支付意愿的非线性变化，模型将最大限度满足消费者需求，同时帮助数据提供者提升市场覆盖率和利润。

(2) 工程目标：

开发功能完备的函数集成工具：在理论框架的基础上，基于构建的效用模型，开发一个函数集成工具，集成数据效用计算、消费者支付意愿建模和多版本定价策略，为数据提供者和消费者提供便捷、科学的定价支持。工具支持多维数据输入、动态效用计算和智能定价策略生成，满足市场需求的多样化和动态化。工具的开发将使得数据定价不仅限于理论层面的推演，还能够直接应用到数据交易平台或企业数据市场中，真正发挥其商业价值。

2.2 研究效果

本研究通过理论与工具的结合，在多个方面取得显著成果，为数据市场的健康发展提供支持。具体效果如下：

(1) 为数据定价提供新思路

本研究提出了基于多维度效用的数据价值评估与定价模型，与传统单一维度评估和定价方法相比，更加全面地反映了数据的实际价值。这一新思路丰富了现有的理论框架，为数据交易提供了更科学的决策依据。

(2) 多维因素整合评估体系

与以往研究中仅依靠单一因素（如信息熵或数据质量）进行定价不同，本研究将数据容量、质量、隐私保护及机器学习预测效果等多维因素整合为一个效用模型。通过量化这些因素的综合贡献，提出了价值分数的概念，并论证了其合理性，使得数据价值评估更加科学和系统。

(3) 注重消费者利益与市场透明度

通过构建消费者支付意愿的非线性模型，研究提出了能够更好匹配消费者需求的定价策略。该策略充分体现了消费者的购买偏好和支付意愿，使消费者能够根据效用选择最合适的数据版本，提升了市场透明度和公平性，同时优化了消费者购买决策的效率。

(4) 提供具体数据定价解决方案

以机器学习中的数据集为例，本研究结合实际数据形式，提出了适应具体场景的数据定价方法。通过实验验证，证明了模型的有效性和适用性，为数据交易平台和企业内部数据管理提供了实用工具。

(5) 提升市场覆盖率，实现利润最大化

研究提出的多版本定价策略能够在满足不同消费者需求的同时，优化市场覆盖率，扩大数据交易规模。数据提供者可以通过高质量版本实现利润最大化，同时利用低版本数据吸引更多消费者，形成良性的市场循环，促进数据市场的持续发展。

(6) 推动数据要素市场的规范发展

通过科学的数据价值评估与定价工具，本研究成果提升了数据交易的效率与透明度，减少了市场信息不对称问题，有助于数据资源的高效流通和优化配置。此外，模型的实际效果可为政策制定提供科学依据，助力数据要素市场的规范化和透明化发展，进一步推动数字经济的高质量发展。

3 拟解决问题

(1) 填补数据价值评估理论的空白

当前的数据价值评估方法多以单一因素为核心，例如数据的规模、质量或信息熵。然而，数据的价值是多维度的，不同维度的特性（如容量、质量、隐私、预测效果等）对数据整体价值的贡献存在差异，单一维度的评估往往难以全面反映数据的真实经济价值。本研究提出基于多维度效用的评估框架，将数据容量、数据质量、隐私保护和机器学习预测效果等多个关键因素综合考虑，通过多因素加权的方式构建统一的效用函数。这种模型不仅解决了传统评估方法的单一性问题，还能够更加科学地量化数据的整体价值，为数据定价提供坚实的理论基础。

(2) 解决市场需求与消费者支付意愿之间的关系问题

消费者支付意愿是数据定价中的核心因素之一，而传统的定价模型多依赖于简化的线性假设，忽视了支付意愿随数据效用变化的非线性特征。不同消费者对数据效用的敏感程度不同，且随着数据效用的提升，支付意愿呈现出复杂的变化趋势。例如，当数据效用接近消费者的预期时，支付意愿急剧增加，而当效用超过预期时，支付意愿的提升速率会逐渐减缓。这一非线性支付意愿的特征目前尚未得到充分的研究和建模。本研究通过引入消费者效用理论，基于数据效用与消费者支付意愿之间的非线性关系，构建了一个更加符合实际市场需求的定价模型。这一模型能够精确刻画消费者的支付行为，帮助数据提供者实现精准定价，并有效提高市场覆盖率和利润最大化。

(3) 解决多版本数据定价策略的科学问题

在数据市场中，不同消费者对数据的需求和偏好差异巨大，数据提供者通常会根据消费者需求推出多个版本的数据产品，例如标准版、高质量版、定制版等。如何在多版本产品中进行合理定价，既能满足市场需求，又能最大化利润，一直是数据定价中的难题。目前，大多数研究集中于单一数据集的定价策略，缺乏对多版本定价机制的深入探讨。本研究提出的多版本定价策略，结合数据效用评估与消费者支付意愿模

型，能够为不同版本数据定价提供科学依据。在此基础上，本研究提出的定价策略能够精准匹配不同消费者的需求，实现市场的细分化和利润最大化。

(4) 推动数据市场的透明度与公平性

当前数据市场存在信息不对称的情况，数据提供者与消费者之间往往缺乏充分的信任和透明度。数据消费者在选择购买数据时，往往无法准确评估数据的真实价值，导致市场上存在不公平的交易。通过科学的定价模型和透明的数据效用评估机制，可以提高市场信息的透明度，增强消费者的信任，从而推动数据市场的公平性和可持续发展。本研究通过建立基于数据效用的定价模型和公开透明的数据评估标准，有助于提高数据交易的透明度，确保消费者能够根据数据的实际效用做出理性选择，进一步提升市场的公平性和效率。

4 论文结构安排

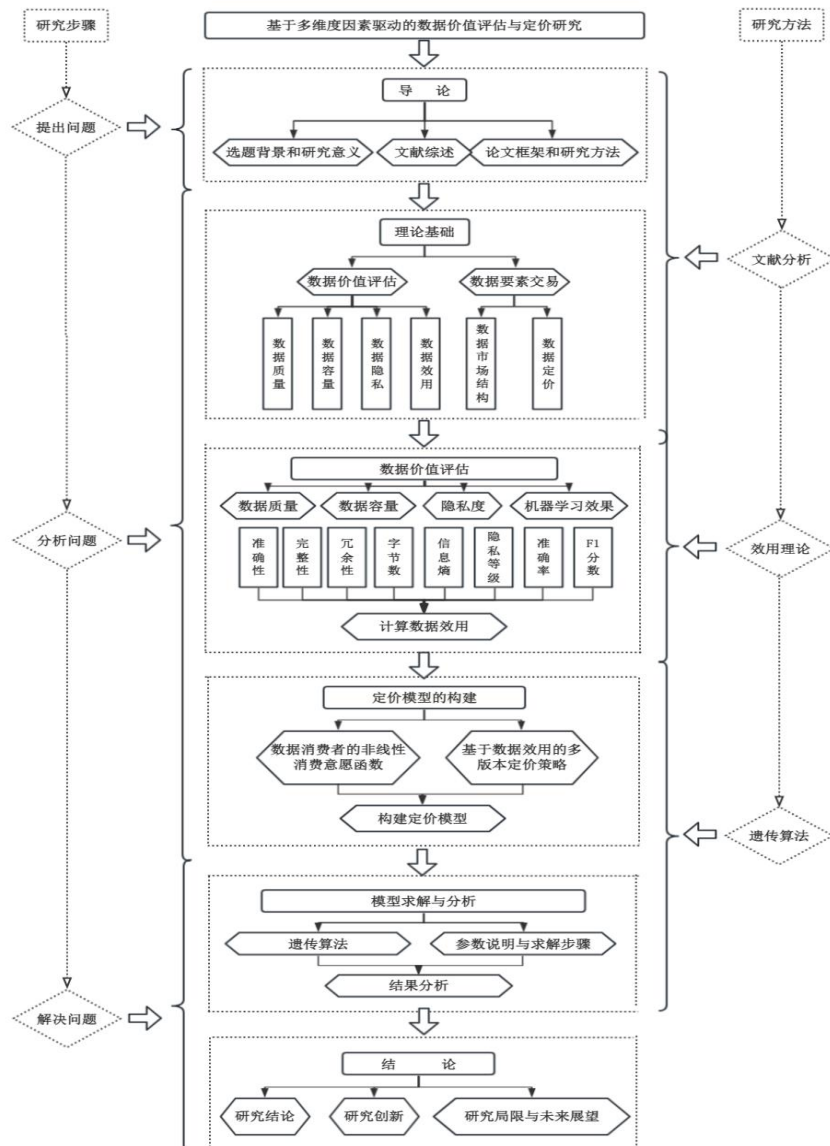


图 2-2 论文结构安排

本论文的研究内容主要分为以下 6 个部分：

绪论部分详细阐述了数据价值评估与定价的研究背景、研究意义，分析了国内外相关领域的研究现状，并介绍了研究目标、技术路线及论文结构。

第一章介绍了数据价值评估与交易的理论基础，分析了数据市场的基本结构、数据定价的关键要素，并阐述了数据容量、数据质量、数据隐私及机器学习预测效果等多维价值评估因素的理论依据。

第二章阐述了数据价值评估的具体方法，详细说明了各评估维度的量化指标，包括准确性、完整性、隐私等级等，并通过多因素加权计算得出数据效用，为后续定价模型的构建提供了科学依据。

第三章构建了基于多维数据效用的定价模型，分析了数据消费者的非线性支付意愿函数，提出了基于多版本数据效用的定价策略，明确了数据定价的核心逻辑与应用场景。

第四章采用遗传算法对定价模型进行求解，并结合模拟实验和实际数据验证模型的效果，通过市场覆盖率和总利润的分析，证明模型的合理性与有效性，同时设计了一个小型数据交易仿真系统，进一步验证模型的实际落地可行性。

第五章总结了论文的研究成果与主要贡献，分析了研究的不足，并对未来数据价值评估与定价策略的发展方向提出了展望。

三、研究方案设计及其可行性分析（包括：研究方法，技术路线，理论分析、计算、实验方法和步骤及其可行性等）（不少于 800 字）

1 研究方法

为了科学地构建数据价值评估模型和定价模型，并验证其在实际市场中的应用效果，本研究采用以下几种研究方法：

（1）文献分析法

本研究首先通过系统梳理国内外关于数据价值评估、定价策略以及消费者支付意愿的相关文献，明确当前研究的理论基础和发展现状。通过分析现有研究的优缺点，提炼出多维度的数据价值评估参数（如数据容量、数据质量、隐私保护、机器学习预测效果等）以及非线性支付意愿模型的关键点，为本研究的创新点和模型构建提供理论依据。

（2）因素分析法

数据价值评估涉及多个关键因素，每个因素从不同维度反映数据的市场价值。本研究通过因素分析法，筛选出影响数据价值的主要参数，包括数据容量、数据质量、隐私保护和机器学习预测效果。进一步地，针对每个因素构建量化指标。通过多维度因素分析，本研究能够全面反映数据价值，为后续效用函数的构建提供支持。

（3）理论分析法

本研究采用效用理论作为模型构建的理论基础，通过效用理论对数据的价值进行量化。效用理论强调商品或服务满足消费者需求的能力，本研究将其应用于数据市场中，将数据的综合效用作为评估其市场价值的核心指标。首先，利用基数效用理论对数据价值进行定量分析，通过多因素加权方式计算综合效用分数。其次，结合消费者支付意愿的非线性特性，构建消费者效用函数，分析不同效用水平的数据对消费者支付意愿的影响。

（4）最优化算法

在模型求解过程中，本研究采用遗传算法对构建的定价模型进行优化求解。遗传算法通过模拟自然选择的过程，在复杂的多目标优化问题中能够探索多版本数据定价策略的全局最优解，避免陷入局部最优，具备全局搜索能力，且适合处理多约束问题，能够快速收敛并提供最优解。通过遗传算法的求解，本研究可以为数据提供者制定更加科学合理的定价策略，满足不同消费者的多样化需求。

2 算法的选择

由于最优化算法是模型求解的核心环节，本研究选择遗传算法作为核心优化工具，为了进一步论证选择的合理性，以下从遗传算法的优势以及与其他常用算法的对比分析出发，阐明其适用性。

2.1 数据定价问题的复杂性与特点

数据定价问题的核心在于通过综合考虑多维度的影响因素，优化数据提供者的利润和消费者的效用。其复杂性主要体现在以下两个方面：

（1）高维多目标优化

数据定价模型需要同时兼顾多个目标：一方面追求数据提供者的利润最大化，

另一方面满足消费者的效用最大化。不同版本数据的效用差异、市场需求等复杂因素使得问题具有典型的多目标优化特性。传统优化算法通常只能处理单目标或低维度问题，在应对多目标优化时效率较低。

(2) 非线性和多约束

数据定价问题中的消费者支付意愿随着数据效用呈现非线性变化，同时模型中还包含多个约束条件，如消费者预期与实际效用的匹配、数据版本间的效用差异等。遗传算法无需线性假设，能够灵活适配非线性、多约束的复杂优化问题，是解决此类问题的理想选择。

2.2 遗传算法的优势及适用性

遗传算法是一种基于自然选择和遗传学原理的全局优化算法，模拟生物进化过程，通过选择、交叉和变异操作，逐步逼近最优解。选择遗传算法作为本研究的核心算法，主要基于以下几点优势：

(1) 全局优化能力强

遗传算法通过种群搜索机制，能够有效避免陷入局部最优问题。在数据定价问题中，效用函数和支付意愿的非线性复杂性导致传统算法容易陷入局部最优，而遗传算法在高维搜索空间中表现出优异的全局搜索能力。

(2) 多目标适配性

数据定价模型需要在数据提供者利润最大化和消费者效用最大化之间寻找平衡。遗传算法能够通过种群多样性，自然适应多目标优化，支持解的 Pareto 前沿求解，是处理多目标问题的理想工具。

(3) 非线性与多约束适应能力

遗传算法无需目标函数的连续性和可微性，能够处理复杂的非线性关系和多种约束条件。在数据效用函数和消费者支付意愿函数中存在非线性特性时，遗传算法能够保持高效的优化能力。

(4) 灵活性与扩展性

遗传算法的参数（如种群规模、交叉率和变异率）可以灵活调整，以适应不同问题场景；其编码方式也能灵活适配本研究的多版本定价策略。

(5) 高效性

遗传算法通过并行化种群搜索，可以在较短时间内探索多个解，提高计算效率。这一特性尤其适用于本研究需要处理的大规模优化问题。

2.3 遗传算法与传统优化算法的对比

算法	特点	适用场景	遗传算法的优势
梯度下降法	依赖目标函数的连续性和可微性，适合单目标优化	可微的低维优化问题	不依赖目标函数连续性，适应非线性复杂问题
博弈论方法	模拟市场主体策略互动，适用于主体行为分析与策略均衡	双边市场策略互动、均衡问题	无需复杂市场假设，直接优化生成最优定价策略
动态规划法	适用于分阶段决策问题，但计算复杂度随问题规模呈指数增长	小规模、多阶段的规划问题	并行化种群搜索显著提高计算效率

(1) 梯度下降法

梯度下降法需要目标函数连续且可微，但数据定价问题中的效用函数和支付意愿具有非线性特性，无法满足梯度下降法的假设。此外，梯度下降法容易陷入局部最优解，缺乏全局搜索能力。

(2) 博弈论方法

博弈论适用于模拟市场主体间的策略互动，但依赖较强的理性假设（如完全理性、信息对称等）。数据定价模型中的多版本策略和消费者支付意愿的复杂性使得博弈论建模过程繁琐，遗传算法通过直接优化方法更高效地生成定价策略。

(3) 动态规划法

动态规划适合分阶段决策问题，但其计算复杂度随着问题规模呈指数增长。在高维多目标优化问题中，动态规划效率较低，而遗传算法通过种群并行搜索，能够显著提高计算效率。

2.4 遗传算法与其他智能优化算法的对比

算法	特点	适用场景	遗传算法的优势
粒子群算法 (PSO)	搜索效率较高，但在高维问题中易陷入局部最优	单目标优化、低维问题	群体多样性更高，避免局部最优问题
蜂群算法 (ABC)	模拟蜜蜂觅食行为，适合单目标优化	连续优化问题、低约束场景	在多目标和多约束问题中性能更稳定
模拟退火算法 (SA)	退火过程计算量大，优化速度较慢	小规模非线性问题	群体搜索机制提高求解效率，更适合动态调整场景

(1) 粒子群算法 (PSO)

粒子群算法适用于低维连续优化问题，但在高维非线性问题中容易陷入局部最优。遗传算法通过种群多样性和随机变异操作，能够更好地探索高维非线性搜索空

间。

(2) 蜂群算法 (ABC)

蜂群算法在单目标优化中表现较好，但在多目标和多约束问题中表现不稳定，可能出现震荡现象。遗传算法通过交叉和变异操作，能够稳定求解复杂约束下的问题。

(3) 模拟退火算法 (SA)

模拟退火算法在非线性优化中具有一定应用，但计算效率较低，退火过程需要大量迭代才能收敛。遗传算法的种群搜索机制显著提升了计算效率，尤其适合动态调整场景。

通过对常见算法的比较分析可以看出，遗传算法凭借其全局搜索能力、非线性适配性、多目标优化支持以及计算效率等优势，能够更好地满足数据定价问题的复杂优化需求。与传统算法相比，遗传算法突破了对目标函数的连续性和可微性限制；与其他智能优化算法相比，遗传算法在多目标优化、约束处理及搜索效率上表现更为优异。因此，遗传算法是解决本研究问题的最佳选择，为数据价值评估与定价策略的研究提供了强有力的技术支持。

3 技术路线与对应理论分析

为了实现数据价值的科学评估与定价策略的构建，本研究制定了清晰的技术路线。研究以理论构建为核心，结合数据分析、模型构建和实验验证，通过多维数据价值因素的量化、数据效用函数的构建、消费者效用函数的建模以及遗传算法的优化求解，最终实现数据交易模型的有效性验证。以下为关键技术路线及对应的理论分析。

3.1 数据价值评估因素的衡量

本研究从数据容量、数据质量、隐私保护和机器学习预测效果四个维度入手，对每个因素进行定义和量化，具体的量化方法见表 3-1。

(1) 数据容量

本研究采用对数函数对容量值进行平滑处理，以更科学地反映数据容量与价值之间的关系。根据香浓信息论，信息量随数据规模的增长逐渐趋于饱和，边际信息量的增速会递减，这表明数据容量对价值的影响并非线性。类似地，经济学中的边际效用递减规律也指出，随着资源规模扩大，其边际效用会逐步降低。采用对数函数处理容量值，可以有效捕捉这一递减效应，使容量对数据价值的影响更符合实际规律。此外，数据集的容量跨度往往较大（如从 MB 到 TB），直接使用容量值可能导致尺度不一致，影响模型计算的稳定性，而对数函数的平滑特性能够解决这一问题，降低尺度差异，增强计算的鲁棒性。

(2) 机器学习训练效果

在量化机器学习训练效果时，本研究采用模型训练的准确性作为主要度量指标。准确性反映了模型在特定任务中正确预测的比例，是评估数据对模型性能贡献的重要指标。从理论上讲，机器学习模型的性能直接依赖于训练数据的质量与特性。高质量的数据能够为模型提供更好的特征分布与信息支持，从而提升预测的准确性。准确性作为一种易于理解且被广泛应用的评价标准，能够清晰地反映数据在具体应用场景中的实际价值。

此外,从效用理论的角度,数据的价值体现在它对特定任务结果的改善程度上,而准确性是这一效用的直观体现。采用准确性作为量化指标,能够真实刻画数据在实际应用中的价值增益。而在实践中,模型的训练准确性较高,意味着数据对机器学习模型的支持效果越好,其市场价值也随之提升。为了便于与其他因素整合,准确性值将被标准化到统一的数值范围,确保其量化结果具有良好的适配性和对比性。

(3) 数据质量

数据质量的度量综合考虑了数据的准确性、完整性和冗余性三个关键维度,通过加权整合方式对数据质量进行全面评估。这些维度分别从数据的真实性、完整性和有效性角度反映了数据在交易和使用中的可靠性与价值。

①准确性:数据的准确性衡量数据值与真实值之间的接近程度。通过计算数据的误差值与总值的比例,可以得出一个准确性评分。准确性越高,数据的真实性和信任度越高,其市场价值也越显著。在量化过程中,本研究通过“1 减去误差比率”的方式,确保数据误差越低,评分越高,直观反映数据的实际可靠性。

②完整性:数据完整性是衡量数据集中是否存在缺失值的重要指标。完整性较低的数据可能会导致分析结果的偏差,影响数据的使用价值。在实际量化中,本研究通过统计缺失值占总记录数的比例进行计算。完整性评分通过“1 减去缺失比例”的方式体现,缺失值越少,评分越高,表明数据的覆盖范围更全面。

③冗余性:数据冗余性反映数据集中重复记录的比例。高冗余性会增加数据处理成本并降低其有效性。为量化冗余性,本研究计算重复记录占总记录数的比例。冗余性评分采用“1 减去重复比例”的方式,重复率越低,评分越高,反映数据的高效性和独特性。

上述三个维度的评分结果根据权重参数进行加权整合,得到数据质量的综合评分。这种方法能够全面衡量数据在准确性、完整性和冗余性方面的表现,为数据效用的进一步计算提供了科学依据。

(4) 数据隐私

在量化数据隐私时,本研究采用信息熵作为主要度量指标,以衡量数据隐私的复杂性和敏感性。信息熵来源于香浓的信息论,其核心思想是通过数据的不确定性来量化信息量。在数据隐私的量化过程中,信息熵能够有效地反映数据集中每个数据元组的隐私贡献,最终累积形成整个数据集的隐私评分。

信息熵通过计算数据集中各元组的概率分布来衡量数据的隐私复杂性。当数据分布较为均匀且信息较难预测时,信息熵值较高,表示数据隐私性较强;而当数据具有较强的规律性或可预测性时,信息熵值较低,隐私性相对较弱。这种方式能够科学地反映数据隐私在不同场景下的特性。具体量化时,将数据集视为多个元组的组合。通过计算每个元组的熵值,并将其进行累积,最终得出整个数据集的熵值作为隐私评分。这一过程不仅考虑了数据集中每个元素的隐私特性,还能够体现出数据集整体的隐私复杂度。

表 3-1 数据价值评估维度的量化方法

因素	描述	量化方法	公式
数据容量	数据集的大小，用记录数或字节数衡量	对数函数平滑容量值	$V_{容量} = \log(N)$ ，其中 N 为记录数
机器学习训练效果	模型预测的准确性（准确率、F1-score 等）	准确率或其他评价指标	$V_{机器学习准确性} = \frac{\text{正确预测数}}{\text{总预测数}}$
数据质量	包含准确性、完整性、冗余性	各子指标综合评分	$V_{质量} = \omega_1 V_{准确性} + \omega_2 V_{完整性} + \omega_3 V_{冗余性}$
- 准确性	数据真实度或误差范围	与真实值偏差	$V_{准确性} = 1 - \frac{\text{错误值数}}{\text{总值数}}$
- 完整性	数据缺失情况	缺失率	$V_{完整性} = 1 - \frac{\text{缺失值数}}{\text{总值数}}$
- 冗余性	数据重复程度	去重比例	$V_{冗余性} = 1 - \frac{\text{重复记录数}}{\text{总记录数}}$
数据隐私	数据的敏感性，基于信息熵评估	熵公式	$V_{隐私} = H(X)$ $= - \sum_{i=1}^n p(x_i) \log(p(x_i))$

3.2 效用函数的构建

(1) 多维数据效用函数

多维数据效用函数通过综合多种数据特性（如数据容量、数据质量、数据隐私、机器学习效果）来评估数据的整体价值。该函数以加权线性组合的方式，将各特性量化结果整合为统一的效用值，其计算方法如下：

$$U_{data} = \omega_1 V_{容量} + \omega_2 V_{准确性} + \omega_3 V_{质量} + \omega_4 V_{隐私含量}$$

(2) 消费者数据效用函数

在构建消费者的效用之前，首先要明确消费者的非线性支付意愿。假设消费者对数据产品的效用敏感，当数据产品被提供给消费者之后，消费者按照一定的标准从中选择适合自己的产品。敏感度高的消费者对于数据价值的追求要高于敏感度低

的消费者，当需要在两个数据效用不同的版本之间做出选择时，敏感度高的消费者会倾向于选择效用更高的数据。

假设 e 为消费者对价值分数的敏感程度， U_e 为数据敏感度为 e 的消费者预期的数据效用，其函数特征如下：

①当数据的效用 $U_{data} < \theta_1 U_e$ 时，支付意愿函数 $WTP = 0$ ：如果数据的效用远低于消费者的期望，消费者将很难从中获得足够的满足感或价值，导致其支付意愿几乎为零。这意味着，消费者在选择数据时对最低效用的数据几乎没有兴趣，进而不会愿意为其支付较高的价格。

②当数据的效用 $\theta_1 U_e < U_{data} < U_e$ 时，消费者支付意愿 WTP 急剧上升：当数据效用接近或达到消费者的预期时，消费者会感知到较高的价值，从而产生较强的购买意愿。在这个阶段，消费者支付意愿的增加是比较迅速的，因为数据提供的效用与消费者的期望匹配或接近，消费者认为该数据能够有效满足其需求，因此愿意支付更高的价格。

③当数据的效用 $U_e < U_{data}$ 时，消费者支付意愿 WTP 的提升速率逐渐减缓：当数据的效用明显超过消费者预期时，消费者仍然会感受到数据的高价值，但其支付意愿的增加速率会逐渐减缓。此时，消费者可能已经满足了他们的主要需求，对额外效用的需求变得较为有限。因此，虽然数据效用高于预期，消费者的支付意愿增速不会再像前两个阶段那样迅猛，且消费者可能因为价格上升而选择购买价格较低版本。

基于上述方法得到消费者的支付意愿后，引入消费者的效用函数如下：

$$U_c(e, U_{data,k}, p_k) = WTP - p_k$$

其中， U_c 为产品对于消费者的效用， k 为数据产品的版本号， $U_{data,k}$ 为第 k 个版本产品的数据效用， p_k 为第 k 个版本产品的出售价格， WTP 为与消费者敏感度和数据效用相关的函数，表示消费者的支付意愿。当效用 $U_c > 0$ 时，表示消费者购买产品能够带来一定的效用，有可能购买此产品，相反，当 $U_c \leq 0$ 的时候，消费者一定不会购买此产品。当存在多个产品能够为消费者带来效用的时候，消费者会选择效用最大的产品。

3.3 建模求解

(1) 模型的构建

结合上述讨论的数据多版本及消费者多样化选择的定价策略，构建相应的数学模型。

首先，假设交易市场中有数据效用不同的 k 个数据版本和对数据敏感度不同的 n 个消费者，根据消费者的选择意愿，对于每一位消费者来说，其目的为效用最大化，则其约束条件如下：

$$\begin{aligned} \max_{x_{ij}} \mu_j(x) &= \sum_{i=1}^k U_c(e_j, U_{data,i}, p_i) x_{ij} \\ i &\in [1, k], j \in [1, n] \\ U_c(e_j, U_{data,i}, p_i) &\geq 0 \\ x_{ij} &= 0 \text{ or } x_{ij} = 1 \\ x_{i_1 j} x_{i_2 j} &= 0, i_1 \neq i_2 \end{aligned}$$

其中， μ_j 为第 j 位消费者的总效用， U_c 为消费者效用函数， e_j 为第 j 位消费者的

敏感度, $U_{data,i}$ 为第 i 个版本产品的数据效用, p_i 为第 i 个版本产品的出售价格, x_{ij} 为消费者的选择结果, 取值为 0 或 1 并且消费者只能选择一个版本进行购买。

对于数据提供者来说, 需要觉得具体的定价策略, 即确定提供版本的数量和每个版本所对应的价格, 其目的是实现利润的最大化, 则约束条件如下:

$$\begin{aligned} \max_{\{U_{data,i}, p_i\}} \varphi(U_{data}, p, x) &= \sum_{i=1}^k \sum_{j=1}^n p_i x_{ij} \\ x_{ij} &= 0 \text{ or } x_{ij} = 1 \\ U_{data,1} &< U_{data,2} < \dots < U_{data,k} \\ p_1 &< p_2 < \dots < p_n \end{aligned}$$

φ 为数据提供者的总利润, 模型的最终目的是实现利润最大化, 产品的售价总是高于成本, 每一件商品的数据效用及其对应的销售价格 $\{U_{data,i}, p_i\}$ 是决策变量。不同版本的数据产品在数据效用上存在着差异, 并且序号越大的版本数据效用的值越高。同时随着产品数据效用的提升, 该版本的售价也随之增加。

(2) 遗传算法求解

为了实现模型的最优解, 本研究采用遗传算法 (GA) 进行求解。遗传算法 (GA) 是一种模拟自然选择和遗传学原理的全局优化方法。GA 通过模拟自然界的进化过程 (选择、交叉、变异), 不断生成新的候选解, 直至找到最优解。其主要优势包括:

①**全局搜索能力:** 遗传算法能够探索广泛的搜索空间, 避免了传统优化方法可能陷入局部最优的问题。

②**适应复杂问题:** GA 不依赖于问题的可微性或线性假设, 能够适应复杂的非线性、组合优化问题, 适合多版本定价问题。

③**处理约束:** GA 可以通过适应性选择和变异操作, 自适应地满足问题中的约束条件。

在数据定价问题中, 遗传算法可以帮助寻找最优的定价策略, 确保在满足不同消费者需求的同时, 实现数据提供方的利润最大化。遗传算法通过模拟生物进化过程的选择、交叉和变异等操作, 迭代优化定价方案。其适应度函数将综合考虑数据效用、消费者支付意愿以及市场竞争等因素, 通过多代进化, 寻找最优解。具体操作步骤如下:

①**初始化种群:** 初始种群由 N 个定价方案组成, 每个方案表示为一个向量 $r_k = [(U_1, P_1), (U_2, P_2), \dots, (U_K, P_K)]$ 。

②**适应度函数:** 适应度函数是计算每个个体 (定价方案) 在给定的市场环境下的总收益。即给定每个版本的价格 P_K 和效用 U_K , 通过计算每个版本的销售量和总收益, 得出个体的适应度。

③**选择操作:** 根据适应度函数值选择较优的个体进入下一代, 常见的方法有轮盘赌选择法和锦标赛选择法。

④**交叉操作:** 通过交叉操作将父代个体的效用和价格信息结合, 生成新的后代个体。交叉方法可以是单点交叉、多点交叉等, 即将父代个体的价格和效用对 (U_K, P_K) 交换。

⑤**变异操作:** 变异操作用于防止算法陷入局部最优解。通过变异操作, 随机调整某个版本的价格。变异方式可以是随机增减价格 P_K 大小。

⑥**终止条件:** 当算法的迭代次数达到预设的上限, 或者适应度函数在多个代之间

趋于稳定时，遗传算法终止，输出最优解。

通过遗传算法的迭代过程，将获得最优的定价方案，数据提供者提供的最优版本数量和每个版本的最优价格。

3.4 数据验证

本研究通过数据验证与仿真系统设计两部分展开实验与分析，以验证多维数据效用函数和数据定价模型的科学与适用性。数据验证评估模型的理论合理性和精度，仿真系统则测试模型在模拟交易环境中的实际应用效果。

(1) 数据验证

数据验证的主要目的是评估本研究提出的多维数据效用函数和定价模型在实际计算中的可行性。通过选取公开可用的数据集（如 UCI Machine Learning Repository、Kaggle 等平台数据集），测试数据效用函数对数据价值的量化效果和定价模型的计算性能。验证过程重点评估数据效用函数是否能够合理整合数据容量、数据质量、数据隐私和机器学习效果等多维度特性，并准确反映数据的综合价值。同时验证定价模型是否能基于效用函数的计算结果生成合理的价格，确保模型具有良好的可操作性和鲁棒性。数据验证的结果将直接体现模型的科学性和应用可行性，为下一步的系统化实现奠定基础。

(2) 仿真系统设计

为了进一步验证模型的应用价值，本研究设计了一个数据交易仿真系统，主要通过一个可交互的前端页面实现数据价值评估和最优定价策略的生成。系统支持用户上传数据集或输入相关参数，通过多维数据效用函数计算出数据的综合价值。用户还可以调整数据特性（如容量、质量等），观察数据效用值的动态变化，并实时生成调整后的最优定价策略。通过这种方式，仿真系统不仅能够展示数据价值评估与定价的全过程，还为用户提供了直观的交互体验，帮助验证效用函数和定价模型在实际交易场景中的可行性和适应性。

四、本研究课题可能的创新之处（不少于 500 字）

1 理论创新

在理论层面，本研究通过引入多维数据效用函数，突破了传统数据价值评估方法单一化的局限性。当前主流的数据价值评估方法多集中于单一指标（如数据容量或数据质量）的计算，未能全面考虑数据的多维属性。本研究综合数据容量、数据质量、数据隐私和机器学习预测效果四个维度，以多维效用理论为基础，构建出更加精准的数据价值评估模型。这一创新不仅为数据价值评估提供了更全面的理论框架，还通过效用函数的量化方式实现了对数据特性的科学整合。

此外，与现有以博弈论为基础的研究不同，本研究创新性地采用了遗传算法对数据定价问题进行求解。传统的数据定价方法通常以博弈论为核心，通过构建生产者和消费者之间的策略互动模型，探讨最优定价方案。然而，博弈论在处理多维度、多约束的非线性优化问题时可能存在计算复杂度高、模型难以收敛等问题。本研究采用遗传算法，通过模拟自然进化过程（选择、交叉、变异），高效地求解多版本数据的最优定价策略。遗传算法不仅能够适应多目标优化问题，还能够处理数据效用和支付意愿的非线性变化，为数据提供者的利益最大化和消费者效用最大化提供了更优解。理论上的这一创新，不仅为数据交易市场提供了新的定价思路，还为复杂经济问题的优化求解提供了新方法。

2 工具创新

在实践层面，本研究开发了一种用于数据价值评估与定价策略生成的工具。现有的数据交易市场虽然对数据的交易机制有一定探索，但普遍缺乏科学的价值量化工具和高效的定价模型。本研究填补了这一市场空白，设计并实现了一款能够动态评估数据价值并生成最优定价策略的工具。

这一工具的创新性体现在以下几个方面：首先，它基于效用理论，对数据容量、质量、隐私和机器学习效果等多维特性进行量化，提供了一个系统化的评估框架。其次，工具实现了数据效用的动态调整功能，用户可以通过修改数据特性实时观察效用值的变化，为数据改进和优化提供依据。最后，工具内嵌了基于遗传算法的最优定价策略生成模块，能够根据不同数据效用生成对应的价格方案，不仅提高了市场定价的合理性，还增强了数据交易的透明度。

这一工具为数据交易市场提供了一种高效且易用的解决方案，特别适用于需要快速评估数据价值并制定定价策略的场景。它的推出有助于提升数据市场的规范化程度，促进数据要素的合理流通，为数据提供者和消费者提供科学的决策支持。

五、研究基础与工作条件 (1. 与本项目相关的研究工作积累基础 2. 包括已具备的实验条件, 尚缺少的实验条件和拟解决途径) (不少于 500 字)

1 与本项目相关的研究工作积累基础

本研究基于对大量文献的深入阅读与分析, 明确了数据价值评估与定价领域的研究现状及其不足。通过对多维数据效用理论、遗传算法及数据市场交易策略等领域的研究成果进行系统梳理, 研究者掌握了相关理论方法并提出了具有创新性的研究思路。相比现有方法, 本研究的多维数据效用函数能够更加全面地量化数据价值, 同时基于遗传算法的定价模型克服了传统博弈论方法的局限性, 在理论和实践上具有重要的探索价值。

研究者在阅读文献的基础上提出的研究方法已经过初步推导和理论验证, 证明了该方法的可行性和潜在贡献。本研究结合当前数据交易市场的实际需求, 对多维数据特性进行量化整合, 并利用遗传算法优化定价策略, 为数据交易市场提供了新的理论工具。这些理论积累与创新构成了本项目顺利开展的重要基础。

2 已具备的实验条件

目前, 本研究已具备以下实验条件, 为项目的实施和推进提供了基础保障:

(1) **硬件条件:** 研究所依托的计算设备具备高性能处理能力, 配备 GPU 加速器和多核处理器, 能够高效处理大规模数据集和复杂的算法计算任务, 为遗传算法求解和多维效用函数的计算提供了强有力的支持。

(2) **软件环境:** 研究者已构建了完善的软件开发与测试环境, 集成了多种数据分析与优化工具, 包括 Python、MATLAB 等编程语言以及相关机器学习和数据处理库。这些工具能够支持模型构建、算法求解和结果验证的全过程。

(3) **理论支撑:** 在多维数据量化、效用理论应用和遗传算法优化领域已有较为充分的理论积累, 为后续模型的实施与优化提供了方法论基础。

3 尚缺少的实验条件与拟解决途径

尽管研究条件已较为完善, 但本项目在数据资源方面仍存在不足。模型验证与仿真实验需要多样化的高质量数据集, 而现有公开数据集尚无法完全满足需求。针对这一问题, 研究计划通过以下方式进行补充:

(1) **在线获取数据集:** 通过公开平台 (UCI Machine Learning Repository、Kaggle 等) 查找适用的数据集, 优先选择具备明确属性且结构完整的数据。

(2) **数据后期处理与多版本设计:** 在基础数据集的基础上, 通过后期处理生成多个版本的数据 (如不同质量或容量), 以满足实验中多维数据效用和多版本定价策略的验证需求。

学位论文工作计划

时间	研究内容	预期效果
2024. 11-2024. 12	完成开题报告撰写与答辩，明确研究方向和技术路线	确定论文研究目标、技术路线和研究内容；完成开题报告答辩
2024. 12-2025. 2	多维数据效用函数构建；消费者效用函数设计；定价模型设计	完成效用函数与定价模型的数学框架构建，初步完成理论验证。
2025. 3-2025. 5	遗传算法求解定价模型的研究	完成遗传算法实现，测试其性能与收敛性；撰写算法部分初稿。
2025. 6-2025. 8	数据验证与仿真系统设计	基于验证结果优化模型并设计开发仿真系统，验证模型的合理性与可行性。
2025. 9-2025. 11	完成中期检查与答辩	完成中期检查；优化研究方案。
2025. 12-2026. 2	撰写论文初稿	完成初稿撰写。
2026. 3-2026. 5	论文修改与定稿；准备答辩材料	提交论文终稿，完成论文答辩准备。
2026. 6	学位论文最终答辩	完成学位论文答辩



评 定 小 组 成 员	姓 名	职 称	导师类型	单位名称	职务
导师意见：					
导师（签名）：					
日期： 年 月 日					
开题报告小组意见：					
组长（签名）：					
日期： 年 月 日					
学院意见（签章）：					
负责人：					
日期： 年 月 日					