

数据经济学

第二章：数据经济学的基本原理

陈希路

暨南大学经济学院

2026 年春

章节目录

- 1 第一节：数据要素具有非物质性
- 2 第二节：数据打破了资源稀缺性的假设
- 3 第三节：数据一定程度上仍具有特定稀缺性
- 4 第四节：数据具有非竞争性
- 5 第五节：数据在经济活动中体现出有限排他性
- 6 第六节：数据要素的边际收益可实现递增
- 7 第七节：数据产品/服务的边际成本趋于零
- 8 第八节：数据要素的规模报酬递增
- 9 第九节：数据要素价值的非独立性
- 10 第十节：数据要素对传统生产要素的增益

经济形态与关键生产要素的演变

- **农业经济时代**：劳动力、土地。
 - **工业经济时代**：资本、技术。
 - **数字经济时代**：**数据**成为新的关键生产要素。
-
- **数据的独特性**：广泛存在、可复制、可共享、可无限增长。
 - **核心影响**：改变了经济学研究的底层逻辑——即**稀缺性约束**条件下的最优资源配置问题。

第一节

数据要素具有非物质性

数据要素的非物质性

传统要素的形态

- 实体性：土地、劳动力
- 虚实兼有：
 - 资本：实体（金属货币、固定资产）+ 虚拟（数字货币、金融产品）
 - 技术：虚拟服务 + 实体产品

数据要素的形态

- 主要是虚拟性/非物质性，这是数据与传统生产要素的关键差异之一
- 存在形式：以数字、符号等形式存在于计算机存储器中

数据内涵与形式的演变

回顾历史，数据的内涵随着生成和加工工具的发展而不断丰富：

- ① **来源**：自然界的运动变化和人类活动产生的痕迹
- ② **演进路径**：
 - 手工纸质时代 → 计算机时代 → **大数据时代**
- ③ **大数据时代的特征**：
 - **数量**：几何式增长
 - **含义**：泛指所有定性或定量的描述
 - **形式**：不仅仅是数字，还包括**文字、图像、声音**等

数据的非物质性具体体现在哪里？

1. 依附性

数据要素无法独立存在，需要依附于现代信息网络等载体

- **后果：**导致数据资源很难成为传统物权的客体，难以遵循传统物权规则

2. 无形损耗性

数据资源不会发生有形损耗

- **特征：**
 - 近乎零成本复制
 - 快速、无次数限制地复制
 - 可跨越时空限制被共享和共用

数据的非物质性具体体现在哪里？

3. 可变性

数据资源在流通过程中可能形成新的数据

- 对既有数据进行替换、迭代更新

4. 结合性

非物质形态的数据要素仍需与其他要素结合

- 必须与**劳动力**和**技术**结合才能发挥对经济增长的作用
- 数字经济发展需要与制造业、服务业等**实体经济部门**相结合

案例分析：数字导航地图与纸质地图

传统纸质地图：实体性要素

- 会发生物理破损和折旧
- 只能供单人或少数人同时查看
- 更新测绘数据的成本极高

数字导航地图（如高德地图、百度地图）：数据要素的非物质性

- **依附性**：必须依赖手机等智能硬件和移动网络基站才能发挥作用
- **无形损耗性**：用户同时在线导航，地图数据不会磨损，且能以近乎零成本无限复制
- **可变性**：根据用户的行驶轨迹和实时拥堵情况，底层数据不断迭代更新
- **结合性**：单纯的路网数据无法直接产生巨额经济价值，必须与路径规划算法、网约车或本地生活等实体服务结合

第二节

数据打破了资源稀缺性的假设

数据打破了资源稀缺性的假设

- 传统经济学中的“稀缺”:

- 描述资源的相对有限性
- 数量和质量在特定时间和地点是相对固定的

- 传统要素的损耗特性:

要素	稀缺性与损耗表现
土地	肥力下降、污染导致生产力损耗
货币/金融产品	对同一个人无法重复使用
厂房/机器	因折旧而减损
劳动力	生产能力和支付手段是稀缺的

数据打破了资源稀缺性的假设

- 不同于传统要素，数据量随着活动持续增长。

大数据摩尔定律

全球新产生的数据量大约**每两年翻一番**。

- **可再生性**：数据具有非稀缺性的属性
- **三大特性打破稀缺**：
 - ① **可复制**：数据库备份生成完全相同的副本（相同结构、内容、质量）；通过管理可增加价值、降低稀缺性
 - ② **可共享**：通过交换、集成、协作分析产生新数据，相互补充和增强
 - ③ **可加工再生**：通过增删改使数据更准确；通过统计分析、机器学习提取新知识

案例分析：开源代码托管平台与传统工业设备

传统工业设备：受制于稀缺性约束

- 一台实体高精度机床在同一时间只能被一家工厂投入生产
- 机器设备随着高强度运转必然产生部件磨损，最终面临报废

开源代码托管平台（如 Github）：打破稀缺性假设

- **可复制、可共享**：数据代码被同时下载调用时，代码不会发生任何损耗
- **可加工**：科研人员在使用数据代码时会修复漏洞，并提交全新的功能分支
- **可再生**：优化后的新代码反哺给平台，促进数据质量与数量的共同增长

数据打破了资源稀缺性的假设

- **不耗尽**：数据不会因使用而耗尽
- **越用越多**：
 - 在使用过程中通过持续收集和更新得到新数据
 - 数据量将无限增长
- **结论**：这在一定意义上打破了数据资源具有稀缺性的说法，改变了经济学的底层逻辑

第三节

数据一定程度上
仍具有特定稀缺性

数据一定程度上仍具有特定稀缺性

现象与悖论

虽然数据看似可以无限供给，但**具有特定使用价值的**数据资源仍然稀缺。

- **现实感受**：我们总是感到缺少所需要的数据，为了特定目标往往需要专门生产或购买
- **本质原因**：
 - 数据是事实的表现，但不等于事实
 - 数据是对事实某种程度的抽象
 - 制定规则和掌握规则的人对数据质量（可信度）有巨大影响

数据稀缺性的来源

数据是由人和计算机参与生产的初级产品，其生产受制于三个因素：

1. 受制于算力

- 数据的产生依赖算力，算力依赖数字基础设施与能源
- 这一稀缺性与**资本与能源的稀缺性**相关联
- **2025–2030 数据中心预测增长趋势：**
 - 容量翻倍（来源：仲量联行）
 - 电力需求翻倍（来源：国际能源署）

数据稀缺性的来源

2. 受制于算法

- 数据要素的开发依赖算法，即数据记录和处理工具
- 算法开发的技术门槛高，合格劳动力稀缺
- 这一稀缺性与**劳动力和技术**的稀缺性相关联
- 直接影响数据的可获得性、质量、准确性和时效性

3. 受制于个人隐私保护规则

- 与人相关的数据采集和使用受法规限制
- 制度约束造成了数据资源的稀缺

案例分析：自动驾驶（如特斯拉端到端模型）中的数据稀缺性

1. 算力与基础设施约束

- 训练先进的端到端自动驾驶大模型，需要斥巨资建设超级智算中心
- 处理海量高质量路测视频需要庞大的 GPU 算力集群支撑

2. “优质数据”的极度稀缺（算法数据约束）

- 尽管日常行驶视频海量，但包含罕见边缘场景的高质量人类驾驶数据稀缺
- 缺乏优质的结构化视频数据，端到端模型就无法学会处理复杂的真实路况

3. 隐私与合规约束

- 数据安全法规严格限制人脸、车牌及敏感地理测绘坐标的直接使用
- 繁琐的脱敏步骤进一步加大了有效数据的获取难度

第四节

数据具有非竞争性

数据具有非竞争性

首先，我们需要理解经济学中竞争性的概念。

竞争性 (Rivalry)

- 一个人使用某种物品会**减少**其他人对该物品的使用。
- 消费者使用增加 → 生产成本增加。
- **例子**：衣服、汽车、文具（别人不能同时使用）。

非竞争性 (Non-rivalry)

- 一些人对产品的消费**不会影响**另一些人的消费。
- 受益者之间不存在利益冲突。

数据的非竞争性特征

与传统生产要素相比，数据是无限可用的。

- **无限复用：**

- 现有数据可以被任意数量的经济主体同时使用
- **不会导致**数量减少或价值降低
- **例子：**100 万个标记图像或汽车行驶数据，可供多家公司同时用于训练算法

- **边际成本 (Marginal Cost)：**

- 数据生产和传输中，每增加一单位数据，边际成本**接近于零**
- 软件代码、数字媒体可以被无限次复制而不增加额外成本

案例分析：实体唱片与数字音乐

实体唱片：具备竞争性

- 一张实体光盘同一时间只能放入一台播放器中供特定人群听
- 每多生产和销售一张实体唱片，都需要增加相应的塑料材质、包装印刷以及物流运输等成本

数字音乐（如 QQ 音乐、网易云音乐）：具备非竞争性

- **无限复用**：存储在云端服务器上的同一首数字歌曲，可以被全球千万用户同时在线收听，且音源数据不会产生任何质量损耗
- **边际成本趋零**：多增加一位听众点击播放，音乐平台只需付出微乎其微的服务器带宽费用，内容本身的额外复制与生产成本几乎为零

非竞争性带来的机遇与挑战

- **机遇（正面）：**

- 降低了信息技术应用和门槛，带来更多创新机会
- 使用和共享变得更容易、便宜
- **应用：** 广告营销（了解需求）、医疗保健（个性化服务）

- **挑战（负面）：**

- 数据的无限复制可能导致数据质量下降
- 隐私泄露的风险增加

第五节

数据在经济活动中 体现出有限排他性

数据在经济活动中体现出有限排他性

排他性

一种商品或服务只能被一个消费者或一组消费者使用，其他消费者无法使用。

- **原因**：通常受所有权或许可证限制
- **典型例子**：**专利**（授予发明者独占权利）
- **作用**：促进创新，但也可能导致垄断和高价格

为什么说数据具有“有限”排他性？

- **数据本身的性质：非排他性**
 - 多个主体可同时使用同一组数据，不会产生干扰（互斥）
 - 例如：劳动力市场统计数据可被多个雇主共同使用
- **现实中产生有限排他性的原因：**
 - ① **交易动力不足：**非竞争性导致边际成本趋近于零，压低交易价格，企业倾向于囤积而非交易
 - ② **安全与竞争考量：**数据泄露会导致免费泛滥
- **结果：**平台企业收集、拥有并控制数据，将竞争对手**排斥在外**，从而在经济活动中形成了有限排他性

案例分析：AI 时代互联网平台的数据围墙与 API 收费

数据的天然非排他性

- 论坛上的公开讨论或社交媒体帖子，在物理层面上可以被全网的 AI 爬虫同时抓取
- 多个大语言模型同时读取同一份高质量语料，不会发生物理互斥或损耗

平台企业（如 Reddit、X 等）人为制造的有限排他性

- **关闭免费 API**：为防止 AI 公司免费抓取高质量的人类对话数据去训练大模型，各大内容平台纷纷关闭免费 API 并大幅提高数据访问价格
- **建立数据围墙**：头部平台通过修改协议和反爬技术，将海量优质语料转化为锁在自有生态内的核心数字资产
- **最终结果**：缺乏资金购买版权数据的初创企业被排斥在外，形成了事实上的数据有限排他性

第六节

数据要素的边际收益可实现递增

数据要素的边际收益可实现递增

什么是边际收益？

在技术水平不变的情况下，增加任何一种要素的投入，当该要素投入数量达到一定程度后，再增加单位该要素所带来的产量的增加量。

- **传统规律：** 边际收益递减
- **数据要素的新特征：** 边际收益因数据要素投入量的变化而变化

新增一单位数据要素收益的变化阶段

① 第一阶段：

- 难以发挥作用，创造价值较低
- 可能导致模型结果偏误，带来**负价值效应**
- 例子：之前版本的 ChatGPT 缺少中文数据，导致回答错误百出

② 第二阶段：

- 价值凸显，创造的价值量**急剧增加**
- 2023 年数据：领先的大语言模型参数规模量级须达到 5 万亿

③ 第三阶段：

- 知识质量已达一定高度，继续提高空间不大
- 新增价值较低，边际收益递减

为什么说数据要素具有边际收益递增效果？

虽然单体数据边际收益呈 S 形，但数据有独特的**价值属性**——**连接性**。

- 单个数据可能价值不大
- 当它与其他相关数据结合时，价值可以迅速成倍增加
- 数据规模越大、种类越多，越能焕发出强大的生产力
- **结论**：数据资源有边际收益递增的效果

案例分析：在线社交平台个性化推荐系统的演进

现实企业：抖音、小红书

边际收益的三阶段变化

- **第一阶段：**新平台刚上线时积累的浏览记录少，推荐算法如果推错内容，会引起消费者反感，呈现出负价值效应。
- **第二阶段：**随着用户的点击、点赞、收藏行为不断沉淀，推荐模型精度发生质变，从而实现千人千面的情景化营销。
- **第三阶段：**核心用户画像已经被精准刻画，再新增几条常规点击记录对算法精度的提升微乎其微，边际收益逐渐递减。

连接性带来的倍增效应

- 孤立存在的点击、点赞、收藏记录的商业价值极为有限。
- 将站内的用户行为与外部的商品购买趋势、实时天气情况进行交叉结合后，数据价值迅速放大，企业可以实现高转化率的情景化精准营销

第七节

数据产品/服务的 边际成本趋于零

数据产品/服务的边际成本趋于零

什么是边际成本？

每一单位新增生产的产品带来的总成本的增量。

- **趋于零的含义：**在不增加总生产成本的情况下，可以生产额外单位的产品
- 在数据产业中，不同阶段付出的边际成本几乎相同
- **具体表现：**
 - **生活：**大部分人几乎免费从互联网获取和分享信息
 - **生产（基础设施）：**
 - **云计算：**提供服务的同时回馈数据，丰富发展
 - **物联网：**处理持续大数据流，降低价值链成本
 - **公共产品：**弱化空间限制（如远程教育、医疗），实现均衡供给

“趋于零”的例外：什么时候需要付出成本？

必须承认，在某些情况下，每新增一单位数据要素也可能要付出一定的成本。

1. 不确定性带来的成本

- 数据具有非均质性、敏感性、安全性等特点
- 加工分析过程不确定性强
- 对**技术能力**和**信任**要求较高

2. 时效性带来的成本

- 数据具有时效性特征
- 随着时间推移，信息可能**过时**、**失效**
- 类比：技术进步带来的资产折旧

案例分析：金融数据终端的边际成本

现实企业：Wind、Bloomberg

边际成本趋近于零的行业特征

- 金融数据提供商将同一份处理好的宏观经济和股票行情数据分发给第 10000 个客户与第 10001 个客户，所需增加的额外网络传输与数据复制成本几乎为零。

必须付出的额外处理成本

- **不确定性带来的成本**：全球各大交易所及宏观机构的原始数据格式非均质化严重，且存在错误风险。为了向金融机构提供安全、标准且可信的数据，服务商需要投入人力与算法资源进行清洗和核验。
- **时效性带来的成本**：金融市场瞬息万变，上一秒的高频交易数据在下一秒可能直接失效。为了维持极低延迟的数据供给，企业必须持续投入资金升级光纤专线与计算节点。

第八节

数据要素的规模报酬递增

什么是规模报酬？

在其他条件不变的前提下，企业内部要素投入按相同比例变化时所带来的产出的变化。

- **传统要素：**

- 受限于工厂空间、设备效率等
- 增加要素组合对产出的边际贡献通常也**递减**

- **数据要素：**

- 具有非竞争性（Romer）
- 随着规模扩大、种类丰富，所蕴含信息越多
- 不同信息形成互补，更好地支持决策优化
- **结论：**总产出增长幅度 **超过**投入增长幅度

数据要素规模报酬递增的理论逻辑

① 平均成本效应

- 数据要素投入的固定成本较高
- 随着数据规模扩大，平均成本逐渐降低

② 网络效应

- 数据天然适合共享
 - 平台企业能降低行业平均成本
 - **金融行业例子**：少量数据 → 基本结论；海量数据 → 检测欺诈、预测趋势
-
- **核心效果**：价值创造实现“ $1 + 1 > 2$ ”
 - **条件**：需关注数据质量、隐私和伦理问题

案例分析：云计算与 AI 大模型生态的规模报酬递增

现实企业：阿里云、亚马逊

1. 平均成本效应

- 建设超大规模智算中心、研发底层云架构以及预训练千亿参数的大模型，需要投入数百亿元的初始资金。
- 当服务接入全球数百万家企业（开发者）时，每增加一次云端 API 调用，所需分摊的基础设施成本极速下降，实现规模报酬递增。

2. 网络效应

- 云平台上汇聚了金融、医疗、零售等千行百业的脱敏使用反馈，不同维度的场景数据大幅提升了 AI 大模型的通用泛化能力。
- 平台积累的数据种类越丰富 → 模型输出质量越高 → 吸引更多企业入驻带来新数据。总产出的增长远远超过单种数据投入的增长。

数据赋能实体经济

- **独特的经济学属性：**

- 非消耗性、易复制性
- 天然适合共享经济，激发网络效应

- **多维价值：**

- 不同维度数据在不同场景下具有不同价值
- 投入生产函数时更能体现规模报酬递增规律

- **最终目标：**

- 促进数据、技术、场景在实体经济中的深度融合
- 通过放大、叠加、倍增作用，提高**全要素生产率**

第九节

数据要素价值的非独立性

数据要素价值的非独立性

“数据是新的石油”？

这种说法指出了数据的重要性，但也存在误区。

- **不同点**：数据的价值**不取决于数量**，而是取决于提取价值的能力
- **非独立性**：当数据不与其他生产要素结合时，自身价值难以实现

决定因素：

- ① **科技应用范围**：取决于不同对象对其使用的程度（具有主观性）
- ② **数字基础设施**：依赖硬件、软件、云、网等支持

因素一：价值取决于科技应用范围

数据如何通过科技应用实现价值？以下是典型领域：

1. 电子商务

- 利用大数据和机器学习
- 分析购买历史与行为模式
- 提供个性化推荐

2. 医疗行业

- 利用大数据和人工智能技术
- 分析病历、症状、医学图像
- 提供准确诊断和治疗方案

3. 智能交通

- 收集流量、拥堵、定位信息
- 智能优化信号灯配时
- 提升出行效率

4. 农业领域

- 传感器收集土壤、气候数据
- 精准决定浇水施肥时间
- 提高产量和质量

因素二：价值实现依赖于数字基础设施

- **背景：**全社会需求激增，基础设施（IaaS、PaaS、SaaS）建设加快。
- **四大支撑作用：**

1. 数据收集和存储

- **工具：**传感器、物联网设备。
- **场所：**云平台、数据中心。
- **作用：**确保安全存储和高效管理。

2. 数据传输和连接

- **工具：**高速网络、5G。
- **技术：**云计算、边缘计算。
- **作用：**实现实时分析和决策支持。

因素二：价值实现依赖于数字基础设施

3. 数据安全和隐私保护

- **机制**：加密技术、身份验证、访问控制。
- **目标**：保护机密性、完整性、可用性。
- **作用**：建立可信赖的交换机制。

4. 数据处理和分析

- **能力**：云计算、大数据技术、人工智能算法。
- **过程**：大规模高效处理。
- **作用**：将数据转化为有意义的**洞察**。

案例分析：美团超脑系统的数据价值

原始数据的非独立性

- 孤立存在的骑手定位坐标、商家出餐读秒记录以及用户收货地址，如果脱离了特定的业务场景与处理工具，则毫无直接经济价值。

激活数据价值的两大决定因素

- **科技应用范围**：企业引入运筹优化与机器学习算法，将定位坐标与时间数据进行融合计算，瞬间生成最优配送路径，提升了骑手的履约效率。
- **数字基础设施**：千万级日订单的实时分发依赖强大的云计算平台提供算力支持，高速通信网络确保了骑手端与服务器端数据的毫秒级交互。

第十节

数据要素对传统生产要素的增益

数据要素对传统生产要素的增益

- **经济形态的演变：**
 - 农业经济：劳动力 + 土地
 - 工业经济：资本 + 技术
 - **大数据时代：**数据成为关键生产要素
- **增益机制：乘数效应**
 - 数据要素对资本、劳动力、土地、技术均给予赋能
 - 提高每一种资源的**配置效率**
 - 在生产函数中表现为对各传统要素的**乘数效应**

增益表现一：减少信息不对称，降低市场失灵风险

- **优化供应链管理：**
 - 实时了解市场需求，预测变化
 - 调整定价，降低库存积压和缺货风险
- **优化生产流程：**
 - 利用数据分析和智能技术实现**自动化**
 - 更有效地利用资本和其他资源
 - 减少因信息不对称带来的**福利损失**

增益表现二：提高资源要素的使用率

数据应用能显著提升各领域的产出效率：

领域	具体增益
农业	土地产出大幅提高，减少水、化肥投入
工业	数字化设备提高劳动力产出，降低劳动强度
生产运行	实时监控减少停机时间，提高设备利用率
交通	数字平台提高道路通行效率和车辆使用率

增益表现三：增进消费者福利

揭示“未被满足的需求”

数据的应用可以揭示消费者未被满足的需求，通过增加供给增进福利。

- **个性化生产：**
 - 揭示偏好，实施个性化生产。
 - 提供符合需求的产品和服务。
- **推动创新：**
 - 通过分析市场数据发现新机会。
 - 迅速将新产品引入市场。
- **改进质量：**
 - 监测用户体验，不断改进产品功能。

增益表现四：加快技术创新速度

案例：生成式 AI 与生物计算大模型助力医药研发

- **赋能环节**：蛋白质结构预测、药物靶点发现、化合物分子生成
- **显著成效**：
 - ① **底层逻辑**：如 AlphaFold 3 等模型实现了对几乎所有生命分子结构的精准预测，改变了传统盲人摸象式的湿实验
 - ② **缩短周期**：将原本需要数年甚至数十年的前期靶点发现与筛选时间缩短至数月
 - ③ **经济增量**：2024 年，据麦肯锡全球研究院测算，生成式 AI 每年预计可为全球制药和医疗产品行业带来 **600 亿至 1100 亿美元**的经济价值
- **结论**：高质量数据和前沿算法的结合提升了研发效率，为产业降本增效提供了革命性机遇

案例分析：京东物流的数据乘数效应

数据赋能资本与土地：提升资源使用率

- 传统物流仓储高度依赖人力步行拣货和庞大物理空间。
- 智能物流园区引入海量数据驱动的仓储搬运机器人。
- 数据算法为机器人实时规划最优移动路径，使单位土地面积的仓储坪效和固定资本的运转效率获得极大提升。

数据赋能供应链：消除信息不对称

- 传统零售供应链往往面临严重的库存积压或区域性断货损失。
- 物流平台利用庞大的消费历史数据进行区域销量精准预测。
- 在大型促销活动开始前，系统主动将特定商品前置调拨到离潜在消费者最近的配送站点，极大降低了全社会的库存周转成本，增进了消费者福利。

本章总结一：数据要素的属性与性质

第一层：形态与稀缺性

- **非物质性**：体现在依附性、无形损耗性、可变性、结合性等方面
- **稀缺性的辩证统一**：打破传统生产要素稀缺属性，但在算力、算法和隐私合规的现实约束下，高价值数据依然存在特定稀缺性。

第二层：竞争性与排他性

- **非竞争性**：多主体同时调用同一底层数据不会引发消耗和互斥。
- **有限排他性**：具有天然非排他性的数据在商业实践中，常因互联网巨头的技术封锁与数据囤积行为，呈现出有限排他性。

本章总结二：数据要素的属性与性质

第三层：成本与收益性质

- **边际成本趋近于零**：数据分发与复制的额外成本极低，仅受制于时效性与不确定性等带来的成本。
- **边际收益与规模报酬递增**：基于连接性、平均成本效应、网络效应，数据要素在生产投入中展现出边际收益递增与规模报酬递增的规律。

第四层：价值依赖与乘数赋能

- **价值非独立性**：原始脱敏数据无法直接变现，其价值取决于科技应用范围和数字基础设施。
- **对传统生产要素的增益**：数据可以对资本、劳动、土地等传统要素赋能，全面提升实体经济的全要素生产率。

谢谢大家!