数据市场数据定价综述

张梦晓、Fernando Beltr´an 和刘家谋*

摘要——数据市场是一种在线平台，将数据所有者、数据经纪人和数据消费者汇聚在一起，促进彼此之间的数据商品化。数据定价作为数据市场的一项关键功能，需要量化数据的货币价值。文献中已有大量关于数据定价的研究。本文旨在全面回顾现有数据定价研究的最新进展，以便深入了解这一新兴研究领域。我们的主要贡献在于提出一种新的数据定价研究分类体系，统一了决定数据价格的不同属性。该框架首先依据市场结构类型对这些研究进行分类，包括卖方市场、买方市场和双边市场。随后，在卖方市场中，研究进一步按查询类型划分，查询类型界定了数据消费者访问数据的方式；在买方市场中，研究则根据隐私概念划分，隐私概念界定了量化数据所有者隐私的方式。在双边市场中，隐私概念和查询类型均作为分类标准。我们系统地考察了分类体系中各类别下的研究。最后，我们讨论现有研究中的不足，并提出未来的研究方向。

索引词—数据定价、定价方法分类体系、数据市场

arXiv:2303.04810v1 [cs.GT] 7 Mar 2023

1 引言

"THE 正如《经济学人》在 2017[1]所言，世界上最有价值的资源不再是石油，而是数据。Web2.0、云计算和物联网等信息技术的发展，催生了前所未有的数据供应。个人、企业和政府日益采用数据驱动决策，释放了数据巨大的经济价值。事实上，阿里巴巴和 Alphabet 等一些全球规模最大、最成功的公司，其核心业务都离不开数据。例如，在美国医疗保健行业，数据驱动技术每年可创造额外的 3000 亿美元价值，其中三分之二来自医疗支出的减少 [1]。在零售业、政府行政管理和金融等其他领域，数据创造的价值也同样可观 [2], [3], [4]。数据密集型技术所需的输入中，很大一部分来自私人数据，即通常预期会被个人保密、不为公众所知的信息。一方面，隐私被视为一项基本个人权利，缺乏隐私会直接损害个人的安全与福祉。另一方面，为了让大众受益，数据挖掘不可避免地需要牺牲一定程度的隐私。维护个人权利与促进集体效用之间的这种内在冲突，构成了当今数据密集型技术面临的核心挑战。

过去三到五年间，数据市场的兴起为解决这一冲突提供了契机。通过数据变现，数据市场旨在平衡个人的隐私损失与私人数据的使用。例如，总部位于澳大利亚的 Data Republic[2] 向一家快餐连锁店提供去标识化的银行交易数据。将外部数据与内部数据相结合后，这家快餐连锁店制定出了比竞争对手更有效的促销策略[3]。另一个例子是，总部位于英国的 Citizenme[4] 为手机品牌 Wileyfox 收集数据，以改进其业务策略[5]。数据市场业务的其他真实案例包括 AggData[6], Data & Sons[7], Snowflake[8],Datum[9], digi.me[10] 和 Lota me[11]。在市场的一端，数据所有者 （如持有私人数据的个人或组织）掌握着自己的个人信息，并愿意以一定的补偿作为交换；补偿金额应体现交换造成的隐私损失。在市场的另一端，数据消费者 （如广告商、零售商、制造商和服务提供商）为分析任务寻求私人数据，并愿意为数据付费，以获取由此产生的经济价值。数据市场是一种在线平台，将数据所有者和数据消费者汇聚在一起，促成双方之间的数据买卖。在许多情况下，数据经纪人 会充当数据所有者与数据消费者之间的中介，负责收集、整合、存储和再分发数据，并从这些业务中获利。数据经纪人也可能是.....不可或缺的组成部分

• 张梦晓现任职于中国电子科技大学计算机科学与工程学院，位于中国四川成都。电子邮箱：*mengxiao.zhang@uestc.edu.cn*

• *Fernando Beltr´an*现任职于奥克兰大学商学院，位于新西兰奥克兰。电子邮箱：*f.beltran@auckland.ac.nz*。

• 刘家谋现任职于奥克兰大学计算机科学学院，位于新西兰奥克兰。电子邮箱：*jiamou.liu@auckland.ac.nz*。本文部分得到中国国家自然科学基金委员会项目（编号：*62172077*）的资助。1. Parkins, David. “世界上最宝贵的资源不再是石油，而是数据。”《经济学人》6 (2017年)。 2. https://www.datarepublic.com/ 3. https://www.datarepublic.com/resources/team-blog/project-highlights-across-our-data-ecosystem 4. https://www.citizenme.com 5. https://www.citizenme.com/case-studies/measuring-brand-nps/ 6. https://www.aggdata.com 7. https://www.dataandsons.com 8. https://www.snowflake.com/ 9. https://datum.org/ 10. https://digi.me 11. https://www.lotame.com/

该平台或外部第三方。

数据定价 旨在量化数据的货币价值，是数据市场的一项关键功能。例如，AggData 以固定价格出售位置数据，以折扣价出售数据集组合，按订阅费提供所有数据集的无限制访问权限，并通过协商提供定制数据。另一方面，Datacoup 向个人支付月费，条件是他们允许其他用户访问其在线账户的信息，例如 Facebook、LinkedIn 和 Google+。这些不同的数据定价策略需要系统研究数据的供需情况，以及数据价值在不同情境下如何实现。

市场结构在确定数据价值方面发挥着关键作用。数据市场的市场结构各不相同。数据市场主要分为三种类型：卖方市场、买方市场和双边市场 [5]。卖方市场整合来自多个来源的数据，并将数据出售给数据消费者。例如，AggData 提供位置数据的卖方市场。另一方面，买方市场使个人和组织能够将其内部数据货币化，让数据经纪人从数据所有者处采购数据。双边市场则连接双方，允许数据所有者向数据消费者提供数据。例如，Data & Sons 提供双边数据市场。在卖方一侧，数据的货币价值取决于消费者预期数据能够带来的额外价值。在买方一侧，数据价值体现了数据所有者因数据交易而放弃了多少隐私。当数据经纪人连接双方时，数据价格应反映数据消费者和数据所有者双方的考量。自然，隐私要求较低的数据所有者应获得更高的补偿，而信息需求较高的数据消费者则应支付更高的费用。

为体现数据所有者的隐私要求，Dwork 在 [6], 提出的差分隐私提供了一个全面的理论框架。该概念通过推断者从聚合查询输出中推断出个人数据记录的概率，严格定义了隐私。这一定义及其变体引起了广泛研究，并在业界得到应用 [7], [8]。另一方面，差分隐私远非体现隐私的唯一方式。例如，Gkatzelis 等人 [9] 以访问数据的数据消费者数量来量化隐私。Parra-Arnau [10]则以真实记录被披露的概率来量化隐私。这些（非差分隐私的）临时隐私定义通常聚焦于特定的应用场景或数据类型。尽管如此，它们仍构成了一批颇具分量的研究，探讨买方市场中的数据价值。

为捕捉数据消费者的信息需求，我们将数据消费者所需的查询大致分为三类。首先，在 AggData 等许多数据市场中，数据交易完成后，数据消费者可直接访问数据集。我们将此类数据请求称为空查询。其次，一次性查询 会返回单个固定的聚合统计量，例如数据的总和、平均值、最小值和最大值。例如，BookYourData[12] 和 DirectMail[13] 允许数据消费者使用查询，筛选出具有目标属性值的记录。最后，在一般查询中，数据消费者可以发出多个查询，而查询答案可以是聚合统计量，也可以是由多行多列组成的视图。

文献中关于数据定价方法的研究相当丰富。Jaisingh 等人 [11] 从运筹学的角度开启了数据定价研究。他们将私人数据视为一种特殊的信息产品，并认为数据的隐私级别是决定其价格的因素。Riederer 等人 [12], 沿用了这一思路，研究了用于销售数据集的无限供给拍卖。同年，Ghosh 和 Roth [13]重新审视了一次性查询下的隐私问题。与 [11], 不同，他们使用差分隐私来量化数据所有者的隐私损失，并将其与数据价值联系起来。后续一系列研究放宽了 [13] 中的假设 [14], [15], [16], [17]。另有研究[18], [19] 探讨了在一般查询下销售数据时的套利问题，并提出无套利定价方案来解决这一问题。此后，一般查询定价中的套利问题引起了学术界的广泛关注。 [19] 中的解决方案仅适用于特定查询类型，而后续研究则提出了适用于更一般查询的解决方案 [20], [21], [22]。Li 等人 [23]通过同时考虑隐私和套利问题，对 [19]进行了扩展。

Muschalle 等人 [24], 、Schomm 等人 [25]和 Stahl 等人 [26]是最早研究数据市场和数据定价的调查之一。这些调查通过实证研究考察了数据市场的特征、参与者以及实践中使用的定价模型。然而，它们未能全面综述这一主题，因为没有讨论数据交易中的信息不对称等关键问题。Liang 等人 [27] 对大数据市场相关研究进行了调查。 [27] 关注的是大数据的生命周期，而非定价本身。Wagner 等人 [28]系统地回顾了影响个人信息感知估值的因素的实证研究。这项研究为确定数据价值提供了见解，但没有提及数据定价方法方面的研究。此外，Fricker 和 Maksimov [29] 回顾了 18 篇研究论文中的定价模型和定价项目。他们采用的分类方法通常适用于任意产品的交易，却忽略了数据交易的独特特征，例如隐私保护的需求。特别是，这种分类未能体现数据估值。最新的文献综述由 Pei 撰写， [30] 总结了现有研究中数据定价策略所需具备的理想属性。这些考量维度包括真实性、收入最大化、公平性、无套利性、隐私保护和计算效率等。然而，以上调查均未通过关联数据定价的各种因素，综合提出一个整合且统一的框架。

本文试图全面综述现有数据定价研究的最新进展，以提供对这一新兴研究领域的总体理解。我们的主要贡献在于提出了一种新的数据定价研究分类体系

12. https://www.bookyourdata.com/

13. https://www.directmail.com/

，该分类体系统一了决定数据价格的不同属性。我们分类框架的基础是，按假定发生数据交易的市场结构类型对这些研究进行分类，即卖方、买方或双边市场。（1）在卖方市场中，数据的价值在于数据为数据消费者提供的信息价值。信息通过查询提供给消费者，而查询类型决定了如何衡量信息价值。这表明应使用查询类型作为区分数据定价研究的关键标准。（2）在买方市场中，数据代表数据所有者的隐私。在不同的隐私概念下，数据所有者的隐私损失以不同方式进行衡量。因此，选择隐私概念作为标准。（3）在双边市场中，数据同时以这两种形式具有价值，查询类型和隐私概念均决定数据价值的衡量方式。因此，这两个标准均被纳入考虑。我们系统考察了分类框架中各类别下的研究。随后，我们回顾了经济学和管理学领域中探讨隐私经济学的工作。最后，我们讨论现有研究中的不足，并界定未来方向。首先，可应用机器学习和强化学习来协助解决复杂的数据定价问题。此外，大多数现有研究局限于关系表及其上的查询。其他形式的数据，例如自然语言，需要新颖的定价方案。

本文结构安排如下：第 2 节介绍数据市场。第 3 节介绍经济学和定价模型的基础。第 4 节根据市场结构、查询类型和隐私概念，讨论数据定价研究的分类体系；随后，第 5、6、7 节将详细回顾各类别中的数据定价研究。第 8 节回顾经济学和管理学领域关于私人数据价值的研究。最后，第 9 节和第 10 节介绍未来机遇并作出总结。

2 数据市场

尽管早在 1998年，Armstrong 和 Durfee 的开创性研究 [31], 就探讨了数字产品（即信息）的交易理念，但“数据市场”这一术语是由 Schomm、Stahl 和 Vossen 在 2013 [25] 中提出的。作者将数据市场定义为一个平台，用户可在此购买不同的许可证，以访问各种数据集。随后，Muschalle、Stahl、L¨oser 和 Vossen [24] 指出，数据市场不仅有助于连接供需双方，还在数据收集与存储以及统一数据格式方面发挥关键作用。最近，Stahl 和 Vossen [5] 将数据市场的功能从数据集交易扩展到各种数据相关服务的交易（例如涉及数据挖掘算法的服务）。

如图 1 [24] 所示，数据市场可以提炼为一种商业模式。参与者主要有三类：数据所有者、数据消费者和数据经纪人。数据所有者 是拥有数据并愿意将其变现的人或机构，例如拥有自身社交、财务、位置和健康数据的个人，以及收集用户数据的公司，包括他们的人口统计信息、偏好和行为。数据消费者是寻求外部数据以改善决策、产品设计、服务和客户管理的个人或机构，例如广告商、软件开发商、零售商、制造商和电信服务提供商。数据经纪人 是中介，负责从数据所有者处采购数据、整合开放数据并将数据出售给数据消费者，并从数据交易中获利。上述数据公司均充当数据经纪人。

2.1 数据市场的结构

虽然数据经纪人的业务包括数据采购和数据销售，但其核心业务在不同的数据市场中各不相同。根据数据交换中发生的双边交易数量，市场结构依据 [5],可称为：

- 单边市场：一种市场交互模式，在这种模式下，数据市场只涉及买方或卖方中的一方，而不是双方。
  - 买方市场：如图 2(a)所示，在买方市场中，数据经纪人从多个数据所有者处收集数据，并相应地向他们支付报酬。买方市场服务于希望将其数据变现的组织和个人。
  - 卖方市场：卖方市场涉及数据经纪人与数据消费者之间的数据交易，如图 2(b)所示。对于寻求外部数据源以执行分析任务的组织和个人而言， [32], 卖方市场可满足其数据需求。
- 双边市场：一种市场交互模型，其中所谓的平台同时服务于买方和卖方。数据市场可视为一种双边市场，其中中间商——数据经纪人——为数据所有者和数据消费者提供数据交易平台，并从数据交易中获利。
  - 中心化双边市场：在中心化双边市场中，所有数据所有者和数据消费者都通过数据经纪人进行数据交易，如图 2(c) 所示。数据经纪人需要同时决定向数据所有者支付的采购价格和向数据消费者收取的销售价格。数据经纪人的利润是数据销售所得收入与数据收集成本之间的差额。例如，Data & Sons 作为数据经纪人，将数据所有者的个人数据变现，并向数据消费者出售聚合信息。由于出售给数据消费者的只有聚合查询结果，该公司能够保护数据所有者的个人隐私。
  - 去中心化双边市场：在去中心化数据市场中，如图 2(d) 所示，数据经纪人为数据所有者和数据消费者提供交易平台，但不参与数据交易，数据消费者和数据所有者则直接进行交易。

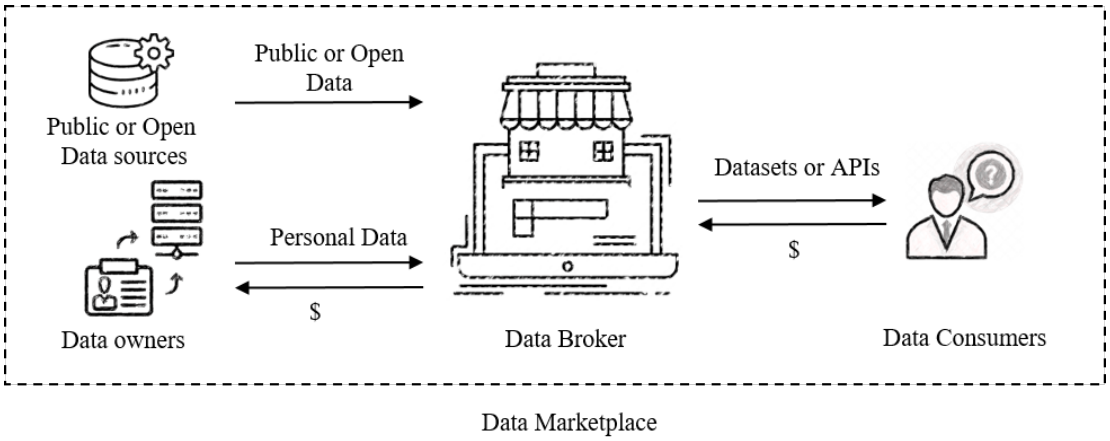


图1. 数据市场商业模式（改编自 [24]）

如果他们是数据交易平台的成员，则可以直接进行交易。Datum、CitizenMe 和 DataWallet[14]都是采用这种市场结构的数据市场示例。

2.2 数据市场交易什么 on?

卖方。卖方提供的商品有两类：数据集和查询。数据集可以是结构化的，例如关系表；也可以是非结构化的，例如图像、音频、图形和文本。大多数现有研究都假设一个数据集对应一个或多个关系表。在表中，每一列描述一个属性，每一行代表一条记录，而记录与某个对象相关联。例如，AggData出售的位置数据采用这种形式：一个数据集就是一张包含多列和多行、用于描述特定行业企业位置的表。每一列对应一个属性，例如地址、城市、州、邮政编码、电话号码等。每条记录对应一个具体的企业位置。与数据集不同，查询用于回答数据消费者提出的具体问题。查询答案可以是一个视图，通常由数据集中的行和列组合而成。例如，DirectMail[15] 出售消费者和企业的邮寄名单。其数据消费者可以使用邮政编码、年龄、性别等属性的过滤器，筛选目标消费者或企业。此外，查询答案还可能包含聚合信息和汇总测量值，例如数据或特定数据子集的总和、平均值、最小值和最大值。

买方。数据经纪人从买方的个人数据所有者处收集数据。数据所有者可以是拥有其用户数据集并将其出售的公司，也可以是拥有关于自身的一条记录的个人。无论哪种情况，买方交易的数据都是个人数据。

3 数据交易中定价方法的概述与基础

本节介绍数据定价中采用的基本经济学模型和定价模型。这些概念为后续章节对数据定价研究的分析奠定了基础。

在卖方一侧，制定面向数据消费者的价格时，会考虑消费者为获得产品所能支付的最高价格，即支付意愿（WTP）。而在买方一侧，重要的是数据所有者分享数据后愿意接受的最低补偿金额，即接受意愿（WTA）。数据经纪人的知识也在数据定价中起着关键作用。需要考虑两种情况：（1）数据经纪人掌握完全信息，即充分了解数据所有者和消费者的偏好；（2）数据经纪人只掌握信息不对称，即只有数据消费者（或数据所有者）自己了解这些偏好。接下来的两个小节将分别讨论完全信息和信息不对称条件下的定价方法。

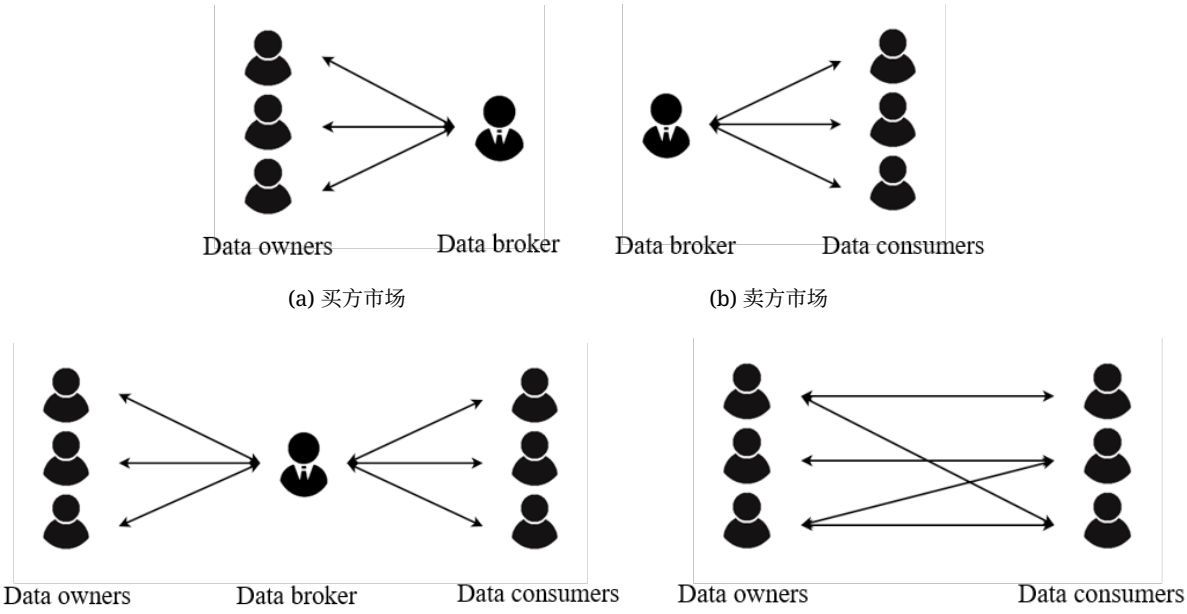
3.1 完全信息情境下的定价方法

数据市场可以采用多种定价方法，例如固定费用、高级定价、按次付费、两部收费制、版本化和捆绑销售。上述方法假设数据经纪人了解数据消费者或数据所有者的偏好，因此可以据此实现最优价格。

固定费用。固定费用通常指与使用限制相关的月度或年度订阅费。在固定费用定价策略中，时间是决定数据消费者需要支付多少费用的唯一因素。例如，Infochimps 收取月度订阅费，并允许数据消费者执行一定数量的查询 [18]。这是最简单的数据定价方法之一。

高级定价。高级定价是订阅选项的一种特殊情况，指按月或按年收取订阅费，并允许无限制使用。在高级定价模式下，消费者可访问的数据集或查询数量不受限制。例如，AggData 为支付高级费用的消费者提供对其 4,500 多个数据集的无限制访问权限。

14. https://datawallet.com/ 15. https://www.directmail.com/



(a) 买方市场 (b) 卖方市场

(c) 集中式双边市场 (d) 去中心化双边市场

图2. 四种数据市场结构

按次付费（或线性定价）。按次付费或线性定价是一种灵活的定价方式，以使用量为依据。价格 $p(n)$ 是使用量 n 的线性函数，即 $p(n) = p_u n$ 其中 p_u 是产品的单位价格。与固定费用相比，按次付费的一大优势是灵活性。对于使用频率不高的消费者来说，这种方式更合适；他们只购买一次或少数几次数据，之后便永久离开市场。

两部收费制。两部收费制是固定费用与按次付费的组合 [33], [34]。应用于数据市场时，数据消费者首先支付固定费用，以获取一定数量的数据集或查询。如果配额用完后还需要更多数据，他们可以按单位价格购买数据，即 $p(n) = p_f + p_u n$ 其中，p_f 是固定部分，$p_u n$ 是可变部分。这种方式适合需求不确定的消费者。

分层定价。分层定价为商品的不同数量设定 k 不同的单位价格 [35]。换言之，它包含k 多个数量层级 t_j 及相应价格 p_j。数量层级与价格的对应关系为$(t_1, p_1) \ldots, (t_k, p_k)$。对于每个层级，$p(m) = p_j m$，其中 $t_{j-1} \leq m < t_j$、$j = 2, \ldots, k$，且 $t_1 \leq t_2 \leq \ldots \leq t_k$ 和 $p_1 \geq p_2 \geq \ldots \geq p_k$。随着数量增加，单位价格下降，这表明存在数量折扣，在某种程度上能够促进销售。在分层定价方法中，需要确定价格 p_j 以及 t_{j-1} 与 t_j 之间的区间长度。

版本控制。版本控制是指以不同价格提供信息产品的多个版本 [36]。不同版本可能在功能、便利性和运行速度等方面有所不同，并面向不同类型的消费者。例如，在数据市场中，数据可以按不同质量或数量提供多个版本出售。例如，金融和经济数据公司 Quandl[16]，提供四种数据集版本。

捆绑销售（或打包）。捆绑销售是版本控制的一种特殊形式。它将多个物品组合在一起，作为一个套装整体出售。套装的价格通常低于其中各件物品价格的总和。数据市场也采用捆绑销售定价方法。例如，Agg-Data 将12 份小型美式餐厅名录打包出售，并以$425的折扣价销售，而不打包的名录则需要$688。

3.2 信息不对称情境下的定价方法

信息不对称是交易过程中常见的问题，指一方掌握其他方并不掌握的信息，从而导致资源分配效率低下。在数据市场中，卖方一侧的数据消费者对待购买的数据有一个预期价值，但数据经纪人并不知道这一信息。在这种情况下，数据消费者可能会策略性地支付低于真实价值的价格，以提高效用。为了减少信息不对称造成的经济损失，数据经纪人需要向数据消费者提供一定的激励，使他们愿意透露其隐匿信息。合同、拍卖和激励机制都是用于解决信息不对称问题的方法。

数据消费者和数据所有者都是会采取策略性行动的理性主体。换言之，在给定交易规则的情况下，他们会预测其他人的策略，并据此选择最优行动。为了揭示其估值中的隐匿信息，数据经纪人需要确保，对于数据所有者而言，真实披露估值是最优策略，且参与交易不会使他们处于不利地位。前一个条件称为激励相容（IC），后一个条件称为个体理性（IR） [37]。此外，数据经纪人通常

16. https://www.quandl.com

有用于数据收集的预算。当所有数据所有者的总支付额不超过预算时，即满足预算可行性（BF）条件。

合同。 在契约理论中，代理人为委托人 工作，双方的利益往往相互冲突。代理人掌握的信息比委托人更多，但委托人无法强迫代理人披露其信息。委托人会向代理人提供一份精心设计的合同菜单，并让代理人从中选择一份。每份合同都列明代理人可能采取的一种行动及相应的报酬。通过观察代理人行动所产生的结果，委托人可以推断出一些隐匿信息。保险契约是契约设计的典型例子。保险公司并不了解潜在被保险人的健康状况，这属于隐匿信息。但保险公司可以通过提供一份保险契约菜单并观察其选择，推断出一些健康信息 [37]。

在数据交易中，委托人可以是数据经纪人，而代理人可以是数据所有者或数据消费者。数据所有者对隐私往往有不同的偏好，数据消费者对数据也有不同的偏好。数据经纪人提供一份由不同数据和价格组成的合同菜单，让数据消费者或数据所有者选择一项方案。

拍卖。自拍卖出现以来，它们便一直在实践中用于确定价格。最常见的拍卖形式是英式升价拍卖：多位竞标者根据拍卖人报出的不断上涨的价格作出回应。形式上，拍卖定义了一套规则，依据来自代理人 [38] 的信息分配资源并设定价格。在拍卖中，拍卖人通过明确界定的拍卖，寻求出售或购买一件或多件物品。一组竞标者遵循拍卖规则，并在规定的时间点分别作出回应，披露一个类似于其对待售物品中某个子集的支付意愿（或购买意愿）的数值。拍卖人主持拍卖，并在满足结束条件时宣布最终价格，随后将物品（或出售权）分配给中标者。

数据市场中的拍卖运作如下：数据经纪人（拍卖人）公布分配和定价规则。数据所有者或数据消费者（竞标者）提交报价，可以是公开报价，也可以是密封报价。随后，数据经纪人根据预先设定的规则和提交的报价，宣布中标者以及中标者需要支付或获得的金额。

激励机制。更一般地说，拍卖属于机制设计所研究的一大类技术，可为解决信息不对称问题提供通用框架。粗略地说，机制设计相当于对推导博弈均衡这一问题进行逆向求解。具体而言，在参与者采取策略行动的博弈中，若给定由博弈均衡定义的期望结果，机制设计旨在寻找能够决定这些结果的博弈规则。激励机制通过将行动集合映射到结果的函数来定义代理人的行动集合和规则。具有激励相容和个体理性这两项性质的机制，能够激励参与者参与其中，并按照真实估值采取行动。

机制设计研究主要采用两种解概念：占优策略均衡和贝叶斯纳什均衡。在贝叶斯纳什均衡中，偏好属于隐匿信息，但市场中的偏好分布为所有代理人所共同知晓，每个代理人根据自己掌握的信息最大化其期望效用。另一方面，占优策略均衡是一种更强的解概念，它不对偏好分布作任何假设。当占优策略存在时，无论其他代理人采取何种策略，代理人都会选择能够使自身效用最大化的策略。在这两种解概念下，激励相容和个体理性约束的定义也相应有所不同。

4 数据定价方法的新分类体系

数据定价是一个具有挑战性的问题，在文献中受到广泛关注。为了全面理解这一主题，我们现在提出一种现有研究的新分类体系，如表1所示。

4.1 市场结构

我们的分类体系以市场结构为基础。具体而言，在卖方市场中，决定数据价格的是衍生信息的效用。例如，当零售商根据人们的购物历史做出销售决策时，购物历史数据的货币价值可以用该决策预期带来的额外利润来衡量。一些指标可用于衡量信息的优劣，例如质量、数量和准确性。相比之下，在买方市场中，数据价值体现为隐私泄露可能给数据所有者造成的货币损失。例如，当某人将自己的收入数据透露给他人时，这些数据可能会被他人甚至公众获取，从而可能对数据所有者造成损害。因此，在这种情况下，数据价值应反映因数据披露而预期给数据所有者造成的潜在损害金额（以货币计）。数据经纪人通常根据数据所有者的隐私偏好及其遭受的隐私损失来确定数据价格。此外，在双边市场中，数据同时体现这两种形式的价值。数据经纪人将数据出售给数据消费者以获取信息价值，同时补偿数据所有者所承受的隐私损失。数据经纪人需要在隐私损失与信息效用之间取得平衡，或在数据采购成本与数据销售利润之间取得平衡。由于双方的数据价值存在本质差异，因此应区分研究不同市场结构下数据定价的文献。为此，我们首先按照市场结构对数据定价研究进行分类。

4.2 卖方：查询类型

卖方的分类标准是查询类型。对于空查询，数据消费者可以访问整个数据集。购买数据集后，数据消费者可以自由探索数据集的内容。在这种情况下，可以使用质量、大小等多个指标来表示数据集的价值，

表1 数据定价方法的新分类体系。我们首先将相关研究分为三类：“买方”“卖方”和“双边”。在买方类别中，研究根据隐私概念进一步划分；在卖方类别中，研究则根据查询类型划分。在双边类别中，隐私概念和查询类型都会用于进一步划分研究。我们将根据该分类体系中的各个类别，考察数据定价领域的现有研究。表格中的单元格标示了讨论相应情形的章节。

双方 / 卖方 / 买方			查询类型		
			空查询	一次性查询	一般查询
隐私概念			5.1	5.2	–
	临时	6	7.1	–	–
	DP	–	–	7.2	7.3

来表示数据集的价值。对于一次性查询，数据消费者可以发起一个固定查询，该查询将数据的总和、平均值、最小值和最大值等聚合统计量作为查询答案返回给数据消费者。查询结果的准确性直接反映其价值。对于一般查询，与数据消费者只能获得特定查询结果的一次性查询不同，数据消费者可以发起多个查询，查询答案可以是聚合统计量，也可以是由多行多列组成的数据视图。数据消费者可以探索视图中包含的信息，并通过组合多个查询答案推断出一些信息。

4.3 买方：隐私概念

买方的分类标准是一种隐私概念，其根源在于个人或群体控制他人访问其个人数据的能力。人们对隐私有着不同的理解，并且已经提出了若干种隐私概念。

差分隐私 是现有数据定价研究中应用最广泛的隐私概念。基本上，差分隐私通过限制在对数据集进行查询并作出回答时，从聚合信息中泄露个人私人信息的程度来定义隐私。为此，我们考虑一个随机化机制 M，它在数据集 D 上执行查询，并随机输出一个实值聚合结果，即 $M(D): \mathbb{R}^n \times \Omega \to \mathbb{R}$，其中$\Omega$ 是一个概率空间。两个至多只有一个元素不同的数据集 D 和 D' 称为相邻数据集。如果 ε 是一个非负参数，则 ε-差分隐私定义如下 [6]：如果对于任意一对相邻数据集D 和 D'，以及输出空间O 中 M的所有可能子集 S，随机化机制 M 都是 ε-差分隐私的，则称该机制满足 ε-差分隐私：

$$\frac{\Pr[M(D) \in S]}{\Pr[M(D') \in S]} \leq e^{\varepsilon}$$

给定两个相邻的数据集 D 和 D'，当上述比值较小时，输出相同结果的概率相近，因此该机制能够保护个人隐私。参数 ε 可解释为隐私程度的量化指标。

与已成为标准化隐私概念的差分隐私不同，一些隐私概念以临时方式使用，例如，匿名化 [39]，允许访问数据的数据消费者数量 [9] 以及披露真实数据的比率[10]。我们将这些隐私概念称为临时概念。这些临时概念针对特定的数据交易场景定义隐私。以 [9]中使用的概念为例。数据经纪人向每位数据消费者出售一个样本数据集。当数据被选取的次数增加，从而被更多数据消费者访问时，相应数据所有者的隐私就更有可能被泄露。然而，这一概念无法应用于一般场景。

在本文其余部分中，我们将先考察单边市场的相关研究，再转向双边市场。综述的文献总结见表2。表2显示，在完全信息情境下，版本控制、按次付费和分层定价被用作定价方法；而在信息不对称情境下，则采用拍卖、合同和激励机制。此外，在完全信息情境下，按次付费、分层定价和版本控制常用于空查询，而版本控制则用于一般查询。

5 卖方市场

卖方市场的常见设置如下：给定一个由垄断数据经纪人和多个数据消费者构成的数据市场，数据经纪人向数据消费者出售查询，并旨在通过为数据消费者的数据集合理定价来实现特定目标。面对不同的查询类型，数据消费者会以不同方式探索信息。我们首先讨论空查询下的数据定价研究，然后讨论一般查询下的相关研究。

5.1 空查询下的卖方市场

在空查询下的卖方市场中，垄断数据经纪人向数据消费者出售空查询，并通过对数据集进行适当定价来实现目标，例如利润最大化。数据消费者对数据集具有异质偏好，例如对数量、质量、估值或隐私级别的偏好。在完全信息的情境下，数据经纪人了解他们的偏好（见 [35], [40]）；而在信息不对称的情境下，经纪人并不了解这些偏好（见 [12], [41], [42]）。

Kushal 等人 [35] 考虑了一个信息完全的卖方市场。他们研究了这样一种情况：数据消费者希望从垄断数据经纪人处购买若干个具有同质属性（如大小和质量）的数据集，并且每次购买都需要支付费用。数据经纪人的目标是根据消费者的支付意愿，为每个数据集设定适当的单位价格，以实现利润最大化。作者采用了两种简单的定价方法：线性定价和分层定价。该研究假设，数据消费者对数据集可能具有不同的边际效用：一组消费者具有恒定边际效用，而另一组消费者具有递减边际

效用。当累计支付意愿大于总价格时，消费者会进行购买。总利润是所有数据消费者支付金额的总和。最优价格可以从总利润函数中推导出来。

除了 [35],Yu 和 Zhang [40]中讨论的数量外，他们还考虑了会影响数据价格的数据质量。数据质量可以通过多个维度衡量，例如准确性、完整性、一致性和时效性，这些维度都会影响数据消费者[43]获得的效用。该研究考虑这样一种情形：数据经纪人面向一群追求效用最大化的数据消费者时，意图实现利润最大化；这些消费者对各维度的数据质量有着不同的期望。这种异质性促使数据经纪人以不同价格提供多个数据质量各异的版本。为刻画数据经纪人和数据消费者的行为，研究提出了一个涉及垄断者和多个潜在数据消费者的双层规划模型。通过求解该模型，可以得出各维度的最优质量水平以及每个版本的价格。这项研究的管理启示是，考虑数据集的多个维度有助于更好地细分市场，从而提高数据经纪人的利润。

Riederer 等人 [12]考虑一个存在信息不对称的卖方市场，具体而言，数据消费者对其数据估值拥有隐匿信息。对于数据经纪人而言，数据集一旦生成，将其提供给任意数量的数据消费者所产生的成本可以忽略不计。因此，作者将数据集视为供给无限，并采用基于指数机制（见 [44]）的无限供给拍卖进行数据定价。该拍卖保证近似激励相容。

Mehta 等人 [45] 研究了信息不对称涉及数据消费者对数据集需求的情况。许多数据市场会提供定制数据集，以满足数据消费者的需求和规格。也就是说，每个数据消费者都可以指定一条“理想”记录，这是一个描述理想数据集中某个条目属性的向量。假设数据消费者可以接受实际数据集与理想数据集之间的细微差异，但随着实际数据集偏离理想数据集，效用会逐渐衰减。此外，只有数据所有者知道理想记录和衰减率。拥有完整数据集的数据经纪人需要考虑提供哪些记录以及如何定价，才能使其收入最大化。这个问题被表述为一个多维机制设计问题，且难以求解 [46]。这项研究在理想记录均匀分布的假设下求解该问题。最优方案是，数据经纪人提供与理想记录距离在一定范围内的记录，并收取相当于数据消费者效用减去信息租金的价格。他们还提出了一种近似算法，在不依赖这一假设的情况下求解该问题。

Li 和 Raghunathan [42] 还考虑了数据消费者需求的异质性。一些消费者需要敏感度较低的“聚合”数据集，而另一些消费者则需要敏感度较高的“个人”数据集。数据经纪人并不知道消费者想要的数据类型。在这种情况下，数据经纪人销售两种类型的数据集：敏感度较低的“聚合”数据集，以及敏感度较高的“个人”数据集。

数据定价需要考虑数据消费者想要的数据类型，而数据经纪人无法获知这一信息。为了解数据消费者想要的数据类型，数据经纪人针对每种数据类型分别提供一种合约。每份合约均采用两部制收费合约，由固定部分和可变部分组成。定价函数写作 $p(s,\ m)\ =\ \alpha_s + \beta_s m$，其中 α_s 和 β_s 分别表示以敏感度级别 s 和数据量 m为依据的固定费用和可变费用。通过求解带有个体理性和激励相容约束的优化问题，可以得到最优定价参数α_s、 β_s 以及数据量 m，进而确定数据价格。聚合数据集和个人数据集交易框架为数据消费者和数据经纪人提供了实践参考。数据消费者更多地利用聚合数据，可以降低成本并减少隐私侵扰。数据经纪人则可以据此提供两种类型的数据集，改进其产品。

Naghizadeh 和 Sinha [47]考虑了数据消费者可能出现的不当行为。数据消费者获得查询答案后，可能会试图滥用数据（例如，泄露数据所有者的隐私）。为防止这种情况，作者针对数据交易设计了合同，其中包括多种类型的诚实数据消费者和一种对抗性数据消费者。每种类型的数据消费者都会收到一份报价，其中说明已发布查询的价格，以及消费者被发现滥用查询答案时需支付的罚款。数据经纪人设计合同以满足不同消费者的需求，并阻止对抗性数据消费者滥用数据。此类数据交易的契约设计问题被表述为一个优化问题，旨在满足激励相容（IC）和个体理性约束（IR）的同时，使预期收益最大化。随后，作者提出了一种近似算法来求解该问题。

5.2 一般查询下的卖方市场

在一般查询下的卖方市场中，常见的设定是：垄断数据经纪人向数据消费者出售一般查询，并致力于设计满足特定属性的定价函数。当允许数据消费者进行一般查询时，数据消费者获得的查询结果将是数据集中满足查询条件的一部分。这部分称为视图。数据消费者获得数据集的视图后，可能会进一步探索该视图中包含的信息。一个查询输出的视图可能包含另一个查询的视图。例如，包含整个人口收入的视图也会包含男性收入和女性收入。现在假设，男性收入的价格 p_1 与女性收入的价格 p_2 之和低于整个人口收入的价格 $p_1 + p_2$，那么数据消费者就可以购买男性和女性的收入数据，再转售整个人口的视图以获利。这就是套利的一个例子。为缓解这一问题，人们引入了“基于查询的定价方案”这一概念 [18]。

形式化地，给定数据集 D、显式价格 p_1、 ...、 p_m，以及 D 的视图 v_1、 ...、 v_m，一个数据查询定价方案 S

由形如 $(v_i,\ p_i)$ [48]的配对集合构成。视图 v 可以由来自 S 的多个视图组合确定，例如 u_1、$\ldots$、u_ℓ，表示为 u_1、$\ldots$、$u_\ell \twoheadrightarrow v$。数据消费者的查询可以通过一个或多个视图来回答，价格则可以根据定价方案 S 自动推导。这种基于查询的定价方案可在某种意义上视为版本控制，因为每个用于回答查询的视图都对应于整个数据集的一个版本。

考虑由一组视图 $v_1,\ldots,v_m$ 确定的视图 v，数据消费者以价格 p 购买 v（或以价格 $p_1+\ldots+p_m$ 购买 $v_1,\ldots,v_m$）。如果 $v_1,\ldots,v_m$ 的价格总和低于 v 的价格，即 $p_1+\ldots+p_m < p$，那么没有人会购买 v，所有消费者都会转而购买该视图集合，甚至可以按价格 p 转售该视图集合。设计查询定价函数的一项关键原则是无套利性 [48]。形式上，基于查询的定价函数 $p(\varphi)$ 以数据集 D 和查询 φ（或一组查询 $\varphi_1,\ldots,\varphi_m$）为输入，并返回一个非负实数价格。给定查询 φ，它由一组查询 $\varphi_1,\ldots,\varphi_m$ 确定，即 $\varphi_1,\ldots,\varphi_m \twoheadrightarrow \varphi$,，若基于查询的定价函数满足 $p(\varphi) \leq \sum_{i=1}^m p_i(\varphi_i)$，则该函数是无套利的。

关于一般查询下卖方市场数据定价的大多数研究，都试图通过设计无套利的基于查询的定价函数来解决套利问题。Koutris 等人 [19] 率先针对一种特定查询类型——广义链式查询开展研究。后续研究 [20], [21], [22], [49], [50], [51], [52], [53] 将结果推广到任意查询。

除了无套利性，Koutris 等人 [19]还认为，基于查询的定价函数应当不存在折扣。这意味着，对于查询价格 $p(\varphi)$，任何用于确定 φ 的查询组合的价格都不应低于 $p(\varphi)$；也就是说，当 φ_1、……、$\varphi_m \twoheadrightarrow \varphi$, $p(\varphi) \geq \sum_{i=1}^m p_i(\varphi_i)$ 时。否则，数据消费者永远不会购买查询组合 φ_1、……、φ_m。作者将所发布的查询限制在一种称为广义链式查询的合取查询子类中。例如，$\varphi_1(x,y) = R_1(x)\ R_2(x,y)\ R_3(y)$ 是一个广义链式查询，其中 φ_1 是一个包含变量 x 和 y 的查询，而 R_1、R_2 和 R_3 是表。此外，他们提出了一种多项式时间算法，可将广义链式查询转换为链式查询，并将链式查询的定价问题归约为最小割问题。边的容量是视图的显式价格，而查询的价格可通过计算图中的最小割成本得出。这项工作 [19]为数据经纪人提供了启示：他们应当关注数据市场中新出现的套利类型；同时，该工作也提供了一种可行的定价方法来解决套利问题。

[19] 中的定价方法是针对一种特定查询类型——广义链式查询而设计的。然而在实践中，数据消费者可能会发出任意查询，例如选择-投影-连接查询、聚合查询和随机抽样查询；理想的定价函数应与这些查询兼容，[20]研究了这一问题。Lin 和 Kifer [20]指出，查询价格反映了该查询能够揭示真实数据集多少信息，并从概率视角研究数据定价问题。当查询的答案有很高的概率揭示整个数据集时，该答案应设定较高的价格。

基于这一直觉，作者针对任意查询研究了两种定价方案：依赖答案的定价和独立于实例的定价。在依赖答案的定价机制中，查询答案的价格是查询和答案的函数；而在独立于实例的定价机制中，查询价格不依赖于数据库实例。令 $\mathcal{D}$ 为所有数据库实例的集合。对于任意 $D \in \mathcal{D}$ 和答案$\omega \in range(\varphi)$，查询$\varphi$的答案为 ω 的概率为 $\Pr[\varphi(D)=\omega]$。依赖答案的定价函数为 $p(\varphi,\omega) = \sum_{D\in\mathcal{D}} w_D \left(1-\Pr(\varphi(D)=\omega)\max_{D'\in\mathcal{D}}\Pr[\varphi(D')=\omega]\right)$，其中 w_D 是为每个数据库实例 D赋予的非负权重。独立于实例的定价函数为 $p\ \varphi$ $[\varphi\ D=\omega]$ $\sum_{D'\in\mathcal{D}}$ $[\varphi\ D'=\omega]$ w_D φ w_D $=$ $\omega]$ ( )

$= \sum_{D\in\mathcal{D}}\sum_{\omega\in range(\ )}$) $\Pr[\varphi(D)\log\Pr(\ )\Pr(\ )$，其中是满足 $\sum_{D\in\mathcal{D}} w_D = 1$的非负权重。这两个定价函数均已被证明具有无套利性。

同样地，Deep 和 Koutris [21] 完整刻画了无套利的独立于实例的定价函数，以及无套利的依赖于答案的定价函数。在依赖于答案的定价方案中，如果数据消费者发出查询 φ 并得到答案ω，她就会知道数据集 D 不可能是任意一个满足 $\varphi(D') \neq \varphi(D)$ 的$D' \in \mathcal{D}$。所有此类数据库实例 $\mathcal{S}_\varphi(\omega) = \{D' \in \mathcal{D} \mid \varphi(D') \neq \omega\}$ 的集合称为冲突集，它决定了查询的价格。给定一个数据库实例 D，令 $\mathcal{S}_D = \{\mathcal{S}_\varphi(\varphi(D)) \mid \varphi \in \Phi\}$，其中 Φ 是查询集；在偏序 $\subseteq$下，它构成一个并半格。定价函数 $p(\varphi,\omega) = f(\mathcal{S}_\varphi(\omega))$ 无套利，当且仅当 f 在每个半格（$\mathcal{S}_D,\subseteq$）上都是单调且次可加的，其中 $D \in \mathcal{D}$。此类定价函数的例子包括加权覆盖函数 $\sum_{D:\varphi(D)\neq\omega} w_D$，以及上确界函数 $\sup_{D:\varphi(D)\neq\omega} w_D$，其中 w_D 是分配给每个 $D \in \mathcal{D}$的权重。类似地，在独立于实例的定价中，对于查询 $\varphi \in \Phi$，作者在 $\mathcal{D}$上定义了一个由等价关系诱导的划分 $\mathcal{P}_\varphi$：当且仅当$\varphi(D) = \varphi(D')$ 时，$D \sim D'$ 成立。如果划分 $\mathcal{P}_{\varphi_1}$ 中的每个块都是另一个划分$\mathcal{P}_{\varphi_2}$中某个块的子集，那么 $\mathcal{P}_{\varphi_1}$ 细化$\mathcal{P}_{\varphi_2}$。按细化关系排序的这些划分构成一个并半格。定价函数 $p(\varphi) = f(\mathcal{P}_\varphi)$ 无套利，当且仅当 f 在这个并半格上是单调且次可加的。此类定价函数的例子包括香农熵函数和 q-熵函数。

使用 [21] 中的定价函数计算某个查询在所有可能数据集上的价格，是一个$\#P$-难问题。为解决这一计算上的局限，Deep 等人 [49], [50] 提出了一个名为 QIRANA 的框架。该框架应用 [21] 中的定价函数，但只考察一个较小的集合 $\mathcal{C} \subseteq \mathcal{D}$。这个集合 $\mathcal{C}$ 称为支持集，其中的每个成员都$\mathcal{C}$需要由数据经纪人预先设定价格。在 QIRANA 中，$\mathcal{C}$ 由一定数量的随机选取的数据库实例组成，这些实例与真实实例 D 相差一到两行。借助该集合，QIRANA 可以高效地计算大型数据集上某个查询的价格。此外，QIRANA 还支持历史感知定价 [52]。数据库本质上是动态的，新的元组会随着时间推移插入数据库。另一方面，数据消

费者可能会在不同时间点对同一个数据库发起多个查询。尽管这些查询可能相同，但由于数据库的动态特性，查询结果可能有所不同。然而，这些查询之间存在一定的重叠，因此消费者可能会为未发生变化的数据部分重复付费。历史感知要求数据消费者对于已购买的任何视图只付费一次。QIRANA [50]将新查询和历史记录中的所有查询视为一个查询捆绑包，并对该查询捆绑包使用无套利定价函数来设定价格。

先前关于基于查询的数据定价的研究，通过提出设计良好的定价函数来应对潜在的套利问题。然而，这些研究都没有考虑基于查询的定价方案中的收入优化问题。Chawla 等人 [53] 将收入纳入考量。他们研究数据经纪人与多个数据消费者之间的数据交易；每位消费者都希望执行一项查询，并为此设定了预算。只有当查询价格不超出预算时，消费者才会购买。此外，由于数据的复制成本可以忽略不计，数据经纪人拥有无限供应。数据经纪人的问题是在定价函数具有单调性和次可加性的约束下，最大化收入，即所有负担得起查询的数据消费者所支付款项的总和。该问题被表述为一个优化问题。

Xia 和 Muthukrishnan [51] 定义了一种称为版本套利的新型套利。在某些情况下，两个查询φ_1，φ_2 会生成数据库中非常相似的视图。例如，查询所有年龄超过 20 岁的人群收入，以及查询所有年龄超过 21 岁的人群收入时，返回的视图可能非常接近。如果前一个查询的价格高于后一个查询，那么打算执行第一个查询的数据消费者会转而购买第二个查询，因为第二个视图中的元组肯定包含在第一个视图中。形式上，对于两个查询 φ_i，$\varphi_{i'}$，将同时满足 φ_i 和 $\varphi_{i'}$ 的元组数量与满足 $\varphi_{i'}$ 的元组数量之比记为 $\Pr(\varphi_i|\varphi_{i'}) \geq \Pr(\varphi_i|\varphi_{i'})p(\varphi_i)$。若基于查询的定价函数满足$p(\varphi_{i'})$，则该函数无版本套利。作者提出了能够保证无版本套利性的基于查询的定价函数。其中一种是统一定价函数，为每项查询设定统一价格；另一种是非统一定价函数，为不同查询设定不同价格。他们还提出了一种贪心算法来求出最优价格。该算法从最优统一定价开始，并在保持无版本套利性质的同时，迭代更新某一项查询的价格。这项工作的主要见解在于，它提出了套利的新定义，并为数据经纪人销售查询提供了一种更灵活的方式。

在某些应用中，整个视图定价被认为过于粗略，因此对数据集中的单个元组定价更为合理。Tang 等人 [22] 解决了这一问题，并提出了一种基于元组的查询定价模型。他们应用数据溯源 技术，找出对查询答案有贡献的最小元组集合，该集合称为最小溯源。数据溯源描述了一条数据的来源及其处理历史 [54]。给定一个数据库 D 和一个查询 $\varphi(D)$，其溯源记为 L ($L \subseteq D$)。在 [22], 的定价模型中，给定一个元组集合 $X = \{t_1, \ldots, t_k\} \subseteq D$，$X$的价格定义为 p-范数，即 $||X||_p = (\sum_{i=1}^{m} s_D(t_i)^p)^{\frac{1}{p}}$, -其中 $t_i \in X$，而 s_D 是价格设定函数。给定数据库D上的查询 φ，用于生成查询答案 $\varphi(D)$的元组集合定义为最小溯源集合 $L_{\min}(\varphi, D)$。随后，查询 $\varphi(D)$ 的价格由其最便宜的最小溯源决定，即 $p_D(\varphi) =$ 最小值$_{L \in L_{\min}} ||L||_p$。

6 买方市场

买方市场的常见设置如下：在一个由垄断数据经纪人和多个数据所有者组成的数据市场中，数据经纪人从数据所有者处采购私人数据，并通过适当补偿数据所有者来实现某些目标。主要问题在于如何补偿数据所有者所感受到的隐私损失。隐私概念会影响隐私的量化方式。所综述的研究买方市场的论文采用了特定隐私概念，没有一篇采用差分隐私。具体而言，隐私通过以下方式衡量：在 [10], 中披露真实数据的概率、[39], 中通过匿名化实现的隐私保护级别，以及 [9]中获准访问数据的数据消费者数量。

数据所有者的隐私态度或偏好各不相同。在完全信息的情境下，数据经纪人知晓这些信息（参见 [10]）；而在非对称信息的情境下，这些信息则是隐藏的（参见 [9], [39]）。

在完全信息假设下，Parra-Arnau[10] 将私人数据视为信息产品，并采用版本控制作为定价方法。与经典信息产品不同，数据具有隐私这一属性。数据所有者拥有多个类别的私人数据，例如健康、宗教等，并对 m 不同类别的数据具有不同的隐私偏好。每个类别的隐私级别 $j \in \{1, \ldots, m\}$ 由参数δ_j表示，该参数表示披露真实数据的概率。此外，数据所有者会声明披露每个类别真实数据所需的补偿要求 p_j，总支付额为$p = \sum_{j=1}^{m} \delta_j p_j$。对于数据所有者而言，问题在于：给定支付额 p，如何通过为 δ_j 设定适当的值来设计最优版本，以尽量减少隐私披露，其中 $j \in \{1, \ldots, m\}$。作者将该问题表述为一个旨在最小化隐私披露的优化问题，并运用 Karush-Kuhn-Tucker（KKT）条件，求出最优的 δ_j、$j \in \{1, \ldots, m\}$及相应价格。这项研究的管理启示包括：为数据所有者提供了一种全面控制隐私披露的方式；还给出了最优披露策略，即在每个类别中披露相同数量的隐私信息。

在不完全信息下，Gkatzelis 等人 [10] 还允许数据所有者控制其隐私披露。为了控制隐私披露，数据所有者可以限制能够访问其数据的数据消费者数量。隐私损失是人数限制的函数。此外，数据所有者具有不同的风险态度。作者利用不同的风险态度，设计了一种称为确定性等价（CE）机制的拍卖机制，要求每位数据所有者报告其估值，并提供两种支付选项：第一种是固定支付，其金额取决于其隐私估值以及访问数据的消费者数量；第二种

是随机支付，其金额取决于其数据被使用的概率。

即使在不完全信息的情况下，Xu 等人 [39] 也认为，数据价值由数据消费者的效用决定，而效用与隐私保护负相关，因为隐私要求越严格，购买的数据就会受到越大的失真或遗漏。根据数据所有者的隐私偏好，数据经纪人提供一份合同菜单，其中包括三个要素，分别代表隐私级别、数据消费者获得的效用，以及向数据所有者支付的转移支付。这三个要素都由数据所有者的隐私偏好决定。为确定最优合同，作者将该问题建模为一个最大化问题，目标是在满足激励相容和个体理性约束，以及数据经纪人对其从数据中获得的总效用的最小要求的情况下，最大化数据经纪人的期望效用。优化模型推导出了最优转移支付函数和生产函数，而隐私级别的价值则通过实验确定。这项研究为数据经纪人设计数据交易合同提供了管理启示。具体而言，最优合同的制定很大程度上取决于数据消费者所需的效用以及数据对数据消费者的效用。

7 双边市场

双边市场的常见设置如下：面对一位垄断性数据经纪人、多个数据所有者和多个数据消费者，数据经纪人从数据所有者处采购私人数据，并向数据消费者出售查询。数据经纪人实际上面临两项数据定价任务，一项针对数据所有者，另一项针对数据消费者。因此，数据经纪人需要同时权衡数据所有者的数据隐私与数据消费者的效用。这是双边市场与单边市场之间最大的区别，也构成了这一场景的主要挑战。处理双方时，既要考虑买方的隐私概念，也要考虑卖方的查询类型。根据这两个标准，现有关于双边市场的研究可分为：基于空查询和特定隐私概念的双边市场、基于一次性查询和差分隐私的双边市场，以及基于一般查询和差分隐私的双边市场。

7.1 空查询与特定隐私概念下的双边市场

在空查询与特定隐私下的双边市场中，数据经纪人向数据消费者出售空查询，并使用特定隐私概念来量化隐私，从数据所有者处采购私人数据。该类别中的四项研究探讨了不完全信息下的双边市场。具体而言，Jaisingh 等人 [11] 和 Zheng 等人 [55] 假设数据所有者的估值未知。Cao 等人 [56] 假设数据所有者和数据消费者的估值均未知。Bergemann 等人 [57] 假设数据所有者的私人数据未知。

Zheng 等人 [55] 考虑一种双边市场，其中数据经纪人知道数据消费者的估值，但不知道数据所有者的估值。具体而言，他们考虑一种众包感知数据市场：某个地点设有多个数据采集点，每个采集点都有多位数据所有者。数据经纪人旨在从数据消费者处实现收入最大化，同时尽可能降低数据收集成本。该问题的解决方案分为两部分，一部分用于实现收入最大化，另一部分用于成本最小化。为实现收入最大化，假设已知每个数据采集点的预期支付，并提出一种贪心算法来选择最具成本效益的采集点。为降低成本，则通过拍卖来选择每个采集点的数据所有者。

Cao 等人 [56]考察了一个所有参与者的数据估值均未知的双边市场，并从系统设计者的角度处理这一问题。系统设计者旨在通过在数据所有者、数据经纪人和数据消费者之间合理分配数据资源，使社会福利最大化。这里，社会福利定义为数据消费者的效用，减去数据所有者因隐私损失而承受的不效用总和，以及数据经纪人的数据收集成本。作者提出了一种迭代拍卖来解决该问题。在迭代拍卖中，每当新一轮拍卖开始时，拍卖人都会公布数据分配、定价和补偿规则，以实现社会福利最大化。随后，每位竞标者，无论是数据所有者、数据消费者还是数据经纪人，都会计算出能够使自身效用最大化的报价，并提交报价及所需数量的数据。拍卖人根据提交的报价制定下一轮拍卖的规则。此任务被表述为一个约束优化问题。规则确定后，新一轮拍卖随即开始。拍卖持续进行，直到社会福利收敛。该拍卖机制被证明满足个体理性，且弱预算可行。

Jaisingh 等人 [11]考虑了一种不同的双边市场：数据市场嵌入在数字平台中，服务提供商为第三方数据消费者收集用户的数据。换言之，服务提供商是数据经纪人，而用户是数据所有者。一方面，数据经纪人需要决定数据收集方式和服务价格；另一方面，她还需要确定向数据消费者收取的价格。作者将合同应用于这样一个问题：由互联网服务提供商（ISP）充当数据经纪人。他们根据隐私成本高低，将互联网用户分为两类。针对用户，数据经纪人有三种数据收集和定价策略：（1）提供一份合同，规定在某项数据收集政策下的互联网服务价格；（2）提供一份不收集数据的互联网服务合同；（3）同时提供上述两种合同。最优策略取决于隐私损失、两类用户的隐私成本以及数据交易总收益之间的关系，这为互联网服务提供商提供了参考。

Bergemann 等人 [57] 也研究了这种双边数据市场。他们考虑了数据所有者数据之间的相关性。也就是说，可以从服务用户所收集的数据中推断出非用户的某些信息，因此非用户也会遭受隐私损失。作者提出了用户与数据经纪人之间，以及第三方与数据经纪人之间的双边合同。

合同规定了数据共享政策，即共享完整数据（包含身份信息）或聚合数据（不含匿名化信息），并规定了相应的价格。与数据生产者签订的最优合同是共享完整数据，并收取数据带来的利润增量；而与用户签订的最优合同则是收集匿名化数据，并向用户支付其个人效用与他人效用之差，这为数据经纪人设计合同提供了管理启示。 [58], [59]也讨论了类似的问题设定。

7.2 一次性查询和差分隐私下的双边市场

在一次性查询和差分隐私下的双边市场中，垄断数据经纪人向数据消费者出售其感兴趣的单个查询答案，并使用差分隐私量化隐私，从数据所有者处采购私人数据。根据预先设定的差分隐私要求，必须在查询答案中加入一定噪声，这会降低查询答案的准确性以及数据消费者获得的信息效用。因此，数据经纪人在为数据所有者设定数据采购价格时，应考虑对数据所有者的隐私保护与数据消费者的信息效用之间的权衡。隐私要求较高的数据所有者应获得较少的补偿，而精度要求较高的数据消费者则应支付更高的费用。此外，在买方市场中，数据所有者的隐私偏好存在异质性，因此愿意以不同价格出售数据。然而，估值虽是数据经纪人有用的信息，却是隐藏的。本类别中的所有研究均基于信息不完全假设。本类别的研究始于 Ghosh 和 Roth 的开创性研究 [13] ，并在 [14], [15], [16], [17], [60], [61], [62], [63], [64]中通过考察一次性查询的不同方面得到了扩展。

Ghosh 和 Roth [13] 将隐私视为定价单位（即，在 ε-差分隐私机制中，交易 ε 单位的隐私），并采用多单位采购拍卖。具体而言，他们研究了这样一个问题：数据经纪人希望通过从多个数据所有者处收集数据来回答计数查询，每位数据所有者都拥有一个私人数据条目 d ，并对隐私有一个估值 θ 。假设当数据在 ε-差分隐私下使用时，估值为 θ 的数据所有者的隐私成本 c 是一个线性函数，即 $c = \varepsilon\theta$。更好的隐私保护通常意味着查询结果的准确性较低，因此数据经纪人需要权衡隐私保护与查询准确性。作者探讨了在两种情形下如何设定采购价格：（1）在数据经纪人的预算约束下最大化准确性；（2）在准确性目标固定的情况下最小化总支付额。在第一种情形下，他们提出了 FairQuery 机制：先按隐私估值升序排序，然后在预算范围内尽可能多地选择数据所有者，并向每位入选的数据所有者支付统一的货币补偿。在第二种情形下，获胜者的数量由精度要求决定。最小成本拍卖机制按数据所有者的成本升序排序。这里的成本是隐私估值与隐私参数的乘积。每位入选数据所有者获得的补偿等于首位未入选者的成本。为保护隐私，分别向这两种机制加入从拉普拉斯分布中抽取的噪声。这些机制保证占优策略激励相容、个体理性、事后预算平衡和准确性。这项工作的管理启示是，要达到特定的隐私级别，数据经纪人需要至少从一定数量的隐私级别高于特定水平的数据所有者处购买数据。

Ghosh 和 Roth [13] 指出，当私人数据 d 与隐私估值 θ 之间存在未知相关性时，会出现一个负面结果：不存在既能保护隐私又能实现非平凡查询准确性的个体理性直接机制。相关性的一个例子是，糖尿病患者往往会对自己的健康数据赋予更高的估值。研究者通过采用间接机制 [14], [15], [60] 或设定额外假设 [61]，重新审视并绕过了这一不可能性结果。

Roth 和 Schoenebeck [14] 允许私人数据与隐私估值之间存在任意相关性，并在贝叶斯设定下重新审视 [13] 问题；在该设定中，估值分布是已知的先验分布。他们提出了一种要么接受要么放弃报价，并结合由某个分布 $\mathcal{G}$定义的彩票机制。对于每个数据所有者 i，从 $\mathcal{G}$中抽取一个随机价格 p_i 。当报告的估值 θ_i 低于 p_i时，使用该数据条目的概率为1，支付金额则是方差水平和报告估值的函数；否则，购买数据的概率为零，支付金额也为零。由于数据所有者的选择具有随机性，即使假设数据条目值与隐私估值之间存在相关性，也没有人能够根据某个数据条目是否被使用来推断任何私人信息。彩票机制带来的一个问题是，由于不同估值的数据所有者被选中的概率不同，查询答案可能会产生偏差。为了解决这一问题， [14]的作者采用霍维茨-汤普森估计量 [65], ，通过将数据价值乘以选择概率的倒数来进行调整，从而确保估计量无偏。 [14]中的机制保证了贝叶斯激励相容、事后个体理性、期中预算可行性和准确性。

同样，Fleischer 和 Lyu [15] 重新审视了不可能性结果，并聚焦于计数查询。为避免数据消费者因隐私估值与数据价值之间的相关性，仅凭个人的参与决策就推断出某些私人信息，个人接受报价的概率应不受其隐私估值影响，始终保持相同。他们假设，数据价值不同的个人所承担的隐私成本分布也不同，并且数据价值为 j 的个人的隐私估值累积分布函数是公开已知的，表示为 F_j。机制首先找出一个值 α_j ，使得$F_j(\alpha_j) = pr$，其中 pr 是预设阈值；然后向每个拥有数据 j 的个人设定预期支付金额 $\varepsilon\alpha_j$，其中 ε 是隐私参数。最后，机制根据收集到的数据进行计算，并在计算结果中加入拉普拉斯噪声。该机制保证贝叶斯激励相容、期中个体理性和准确性。

类似地，Ligett 和 Roth [60] 设计了一种合同，以规避不可能性结果。与 [14], [15],不同，他们不要求基于隐私估值分布的先验知识作出贝叶斯假设。他们提出的合同由三个要素构成：隐私价格、参与决策所对应的隐私保护级别，以及私人数据所对应的隐私保护级别。给定期望的准确性水平后，设定目标样本规模。向每位数据所有者提供报价，直到接受报价的数据所有者人数达到目标规模。随后，在从样本数据集中得出的计数查询答案中加入拉普拉斯噪声。

与通过采用间接机制规避不可能性结果的 [14], [15], [60] 不同，Nissim 等人[61] 提出了一种直接机制，但额外假设隐私估值具有单调性。具体而言，他们假设，拥有某种私人数据价值的数据所有者，通常会对自己的隐私赋予更高的估值。例如，患有某种疾病的人往往更不愿透露自己的医疗信息，因此需要获得更多报酬才愿意公开数据。他们提出的机制如下：给定n 个数据所有者、预算 B 和隐私保护级别ε，对于估值低于 $B/2\varepsilon n$的数据所有者，机制保持其数据不变，并向每位数据所有者支付 B/n；对于估值较高的数据所有者，机制则将其数据改为 0 ，且不向其支付任何报酬。最后，加入从对称几何分布中抽取的噪声。该机制对所有数据所有者均满足事后个体理性，对隐私估值较低的数据所有者满足占优策略个体理性，同时满足预算可行性和准确性。

上述研究，包括 [13], [15], [60], [61], ，探讨了特定查询类型——计数查询的定价问题。然而，现实世界中常用的查询并不限于计数查询。因此，[13] 的另一项扩展是考虑更一般的查询类型 [16], [62]。

Dandekar 等人 [16] 聚焦于线性预测器查询。线性预测器通过探索现有数据所有者与新用户之间的相似性，来预测推荐系统中新用户的行为。具体而言，使用线性预测器 $\varphi(D) = \sum_{i=1}^{n} w_i d_i$对由数据所有者$i$ 的个人数据 d_i 构成的数据集 D 进行评估，其中 w_i 是与 i 相关联的公开权重，用于描述相似性。数据所有者对隐私保护的要求各不相同，隐私估值也存在差异。权重越大，预测越准确，因此需要更多数据所有者参与。然而，这意味着需要支付更高的费用。于是，该定价问题被表述为一个最大化问题，目标是使权重总和最大化，同时满足预算和准确性的约束。该问题可视为一个 0/1 背包问题，并采用贪心方法求解。这项工作的主要见解与 [13]的隐私-准确度权衡相呼应。也就是说，对于线性预测器查询，要达到一定的准确性，隐私保护良好的数据所有者的权重最多只能占总权重的一定比例。

除了线性预测器之外，Zhang 等人 [62] 还允许采用其他常见的查询类型，例如计数查询和中位数查询。此外，作者重新审视了先前研究中的一项假设，即数据所有者愿意按照数据经纪人确定的任何隐私保护级别提供其数据，并假设数据所有者目标单一，只有在满足其隐私要求时才会出售数据。数据所有者向数据经纪人报告其隐私估值和隐私要求。随后，将定价问题表述为一个优化问题，旨在最大化查询准确性，同时满足激励相容（IC）、个体理性（IR）和预算约束，并通过应用 KKT 条件求解。

在 [13], 中，数据所有者可能会谎报其隐私估值，但不能谎报其数据条目，因为这里假设数据条目是可验证的。然而，在某些情况下，不存在真实数据集，数据经纪人也无法在数据收集后验证数据。 [17], [63], [64]研究了这种情形。

Ghosh等人 [17] 考虑了不可验证数据的交易，并设计了一种同伴预测机制。该机制最初用于如实收集一组专家对某些未来事件发生率的意见。不可验证数据被视为同伴预测机制中的意见，而数据经纪人希望根据数据所有者的意见预测事件是否会发生。作者设计了一种支付规则，使成本较低的数据所有者愿意参与该机制，并在其报告的数据与其他数据所有者得出的结果一致时，使其期望效用最大化。因此，该机制能够确保数据所有者如实报告私人数据。这项工作的主要见解是，数据经纪人可以利用同伴预测机制来解决数据不可验证性问题。

尽管数据不可验证， [17] 中的机制仍能激励数据所有者如实报告数据。Cummings等人 [63] 和 Wang等人 [64] 考虑了一个略有不同的情形：数据经纪人对数据所有者而言并不可信，且允许数据所有者报告噪声数据。

Wang等人 [64]采用差分隐私的本地模型，个人可以向自己的数据添加噪声，并以噪声数据换取货币补偿。他们考虑了一个特定情形：数据消费者对某个州感兴趣，而数据所有者掌握一些能够指示该州情况的信息，即一个二元信号。数据消费者的目标是以最低的期望支付获取准确的信息。他们设计了一种支付机制，要求所有个人报告自己的私人数据。如果某人的数据与其他大多数人的数据一致，她就会获得报酬；否则就会受到惩罚。例如，如果数据消费者报告1 ，而根据其他人的数据推断出的州为0，那么该数据消费者就必须支付惩罚金。在这一机制中，一种均衡是所有代理人都以固定概率报告正确的信号，因此该机制满足本地差分隐私。

Cummings等人 [63] 允许数据所有者提供带噪声的数据。数据经纪人为数据所有者提供多个不同方差水平的选项，并要求他们报告各个水平下的隐私估值，目标是通过选择合适的数据所有者及其对应的方差水平，将总支付额降至最低。所需机制的设计问题被表述为一个组合优化问题，其目标是在满足准确性约束的前提下最小化总成本。随后，该问题被转化为整数线性规划，并进一步松弛为线性规划。通过求解这一线性

规划，可以得到最优解，即最优分配规则，该规则表明数据所有者被分配到特定方差水平的概率。随后应用维克里－克拉克－格罗夫斯（VCG）支付规则 [66], [67],[68] 。

与上述考虑数据集特征的研究不同，Zhang 等人 [69] 考虑了正网络效应对数据所有者参与数据交易的影响。也就是说，数据市场中的参与者越多，数据所有者就越愿意加入，因此隐私成本会随着参与者数量的增加而降低。对于估值为 v_i 的数据所有者 i ，当其数据受到 ε_i-差分隐私保护时，成本函数建模为 $c_i = \varepsilon_i v_i k^{-t}$，其中 k 为参与者数量， $t \geq 0$ 是用于描述正网络效应的参数。他们设计了一种拍卖机制，首先将数据所有者随机划分为两组。然后，按出价从低到高对每组进行排序，并计算每组的最优利润。将较低的利润设为目标利润。对于利润较高的组，依次选取前 k 名数据所有者，直至达到目标利润。

7.3 一般查询和差分隐私下的双边市场

在一般查询和差分隐私下的双边市场中，垄断数据经纪人向数据消费者出售一般查询，并通过差分隐私从数据所有者处获取私人数据，以量化隐私。除了上一小节讨论的隐私保护与支付之间的权衡之外，潜在的套利条件使数据定价问题变得更加复杂。在买方方面，数据经纪人需要保护数据所有者的隐私，并针对隐私损失给予适当补偿；在卖方方面，数据经纪人需要设定合适的价格以体现信息价值，从而避免套利。[23],[70], [71], [72], [73]研究了这一问题。

Li 等人 [23] 考虑了一个双边市场，在该市场中，数据消费者可以发起多个线性查询，并指定精度要求。这里，精度要求以方差水平表示。当一项查询被执行多次时，方差会降低，隐私保护也会减弱。在隐私保护下进行多次查询的情形中，无套利性被重新定义：方差较高的查询，其价格应低于方差较低的查询。作者分别为数据市场的双方设计了定价函数和补偿函数。对于数据所有者，给定一个方差水平和数据消费者指定的一项查询，补偿金额与该查询函数对这位数据所有者的数据变化的敏感度成正比，与方差水平成反比。研究证明，该补偿函数具有无套利性。对于数据消费者，查询价格为所有相关数据所有者的补偿总和。只要查询的定价函数是次可加且非递增的，就可以对其进行调整，例如采用线性组合、最大值等形式。研究证明，这些形式的定价函数具有无套利性。本文的主要见解是：本文针对隐私保护数据市场提出了重新定义的无套利性概念，并为数据经纪人提供了实用的定价函数。

与 [23], 不同，在该研究中，数据消费者指定所期望的准确性，由数据经纪人计算最小价格；Nget 等人 [70]研究了这样一种数据市场：数据消费者拥有预算，而经纪人则返回最准确的答案。数据所有者的隐私偏好各不相同，并分为两种类型：保守型和自由型。为了兼顾这两种类型的数据所有者，他们提出了两个补偿函数。所提出的机制由三部分组成：随机选择一组数据所有者、计算补偿，以及选出准确性最高的样本。具体而言，该机制首先按照统计公式计算出的规模，选取一组数据所有者。第二部分中，给定数据消费者的预算，将款项平均分配给选中的数据所有者，并通过补偿函数的反函数确定所有入选数据所有者的隐私级别。根据隐私级别，在查询结果中加入噪声，并测量其准确性。该抽样过程会重复多次，机制将各样本得出的查询结果进行比较，并选择准确性最高的结果。

Niu 等人 [71], [72]考虑了一个双边数据市场，其中个体数据条目之间可能存在相关性。换言之，披露一个数据条目可能导致泄露其他数据所有者的隐私。因此，即使某个数据所有者的数据条目未被使用，只要该条目与所选数据条目相关，她就应获得补偿。补偿函数应满足无套利和依赖公平性。这里，依赖公平性要求：当某个数据所有者及其相关数据所有者均未参与时，该数据所有者获得的补偿应为零。作者提出了自底向上设计和自顶向下设计，两者都会分别确定数据消费者应支付的价格和数据所有者应获得的补偿。自底向上设计首先计算隐私补偿，然后确定数据消费者应支付的价格；自顶向下设计则反之。在自底向上设计中，采用了与 [23] 类似的定价方案。不同之处在于补偿函数。由于数据条目之间存在相关性，他们采用依赖差分隐私，而非差分隐私，来量化依赖数据集上的隐私损失。补偿与依赖敏感度成正比。这里，依赖敏感度定义为修改一个数据条目及其相关数据条目所导致的函数值变化。在自顶向下设计中，给定数据消费者的预算，每个为查询答案作出贡献的数据所有者所获得的补偿，等于其隐私损失占所有参与的数据所有者隐私损失总和的比例。该研究在 [73] 中进一步扩展，考察交易时间序列数据时，时间序列数据的时间相关性，例如对心率和身体活动进行连续监测。Niu 等人 [73] 采用自底向上设计，并运用 Pufferfish 隐私 [74] 的概念来量化时间相关性下的隐私。

参考文献	市场结构	隐私概念	查询类型	定价方法
[35]	卖方	–	空查询	按次付费、分层定价
[40]	卖方	–	空查询	版本控制
[12]	卖方	–	空查询	拍卖
[45]	卖方	–	空查询	激励机制
[42]	卖方	–	空查询	合同
[47]	卖方	–	空查询	合同
[19]	卖方	–	一般查询	版本控制
[20]	卖方	–	一般查询	版本控制
[21]	卖方	–	一般查询	版本控制
[49], [50]	卖方	–	一般查询	版本控制
[52]	卖方	–	一般查询	版本控制
[53]	卖方	–	一般查询	版本控制
[51]	卖方	–	一般查询	版本控制
[22]	卖方	–	一般查询	版本控制
[10]	买方	临时	–	版本控制
[9]	买方	临时	–	拍卖
[39]	买方	临时	–	合同
[55]	集中式双边	临时	空查询	拍卖
[56]	集中式双边	临时	空查询	拍卖
[11]	集中式双边	临时	空查询	合同
[57]	集中式双边	临时	空查询	合同
[13]	集中式双边	DP	一次性查询	拍卖
[14]	集中式双边	DP	一次性查询	激励机制
[15]	集中式双边	DP	一次性查询	激励机制
[60]	集中式双边	DP	一次性查询	合同
[61]	集中式双边	DP	一次性查询	拍卖
[16]	集中式双边	DP	一次性查询	拍卖
[62]	集中式双边	DP	一次性查询	拍卖
[17]	集中式双边	DP	一次性查询	激励机制
[64]	集中式双边	DP	一次性查询	激励机制
[63]	集中式双边	DP	一次性查询	激励机制
[69]	集中式双边	DP	一次性查询	拍卖
[23]	集中式双边	DP	一般查询	版本控制
[70]	集中式双边	DP	一般查询	激励机制
[73]	集中式双边	DP	一般查询	版本控制
[71], [72]	集中式双边	DP	通用 query	版本控制

表2 文献综述摘要

8 隐私经济学的进一步讨论

隐私的经济价值已在经济学领域得到广泛的理论和实证研究。经济学研究的重点是通过分析不同情境下隐私披露或隐私保护所产生的后果，展现私人数据的价值，并为私人数据收集和定价提供管理启示。

私人数据在决策中的价值。Posner[75] 认为，保护私人数据可能会导致市场效率低下。也就是说，保护私人数据会使某些可能有用的信息无法获取，从而可能对需要这些信息的个人或组织的决策产生负面影响，进而导致市场效率低下。Stigler [76] 也提出了类似的观点。由于人们往往倾向于保护其负面信息的隐私，基于有偏信息作出的决策会导致效率低下。

私人数据对消费者的价值。Varian [77]指出，作为数据所有者的消费者希望生产者了解其某些类型的私人数据，例如他们的偏好，因为这有助于降低搜索成本；但他们希望对另一些类型的私人数据保密，例如其支付意愿。Ali 等人 [78], [79] 指出，私人数据收集对消费者有益。

私人数据对在线用户的价值。Friedman 等人 [80] 研究了在线环境中的隐私价值。在这种环境中，用户可以通过不断更换标识符来保护隐私。匿名用户难以获得信任并与他人建立合作，从而造成社会成本。Bapna 等人 [81] 考察了在线交友网站，用户可以匿名查看他人的个人资料。实证结果表明，匿名用户的匹配对象少于非匿名用户。Grad wohl [82] 研究了隐私保护对社交网络用户的影响。社交网络用户一方面享有参与网络所带来的网络效应，另一方面又因彼此之间的相互依存而遭受隐私损失。作者指出，在此类社交网络中，隐私保护能够提升社会福利。

私人数据在价格歧视中的价值。消费者的电子足迹（例如购买历史）有助于动态定价，这是一种价格歧视。Rossi 等人 [83] 表明，使用购买历史能够显著增加公司的收入。 [84], [85], [86]也得出了类似结论。Conitzer 等人 [87]表明，将购买历史用于价格歧视能够提升社会福利。

合同中私人数据的价值。Calzolari 和 Pavan [88] 考察了两个委托人先后与一名代理人签订合同的环境，并表明，三方中

个人数据的共享能够提升社会福利。Liang 和 Marsden [89] 考察了一名委托人与一名代理人之间的合同，其中有两类私人数据可供使用，并表明，披露其中一类数据会提高社会福利，而披露另一类数据则会降低社会福利。

广告中私人数据的价值。De Corniere 和 De Nijs [90] 考察了消费者私人数据在赞助搜索拍卖中的价值，并表明，通过向广告商披露消费者数据，广告平台能够提高广告与消费者需求之间的匹配质量。Goldfarb 和 Tucker [91]开展了调查，结果显示，隐私法规的实施显著降低了广告的有效性。

上述研究通过分析隐私变化的影响，揭示了私人数据的价值。另一些研究则通过明确量化私人数据的货币价值来进行分析。例如，Hann 等人 [92] 表明，美国的数据所有者认为隐私保护的价值介于$30.49 和$44.62之间。 [93], [94], [95] 也研究了隐私保护的价值。Huberman 等人 [96]表明，体重数据和年龄数据的平均货币价值分别为$57.56 和$74.06。随后， [97], [98],[99], [100], [101] 考察了位置数据、浏览历史、移动数据及其他类型私人数据的货币价值。

9 未来机遇

除现有的数据定价方法外，仍然存在以下所述的挑战与机遇：

（1）数据定价问题已受到众多不同领域的关注，包括但不限于经济学、数据库与数据管理、《管理科学》和计算机科学。数据定价问题往往较为复杂，尤其是在数据市场中存在信息不对称的情况下，这会使传统分析方法失效或效率低下。此外，数据交易并非一次性业务，而是需要动态定价，这进一步增加了数据定价问题的复杂性。因此，机器学习和强化学习是解决这些问题的强大辅助工具。 [102] 和 [103]分别利用强化学习来确定最优固定费用和最优合同。机器学习和强化学习有望应用于其他定价方法。

（2）现有大多数关于数据定价的研究都聚焦于原始数据集和查询答案的交易。数据的核心价值在于提供洞见，从而帮助人们做出更好的决策；而利用机器学习进行数据分析，已广泛应用于科学、商业智能和许多其他领域。交易物品不应局限于原始数据集或简单查询，而可以是基于丰富数据训练的学习模型。Chen 等人 [104]提出了适用于此类数据交易的基于模型的定价框架。然而，只有少数研究对此进行了探讨，例如 [104], [105], [106], [107], [108]。基于模型的定价面临的一个关键问题是如何评估学习模型的质量以及其对数据消费者的估值，这对于为学习模型设定合理的价格至关重要，也是有待未来研究的问题。

（3）数据可以分为三种类型：结构化数据、非结构化数据和半结构化数据。结构化数据通常以关系表或统计数据的形式表示，而非结构化数据则是一系列符号，通常以自然语言编码。半结构化数据具有一定的结构，同时又具备一定程度的灵活性，例如 XML 中的数据 [43]。大多数关于数据定价的研究都聚焦于结构化数据，很少讨论非结构化数据的情况，例如自然语言格式的数据、社交网络数据和交通网络数据。然而，后者是自然语言处理和图表示学习等任务的重要输入，因此需要一种能够适应非结构化数据和半结构化数据的数据交易框架。

(4) 随着隐私增强技术（PETs）迅速发展，其应用和部署日益普及。目前有关数据定价的研究采用了差分隐私和 k-匿名的概念。实践中部署的许多其他隐私增强技术，例如同态加密、多方计算、联邦分析和假名化，在这一领域尚缺乏研究。这些技术以各种方式实现隐私保护并定义隐私，因此，在针对不同隐私定义设计数据定价框架时，需要特别加以关注。

(5) 数据消费者（或数据所有者）的支付意愿（或接受意愿）以及需求（或供给）难以建模。对数据消费者而言，数据属于体验品。只有在购买数据并从中获取有用信息之后，数据消费者才能估算所购数据的价值。换言之，数据消费者在做出购买决策时，很难预先判断数据的价值并表达其支付意愿。对数据所有者而言，由于隐私是无形的，他们很少会量化私人数据的价值。因此，他们在做出出售决策时，也很难表达自己的接受意愿。 [92], [96] 等研究通过实证方法考察了私人数据的货币价值，为数据所有者的接受意愿和供给建模提供了参考。还需要从数据消费者的角度研究数据的价值。

(6) 现有研究大多只关注补偿数据所有者隐私损失的成本，而忽略了其他成本，例如数据整合、存储和维护成本；尤其对于大规模数据而言，这些成本不容忽视。这些不可忽视的成本可能会影响从数据所有者处收集数据以及向数据消费者进行数据定价，因此应在未来的研究中予以考虑。

10 结论

数据需求的不断增长催生了一种新型商业模式——数据市场，并向研究人员和创业者提出了一个问题：如何为数据合理定价？为解决这一问题，许多学者从不同角度提出了数据定价方法。本文构建了数据定价方法的分类体系，并对现有数据定价文献进行了系统性综述。

我们首先介绍数据市场的基础知识，然后概述基本的定价方法。随后，我们提出一种新的数据定价研究分类体系，重点关注数据的内在价值。数据在数据市场的双方各有不同意义上的价值，因此市场结构构成了分类体系的基础。对于不同的查询类型和隐私概念，数据价值的衡量方式也各不相同。因此，我们分别将查询类型和隐私概念作为卖方和买方的分类标准。依据该分类体系，我们详细回顾各类别中现有的数据定价研究，并阐述各类别定价方法的适用性与局限性。最后，我们概述数据定价领域面临的挑战与机遇。

总而言之，本调查细致地阐述了数据定价方法的研究现状。这有助于我们理解数据市场中的定价理论，并为如何在实践中运用这些理论提供管理指导。本调查或许能发挥重要作用，指出有助于弥合学术界提出的数据定价方法与业界已采用的现行定价实践之间差距的桥梁。在实践中，数据市场的定价主要依赖简单的定价方法，例如固定费用、高级定价和线性定价。几乎总是，在确定单位数据价格后，数据数量便是决定数据消费者所支付价格的唯一因素。本调查广泛讨论的其他因素，例如隐私、安全和消费者效用，除数据公司交易个人数据的情况外，很少受到考虑。

参考文献

[1] J. Manyika、M. Chui、B. Brown、J. Bughin、R. Dobbs、C. Roxburgh 和 A. H. Byers，《大数据：创新与竞争的下一个前沿》，技术报告，以及生产力。技术报告，麦肯锡全球研究院，2011年。[2] M. Fisher 和 A. Raman，《在零售业中利用数据和大数据》，《生产与运营管理》，第27卷，第9期，第1665–1669页，2018年。[3] M. Maciejewski，《做得更多、更好、更快且成本更低：在公共管理中利用大数据》，《国际行政科学评论》，第83卷，第1期增刊，第120–135页，2017年。[4] S. Yin 和 O. Kaynak,–《现代工业中的大数据：挑战与趋势 [观点],》《*IEEE*会刊》，第103卷，第2期，第143–146页，2015年。[5] F. Stahl、F. Schomm、G. Vossen 和 L. Vomfell，《数据市场的分类框架》，《越南计算机科学杂志》，第3卷，第3期，第137–143页，2016年。[6] C. Dwork，《差分隐私》，载于《第*33*届自动机、语言与编程国际会议论文集——第二卷》，2006年。[7] C. Dwork、A. Roth、等，《差分隐私的算法基础》，《理论计算机科学基础与趋势》，第9卷，第3–4期，第211–407页，2014年。[8] G. Cormode、S. Jha、T. Kulkarni、N. Li、D. Srivastava 和 T. Wang，《大规模隐私保护：本地差分隐私的实践》，载于《*2018*年数据管理国际会议论文集》，第1655–1658页，2018年。[9] V. Gkatzelis、C. Aperjis 和 B. A. Huberman，《私人数据定价》，《电子市场》，第25卷，第2期，第109–123页，2015年。[10] J. Parra-Arnau，《个人数据市场中隐私的优化直接销售》，《信息科学》，第424卷，384，2018年354页。[11] J. Jaisingh、J. Barron、S. Mehta 和 A. Chaturvedi，《个人信息的隐私与定价》，《欧洲运筹学杂志》，第187卷，第3期，第857–870页，2008年。

[12] C. Riederer、V. Erramilli、A. Chaintreau、B. Krishnamurthy 和 P. Rodriguez，《出售：你的数据，由你出售》，载于第十届ACM网络热点主题研讨会论文集，第1–6页，2011年。[13] A. Ghosh 和 A. Roth，《拍卖隐私》，载于第12届ACM电子商务会议论文集，第208–199页，2011年。[14] A. Roth 和 G. Schoenebeck，《以低成本开展真实调查》，载于第13届ACM电子商务会议论文集，第826–843页，2012年。[15] L. K. Fleischer 和 Y.-H. Lyu，《当成本与数据相关时用于出售隐私的近似最优拍卖》，载于第13届ACM电子商务会议论文集，第568–585页，2012年。[16] P. Dandekar、N. Fawaz 和 S. Ioannidis，《推荐系统的隐私拍卖》，载于第八届互联网与网络经济国际会议论文集，第309–322页，施普林格出版社，2012年。[17] A. Ghosh、K. Ligett、A. Roth 和 G. Schoenebeck，《无需验证即可购买私人数据》，载于第十五届ACM经济学与计算会议论文集，第931–948页，2014年。[18] M. Balazinska、B. Howe 和 D. Suciu，《云端数据市场：数据库社区的机遇》，《VLDB基金会会刊》，第4卷，第12期，第1482–1485页，2011年。[19] P. Koutris、P. Upadhyaya、M. Balazinska、B. Howe 和 D. Suciu，《基于查询的数据定价》，载于第31届ACM数据库系统原理研讨会论文集，第167–178页，2012年。[20] B. R. Lin 和 D. Kifer，《一般数据查询的无套利定价》，《VLDB基金会会刊》，第7卷，第9期，第757–768页，2014年。[21] S. Deep 和 P. Koutris，《无套利数据定价方案的设计》，2016年。[22] R. Tang、H. Wu、Z. Bao、S. Bressan 和 P. Valduriez，《价格正合适》，载于数据库与专家系统应用国际会议，2013年。[23] C. Li、D. Y. Li、G. Miklau 和 D. Suciu，《隐私数据定价理论》，《ACM数据库系统汇刊》，第39卷，第4期，第1–28页，2014年。[24] A. Muschalle、F. Stahl、A. Loser 和 G. Vossen，《数据市场的定价方法》，载于实时企业商业智能国际研讨会，第129–144页，施普林格出版社，2012年。[25] F. Schomm、F. Stahl 和 G. Vossen，《数据市场：初步调查》，《ACM SIGMOD记录》，第42卷，第15–26页，2013年。[26] F. Stahl、F. Schomm 和 G. Vossen，《重新审视数据市场调查》，技术报告，ERCIS工作论文，2014年。[27] F. Liang、W. Yu、D. An、Q. Yang、X. Fu 和 W. Zhao，《大数据市场调查：定价、交易与保护》，IEEE Access，第6卷，第15132–15154页，2018年。[28] A. Wagner、N. Wessels、P. Buxmann 和 H. Krasnova，《为个人信息标价——文献综述》，载于第51届夏威夷国际系统科学会议论文集，2018年。[29] S. A. Fricker 和 Y. V. Maksimov，《数据市场中数据产品的定价》，载于软件业务国际会议，第49–66页，施普林格出版社，2017年。[30] J. Pei，《数据定价调查：从经济学到数据科学》，IEEE知识与数据工程汇刊，2020年。[31] A. A. Armstrong 和 E. H. Durfee，《混合与记忆：信息市场中的涌现合作》，载于自主智能体与多智能体系统国际会议论文集（目录号98EX160），第34–41页，IEEE，1998年。[32] F. Consulting，《数字化正在推动下一代数据市场的发展》，2017年。[33] C. Shapiro 和 H. R. Varian，《信息规则：网络经济战略指南》，哈佛商学院出版社，1998年。[34] O. Shy，《如何定价：定价技术与收益管理指南》，剑桥大学出版社，2008年。[35] A. Kushal、S. Moorthy 和 V. Kumar，《数据市场定价》，技术报告，2012年。[36] C. Shapiro 和 H. R. Varian，《版本控制：销售信息的明智之道》，《哈佛商业评论》，第76卷，第6期，第107–115页，1998年。[37] T. Borgers 和 D. Krahmer，《机制设计理论导论》，牛津大学出版社（美国），2015年。[38] R. P. McAfee 和 J. McMillan，《拍卖与竞价》，《经济文献杂志》，第25卷，第2期，第699–738页，1987年。

[39] L. Xu、C. Jiang、Y. Chen、Y. Ren 和 K. J. R. Liu，《数据收集中的隐私还是效用？一种契约理论方法》，IEEE《信号处理选题期刊》，第9卷，第7期，第1256–1269页，2015年。[40] H. Yu 和 M. Zhang，《基于数据质量的数据定价策略》，《计算机与工业工程》，第112卷，10月号，第1–10页，2017年。[41] S. Mehta、M. Dawande、G. Janakiraman 和 V. Mookerjee，《如何销售数据集？数据变现的定价政策》，载于2019年ACM经济学与计算会议论文集，第679–679页，2019年。[42] X.-B. Li 和 S. Raghunathan，《兼顾隐私意识的客户数据定价与传播》，《决策支持系统》，第59卷，第63–73页，2014年。[43] C. Batini、C. Cappiello、C. Francalanci 和 A. Maurino，《数据质量评估与改进方法》，《ACM计算综述》，第41卷，第3期，第1–52页，2009年。[44] F. McSherry 和 K. Talwar，《通过差分隐私进行机制设计》，载于第48届IEEE计算机科学基础年会（FOCS'07），第94–103页，IEEE，2007年。[45] S. Mehta、M. Dawande、G. Janakiraman 和 V. Mookerjee，《如何销售数据集？数据变现的定价政策》，《信息系统研究》，第32卷，第4期，第1281–1297页，2021年。[46] C. Daskalakis、A. Deckelbaum 和 C. Tzamos，《最优机制设计的复杂性》，载于第二十五届ACM-SIAM离散算法研讨会论文集，第1302–1318页，SIAM，2014年。[47] P. Naghizadeh 和 A. Sinha，《隐私数据商业化的对抗性契约设计》，载于2019年ACM经济学与计算会议论文集，第681–699页，2019年。[48] M. Balazinska、B. Howe、P. Koutris、S. Dan 和 P. Upadhyaya，《关于关系数据定价的讨论》，Springer Berlin Heidelberg，2013年。[49] S. Deep 和 P. Koutris，《Qirana：可扩展查询定价框架》，载于2017年ACM数据管理国际会议论文集，第699–713页，2017年。[50] S. Deep、P. Koutris 和 Y. Bidasaria，《Qirana演示：实时可扩展查询定价》，《VLDB基金会会刊》，第10卷，第12期，第1949–1952页，2017年。[51] C. Xia 和 S. Muthukrishnan，《基于用户的市场中的无套利定价》，载于第17届自主智能体与多智能体系统国际会议论文集，第327–335页，2018年。[52] P. Upadhyaya、M. Balazinska 和 D. Suciu，《利用数据API进行价格最优查询》，《VLDB基金会会刊》，第9卷，第14期，第1695–1706页，2016年。[53] S. Chawla、S. Deep、P. Koutris 和 Y. Teng，《查询定价的收入最大化》，arXiv预印本 arXiv:1909.00845，2019年。[54] P. Buneman、S. Khanna 和 W. C. Tan，《为何与何处：数据溯源的特征描述》，载于数据库理论国际会议，2004年。[55] Z. Zheng、Y. Peng、F. Wu、S. Tang 和 G. Chen，《众包群体中的数据交易：面向移动众感知的利润驱动数据采集》，IEEE《通信选题领域期刊》，第35卷，第2期，第486–501页，2017年。[56] X. Cao、C. Yan 和 K. J. R. Liu，《数据交易的迭代拍卖机制》，载于ICASSP 2017——2017年IEEE声学、语音与信号处理国际会议（ICASSP），2017年。[57] D. Bergemann、A. Bonatti 和 T. Gan，《社交数据经济学》，《兰德经济学杂志》，2022年。[58] S. Ichihashi，《数据外部性的经济学》，《经济理论期刊》，第196卷，第105316页，2021年。[59] D. Acemoglu、A. Makhdoumi、A. Malekian 和 A. Ozdaglar，《数据过多：数据市场中的价格与低效率》，《美国经济学杂志：微观经济学》，第14卷，第4期，第218–56页，2022年。[60] K. Ligett 和 A. Roth，《接受或放弃：隐私有代价时如何开展调查》，载于互联网与网络经济国际研讨会，第378–391页，施普林格出版社，2012年。[61] K. Nissim、S. Vadhan 和 D. Xiao，《重新划定从隐私敏感个体处购买数据的边界》，载于第五届理论计算机科学创新会议论文集，第411–422页，2014年。[62] M. Zhang、F. Beltran 和 J. Liu，《隐私约束下的数据拍卖销售》，载于人工智能不确定性会议，第669–678页，PMLR，2020年。[63] R. Cummings、K. Ligett、A. Roth、Z. S. Wu 和 J. Ziani，《准确性待售：在方差约束下聚合数据》，in *2015*年理论创新会议论文集
计算机科学，第317–324页，2015年。[64] W. Wang、L. Ying和J. Zhang，"向注重隐私的个人购买数据：负向支付的影响"，载于《网络与互联网经济国际会议》，第87–101页，施普林格，2016年。[65] D. G. Horvitz和D. J. Thompson，"从有限总体中进行不放回抽样的推广"，《美国统计学会会刊》，第47卷，第260期，第663–685页，1952年。[66] W. Vickrey，"反投机、拍卖与竞争性密封投标"，《金融学杂志》，第16卷，1961年，第37–44页，第1、8期。[67] E. H. Clarke，"公共物品的多部定价"，《公共选择》，第11卷，第1期，第17–33页，1971年。[68] T. Groves，"团队激励"，《计量经济学：计量经济学会期刊》，第617–631页，1973年。[69] M. Zhang、L. Yang、X. Gong和J. Zhang，"隐私保护众包感知：隐私估值、网络效应与利润最大化"，载于2016年IEEE全球通信大会（GLOBECOM），第1–6页，IEEE，2016年。[70] R. Nget、Y. Cao和M. Yoshikawa，"如何通过个人数据市场中的定价机制平衡隐私与金钱"，2017年。[71] C. Niu、Z. Zheng、F. Wu、S. Tang、X. Gao和G. Chen，"释放隐私的价值：交易私人相关数据的聚合统计量"，载于第24届ACM SIGKDD知识发现与数据挖掘国际会议论文集，第2031–2040页，2018年。[72] C. Niu、Z. Zheng、F. Wu、S. Tang、X. Gao和G. Chen，"Erato：交易私人相关数据的带噪聚合统计数据"，IEEE知识与数据工程汇刊，2019年。[73] C. Niu、Z. Zheng、S. Tang、X. Gao和F. Wu，"小传感器赚大钱：在Pufferfish隐私下交易时间序列数据"，载于IEEE INFOCOM 2019——IEEE计算机通信会议，第568–576页，IEEE，2019年。[74] D. Kifer和A. Machanavajjhala，"严谨且可定制的隐私框架"，载于第31届ACM SIGMOD-SIGACT-SIGAI数据库系统原理研讨会论文集，第77–88页，2012年。[75] R. A. Posner，"隐私经济学"，《美国经济评论》，第71卷，第2期，第405–409页，1981年。[76] G. J. Stigler，"经济与政治中的隐私导论"，《法律研究杂志》，第9卷，1980年，第644–623页，第4期。[77] H. R. Varian，"个人隐私的经济学方面"，1996年。[78] S. N. Ali、G. Lewis和S. Vasserman，"自愿披露与个性化定价"，载于第21届ACM经济学与计算会议论文集，第537–538页，2020年。[79] I. P. Fainmesser、A. Galeotti和R. Momot，"数字隐私"，《管理科学》，2022年。[80] E. J. Friedman和P. Resnick，"低成本化名的社会成本"，《经济学与管理战略杂志》，第10卷，第2期，第173–199页，2001年。[81] R. Bapna、J. Ramaprasad、G. Shmueli和A. Umyarov，"在线约会中的单向镜：一项随机田野实验"，《管理科学》，第62卷，第11期，第3100–3122页，2016年。[82] R. Gradwohl，"网络中的信息共享与隐私"，载于2017年ACM经济学与计算会议论文集，第349–350页，2017年。[83] P. E. Rossi、R. E. McCulloch和G. M. Allenby，"购买历史数据在目标营销中的价值"，《营销科学》，第15卷，第4期，第321–340页，1996年。[84] Y. Chen和Z. J. Zhang，"针对策略性消费者的动态目标定价"，《国际产业组织杂志》，第27卷，第1期，第43–50页，2009年。[85] A. Acquisti和H. R. Varian，"根据购买历史制定价格"，《营销科学》，第24卷，第3期，第367–381页，2005年。[86] R. Cummings、K. Ligett、M. M. Pai和A. Roth，"均衡模型中的隐私：一个奇特案例"，载于2016年ACM经济学与计算会议论文集，第659–659页，2016年。[87] V. Conitzer、C. R. Taylor和L. Wagman，"躲猫猫：重复购买市场中有代价的消费者隐私"，《营销科学》，第31卷，第2期，第277–292页，2012年。[88] G. Calzolari和A. Pavan，"序贯合约中隐私的最优性"，《经济理论期刊》，第130卷，第1期，第168–204页，2006年。

[89] A. Liang和E. Madsen，“数据与激励”，载于第*21*届42，2020 41届*ACM*经济学与计算会议，第–页。[90] A. De Corniere和R. De Nijs，“在线广告与隐私”，《兰德经济学杂志》，第47卷，第1期，第48–72页，2016年。[91] A. Goldfarb和C. E. Tucker，“隐私监管与在线广告”，《管理科学》，第57卷，第1期，第57–71页，2011年。[92] I.-H. Hann、K.-L. Hui、S.-Y. T. Lee和I. P. Png，“克服在线信息隐私顾虑：信息处理理论方法”，《管理信息系统杂志》，第24卷，第2期，第13–42页，2007年。[93] E. Rose，“数据使用者与数据主体：消费者是否愿意为个人信息的产权付费？”，载于第*38*届夏威夷国际系统科学年会论文集，第180c–180c页，IEEE，2005年。[94] J. Y. Tsai、S. Egelman、L. Cranor和A. Acquisti，“在线隐私信息对购买行为的影响：一项实验研究”，《信息系统研究》，第22卷，第2期，第254–268页，2011年。[95] L. Olejnik、T. Minh-Dung和C. Castelluccia，“拍卖隐私”，2013年。[96] B. A. Huberman、E. Adar和L. R. Fine，“隐私估值”，《*IEEE*安全与隐私》，第3卷，第5期，第22–25页，2005年。[97] G. Danezis、S. Lewis和R. J. Anderson，“位置隐私价值几何？”，载于*WEIS*，第5卷，第56页，Citeseer，2005年。[98] D. Cvrcek、M. Kumpost、V. Matyas和G. Danezis，“位置隐私价值研究”，载于第*5*届*ACM*电子社会隐私研讨会论文集，第109–118页，2006年。[99] J. P. Carrascal、C. Riederer、V. Erramilli、M. Cherubini和R. De Oliveira，“巨无霸汉堡消费者的浏览行为：在线个人信息经济学”，载于第*22*届万维网国际会议论文集，第189–200页，2013年。[100] J. Staiano、N. Oliver、B. Lepri、R. de Oliveira、M. Caraviello和N. Sebe，“货币流动：一项以人为中心的个人移动数据经济学研究”，载于*2014*年*ACM*普适与泛在计算国际联合会议论文集，第583–594页，2014年。[101] M. Zhang和F. Beltr´an，“在数据交易市场中发现私人数据价值的实验研究”，2022年。[102] X. Chen、Y. Yang、L. Xu、M. Lei和W. Bi，“基于强化学习的动态隐私定价”，载于*2018*年大数据与计算国际会议论文集，第59–64页，2018年。[103] L. Xu、C. Jiang、Y. Qian、Y. Zhao、J. Li和Y. Ren，“动态隐私定价：一种采用时变奖励的多臂老虎机方法”，《*IEEE*信息取证与安全汇刊》，第12卷，第2期，第271–285页，2016年。[104] L. Chen、P. Koutris和A. Kumar，“迈向数据交易市场中机器学习的基于模型定价”，载于*2019*年数据管理国际会议论文集，第1535–1552页，2019年。[105] Z. Zheng、Y. Peng、F. Wu、S. Tang和G. Chen，“移动群智感知数据市场的在线定价机制”，载于第*18*届*ACM*移动自组网与计算国际研讨会论文集，第1–10页，2017年。[106] Z. Zheng、Y. Peng、F. Wu、S. Tang和G. Chen，“Arete：移动数据市场联合在线定价与奖励共享机制的设计”，《*IEEE*移动计算汇刊》，第19卷，第4期，第769–787页，2019年。[107] A. Agarwal、M. Dahleh和T. Sarkar，“数据交易市场：一种算法解决方案”，载于*2019*年*ACM*经济学与计算会议论文集，第701–726页，2019年。[108] J. Abernethy、Y. Chen、C.-J. Ho和B. Waggoner，“通过主动数据采购实现低成本学习”，载于第十六届*ACM*经济学与计算会议论文集，第619–636页，2015年。



张梦晓 (mengx- iao.zhang@uestc.edu.cn)是电子科技大学算法与逻辑研究实验室的研究助理。她的研究兴趣包括算法机制设计和安全与隐私经济学。



Fernando Beltr´an(f.beltran@auckland.ac.nz)是新西兰奥克兰大学信息系统与运营管理系副教授。Fernando获得了哥伦比亚安第斯大学电气工程学士学位，并在纽约州立大学石溪分校获得应用数学博士学位。他目前的研究方向包括数据隐私保护与数据市场的经济学、竞争的策略模型

以及无线电频谱共享中的合作，还有由网络数据可视化支持的沉浸式分析。BELTRAN 博士目前负责 Aroaro研究空间，即奥克兰大学数据可视化实验室。



刘家谋（jiamou.liu@auckland.ac.nz）是新西兰奥克兰大学计算机科学学院的高级讲师。他于2010年在奥克兰大学获得计算机科学学士和博士学位，之后先后在德国莱比锡大学担任副研究员，并在新西兰奥克兰理工大学担任高级讲师。刘博士的研究兴趣涵盖人工智能、多智能体系统和计算社会网络。