

基于隐私度量的数据定价模型

彭慧波, 周亚建

(北京邮电大学网络空间安全学院, 北京 100876)

摘 要: 有价值的资源充分流动, 可以实现大数据产业的长期、稳定和可持续发展。本文讨论了通过构建以合理定价为中心的数据交易机制来利用经济杠杆来激励数据流的想法。首先, 本文结合信息熵, 提出了一种基于分级的交易数据集隐私度量方法; 其次, 基于数据集中的隐私含量和数据引用指数提出了一种数据定价模型对待交易数据元组进行定价; 最后, 结合一个实际的交易数据集验证了模型的正确性、合理性、有效性。

关键词: 计算机应用技术; 数据定价; 隐私度量; 信息熵; 隐私分级

中图分类号: TP399 **文献标识码:** A **DOI:** 10.3969/j.issn.1003-6970.2019.01.012

本文著录格式: 彭慧波, 周亚建. 基于隐私度量的数据定价模型[J]. 软件, 2019, 40 (1): 57-62

A Data Pricing Model Based on Privacy Measurement

PENG Hui-bo, ZHOU Ya-jian

(Beijing University of Posts and Telecommunications School of Cyberspace Security, Beijing 100876, China)

【Abstract】: With valuable data resources flowing fully, the long-term, stable and sustainable development of the big data industry be realized. The idea of using economic levers to motivate data flow through the construction of a data transaction mechanism centered on reasonable pricing has been discussed in this paper. Firstly, this paper proposes a classification-based transaction data set privacy measurement method based on information entropy. Secondly, based on the privacy content and data index of the data set, a data pricing model is proposed to price the transaction data tuple. Finally, a specific example has been used to verify the correctness, rationality and effectiveness of the pricing model.

【Key words】: Computer application technology; Data pricing; Privacy measurement; Information entropy; Privacy rating

0 引言

随着信息技术的快速发展, 时刻都有海量数据的产生, 但是这些数据都是停止不动的, 形成了一个信息孤岛。数据作为一种在一定程度上不可共享的资源, 逐渐演变成为一种可进行交易的商品, 数据资源成为人类社会一种必不可少的生产要素与战略资产^[1]。为了促进数据的流动这就需要数据交易的支持。但是, 由于缺乏规范的交易渠道和统一的交易规范, 现在并没有统一的数据交易平台。而数据交易的平稳、健康发展又离不开数据定价方法的支持。目前仍未能在全球领域内形成对数据定价方法的统一认识。因此, 如何实现数据定价理论中定价机制的公开化、数据价格的透明化, 一直是困扰国内外科研人员以及大数据从业者的重大问题。

要解决这个问题, 隐私风险分析及评估不失为

一种可行解决方案。但隐私风险分析及评估, 尤其是量化隐私风险, 势必会涉及隐私度量问题。从这些分析来看, 隐私度量的研究具有十分重要的理论意义和应用价值。目前对于隐私度量的研究主要是基于 Shannon 信息论的通信框架, 提出了几种隐私保护信息熵模型, 以解决隐私保护系统的相关度量问题。信息熵作为信息度量的有效工具, 在通信领域已展现出其重要的贡献。为此, 不少学者或多或少进行了探索, 提出了比如事件熵、匿名集合熵、条件熵形式化分析和分布概率数学理论等在内的相关理论^[2]。然而这些方法在实际应用中并不能达到理论上所提到的隐私保护的效果, 并且存在不同程度的不足或缺陷。此外, 目前关于隐私度量的研究还较为零散, 更多地是针对某一特定领域, 如位置隐私保护, 目前还尚未形成统一的模型及体系。其应用范围也受到限制, 特别是隐私是具有时空性的,

作者简介: 彭慧波(1994-), 男, 研究生, 主要研究方向: 数据定价; 周亚建(1971-), 男, 副教授, 主要研究方向: 数据定价, 文本分类。

与人的主观感受也有关系,不同的人对同一隐私的认同可能不同。然而在数据交易的过程中,隐私含量作为一个影响数据定价的重要特征,如何衡量数据集中隐私含量就显得尤为重要。目前尚没有一种针对交易数据集中隐私含量的度量方式^[3]。

在研究过程中本文的研究对象主要是待交易数据集中的各个元组。此外,本文做出以下假设:(1)数据拥有者和数据购买者对数据和其中包含的敏感信息都十分关注,出于法律等因素的约束,可信第三方不会主动泄露数据中的敏感信息;(2)数据拥有者和数据购买者都是理性的,是否出售和购买数据都是通过合理分析得出,包括成本、收益、风险在内的所有因素都可以用价格进行衡量;(3)本文从数据交易平台的角度出发,结合数据本身特性分析交易数据集的价格,得出的结果均不考虑除数据本身之外其它因素对价格所带来的影响;(4)本文所设定的模型不考虑套利和整体折扣的情况。

基于以上论述,本文提出一种基于信息熵的隐私度量方法,并在此基础上建立基于隐私度量的数据定价模型。

1 相关工作

包括微软、亚马逊、贵州大数据交易所、上海大数据交易中心在内的国内外数据交易平台结合多年来的实践经验,提出了包括基于次数定价、基于固定费用定价(只与使用时间有关)、双方税率定价(先收取基本费用,每次使用收取固定费用)、免费增值定价(使用开放数据免费,使用额外服务收取一定费用)在内的不同数据定价方法。

在理论研究方面,国内外科研人员也提出了各式的数据定价模型。这些定价方法可分为人工定价和自动定价两大类。在人工定价方面,张晓玉等人根据博弈论方法,运用“一对一”和“一对多”的讨价还价模型,对数据的价格确定过程进行了详细分析,对数据定价协商过程进行了建模^[4]。Christopher Riederer 等人提出了一种数据拍卖模型,用户根据自身隐私信息含量提出交易底价,数据购买者在数据交易平台上通过拍卖的方式购买经过脱敏处理的数据^[5]。而在自动定价领域中,Dusit Niyato 等人结合了经济学中 Stackelberg 模型和机器学习中的分类算法,将数据交易分为数据提供者、数据交易服务提供者和数据消费者三个部分,数据提供者设定数据价格为 p_b ,数据消费者设定数据价格为 p_s ,数据

交易服务提供者通过最小化 p_b 、最大化 p_s ,使 p_b 尽可能逼近 p_s 来实现各方利益的最大化^[6]。为了解决人工议价速率较慢的问题,Magdalena Balazinska 等人提出在元组(即数据库中的每条信息)这一结构粒度上设定数据的价格,并基于公共数据库中待交易数据集产生的顺序建立定价函数^[7]。定义价格函数 f :

$$f: D \rightarrow R^+ \quad (1)$$

其中, D 为数据集合, R^+ 为数据价格。在考虑非法套利的情况下,包含多个元组的数据集价格,就是各个元组单一价格之和,再减去一个价格差。

除此之外,Yuncheng Shen 等人提出了积极分级和反转定价机制^[8]。数据属性根据其影响数据质量的程度被分为不同等级,每个数据元组的准确价格都是根据数据属性因素和数据集的整体价格进行计算。将数据元组作为最基本的数据度量组合,结合信息熵、权重、数据引用指数、花费等影响数据价值的因素对每个元组进行定价^[9]。

现有的数据交易仍然采用协议定价、拍卖定价、集合定价等定价方式,数据定价方法无法实现人工智能时代自动化定价的目标;再者,由于信息不对称,存在第三方非法套利的情况^[10];此外,由于用户隐私保护的限制,再加上数据本身极易复制的特性,使得从价值上看,数据上具有极高的固定成本和极低的可变成本^[11],导致了传统定价方法在数据领域不再生效。

2 数据集中的隐私度量

在个人隐私保护和个人数据共享使用方面,世界各国都主张采取二者均衡的基本政策,即在一定个人隐私保护基础上,用户授权数据的正常开放、分享和使用,但不允许数据乱用、滥用^[12]。

虽然,数据交易平台已经对待交易数据集进行了脱敏处理,但是为了方便进行下一步数据挖掘,个人隐私数据仍然部分保留在交易数据集中。影响交易数据集价格的最主要因素是数据集中的隐私含量,因此隐私度量就显得尤为重要^[13]。而隐私度量可从应用领域、个人行为数据、隐私程度等三个方面进行分析。数据集中隐私含量的多少根据应用领域的不同,以及个人私事、个人数据、个人行为的不同而存在较大差异。

世界上大多数国家和地区对敏感信息定义可概括为两种基本模型:“欧盟标准”和“欧盟标准加犯

罪记录”。在法律上对 72 个国家和地区的相关敏感信息、数据进行归类、统计^[14]。针对个人的风险和价值的不同,对隐私信息的不同级别进行水平划分,而对不同隐私信息进行一定力度的保护。对隐私信息在应用中的保护价格进行设定,考虑从隐私级别、交易数据集描述对象两个方面进行分析,以此得出待交易数据集中隐私的含量。

将待交易数据集中的隐私级别用 L_k 进行表示,设 L_k ($0 \leq k \leq 9$) 的值在 $[1, 10]$ 之间,其中共有 10 个隐私风险级别, L_0 为最低级别,其中包含的隐私含量为 0,所有的信息均不涉及个人隐私信息,都是可被公开的数据或是在互联网上可以搜集到的数据; L_9 为最高级别,即数据集中的所有信息均为敏感信息。隐私级别的分类根据经验进行设置,假设数据集中第 i 个元组的隐私级别为 L_i ,整个数据集的隐私级别 L 为各个元组的隐私级别之和取平均值,其满足如下约束条件:

$$\sum_{i=1}^n L_i = n * L \quad (2)$$

待交易数据集在不同的信息领域下的作用不同,将使得隐私的需求量增加,使得隐私数据的价值增加,致使最后的交易数据集的价格增加。现有待交易数据集可根据描述对象的不同,分为人、物、组织等三种不同的维度,采用 O_d 进行表示,设 O_d 的值在 $[1, 3]$ 之间。其中,人这一维度中所包含的隐私信息较多,需要进行数据维度的拆分,否则多个维度之下将可以具体定位到某一个人。并且描述“人”这一对象的数据集多用于风投风控和市场营销,因此假设描述“人”这一对象的 O_d 值为 3。其次,相较于描述“组织”这一对象的数据集,描述“物”这一对象的数据集中的隐私含量较多,因此如果交易数据集的描述对象为“物”,其 O_d 为 2;若描述对象为“组织”时,则 O_d 为 1。

此外,所谓交易数据集售卖的是数据集中有价值的信息,即信息售卖者将信息购买者想要获取但不知道的信息贩卖给信息购买者,其中真正有价值的部分就是除了隐私就是未知信息。根据 Shannon 信息论,信息量的多少可以用信息熵来进行衡量,这里的信息内容是与事情发生的可能性相关的相对数量;换句话说,信息内容等于选择概率的对数概率,用其概率的负对数来表示^[15]。

假设 X 是一个离散随机变量,其取值空间为 R ,满足概率分布 $p(x) = P(X = x), x \in R$, 则变量 X 信息熵 $H(X)$ 被定义为:

$$H(x) = - \sum_{x \in R} p(x) \log_2 p(x) \quad (3)$$

信息熵也有可以称作自信息,其描述了随机变量的不确定性度量。自信息表示了信息源 X 发生各种情况所带来的平均信息。当自信息量越大的情况下,其发生的不确定性越大,正确估计其价值的可能性就越小。随机变量的不确定性越大的情况下,决定其价值的信息内容就越多。

假设一个数据集有 n 个数据元组,每个元组有 k 个属性。 x_{ij} 表示第 i 个元组的第 j 个数据项(即第 j 个属性值)。 $p(x_{ij})$ 表示 x_{ij} 出现在数据集中的概率。因此,数据集中第 i 个元组的第 j 个数据项的概率计算公式如式 4 所示:

$$p(x_{ij}) = \frac{x_{ij} \text{ 出现的次数}}{\text{数据集中的元组个数}} \quad (4)$$

数据集中第 i 个元组的熵计算方式如式 5 所示:

$$H(x_i) = - \sum_{j=1}^k p(x_{ij}) \log_2 p(x_{ij}) \quad (5)$$

待交易数据集中所有元组的总信息熵如式 6 所示:

$$H(x) = \sum_{i=1}^n H(x_i) = - \sum_{i=1}^n \sum_{j=1}^k p(x_{ij}) \log_2 p(x_{ij}) \quad (6)$$

根据 Shannon 信息论的定义,随机变量的信息熵越大,不确定性越大,信息内容的有效性越大,其交易价格也就越高。因此,熵的大小就代表了价格的高低。通过式 3-3 计算数据集中第 i 个元组的熵,定义 q_i 表示第 i 个元组的信息内容的大小,则 $H(x_i)$ 分配系数 q_i ; 通过式 3-4 计算整个数据的信息熵,定义 q 表示数据集信息内容的大小,则 $H(X)$ 分配系数 q 。则第 i 个数据元组信息内容的权重系数为 $\frac{q_i}{q}$,其满足如下限制:

$$\sum_{i=1}^n \frac{q_i}{q} = 1 \quad (7)$$

综上,可以得出数据集中某一元组的隐私含量 θ_i 的计算公式为:

$$\theta_i = L_k * O_d * H(x_i) = - \sum_{j=1}^k L_k * O_d * p(x_{ij}) \log_2 p(x_{ij}) \quad (8)$$

整个数据集的隐私含量 θ 的计算公式为:

$$\theta = L * O_d * \sum_{i=1}^n H(x_i) = - \sum_{i=1}^n \sum_{j=1}^k L * O_d * p(x_{ij}) \log_2 p(x_{ij}) \quad (9)$$

此时, 应该注意的是各个元组的隐私含量相加不等于这个待交易数据集的隐私含量, 原因在于多个维度的隐私组合之后能够获取到更多的信息, 数据购买者能够获取的隐私价值就更高。

3 基于隐私度量的数据定价模型

构建一个有效地数据定价模型能够促进传统金融市场的进化, 因此有必要构建一种合理、有效的数据定价模型, 更直接地评估个人数据的内在价值, 并解决各方之间个人数据的使用和销售问题。

目前, 影响数据定价的参数有很多, 通过调研现有数据定价模型, 从中可以提取影响数据定价的数据质量评价指标、数据拆分指标、数据历史成交价指标、数据效用指标等多种现行的数据交易定价指标。其中数据质量评价指标又包含数据样本量、数据品种、数据完整性、数据时间跨度、数据实用性、数据实时性、数据深度、数据样本覆盖度、数据稀缺性等多种指标^[16]。但是, 这些数据交易定价指标往往存在各种各样的内在弊端及缺乏对数据产品成本的考虑。

具体来说, 数据质量指标、数据拆分指标虽然已有很多细分指标, 但这些指标只能评判数据产品的价格高低, 无法给出数据产品的准确价格区间; 数据历史成交价指标虽然具有很重要的参考价值, 是数据市场供需情况的准确反映, 但由于数据交易市场本身的不成熟性、数据产品价格自身的复杂性, 这种指标也难以有效发挥其作用^[17]; 数据效用指标虽然切切实实指出了数据交易的关键之处, 在一定程度上表示了数据的价值, 也是人们最关心的一种指标, 但它也存在几个不足之处: 虽然数据买方在使用数据产品前的收益较好确定, 但买方受自身能力、外部机遇等因素影响, 使用数据产品后能不能取得收益、取得多少收益都是未知因素, 如此一来, 数据效用指标便成了纸上谈兵^[18]。因此, 选取合适的影响数据定价特征势在必行。

最终目标是确定可用于在开放市场中对数据进行定价的模型, 作者研究了其他数字资产的交易方式。本文将数据元组作为数据度量的基本单位, 使用它来评估个人数据的价值, 以便在数据市场中校准其价格。结合前人的相关工作和之前的研究调查, 除了前文提到的隐私度量信息和信息熵, 本文还使用了数据引用指数作为影响数据定价的相关特征。

数据引用指数的思路来自于 H 指数, 是 2005

年由加州大学圣地亚哥分校的物理学家 Jorge E. Hirsch 提出, 作为确定理论物理学家相对质量的工具, 有时也被称为 Hirsch 指数或 Hirsch 数。 H 指数反映了出版物的数量和每个出版物的引用次数。 H 指数的定义是指数为 h 的学者发表了 h 篇论文, 其中每篇论文至少在其他论文中被引用过 h 次。使用 H 指数来衡量出版的权威性, H 指数越大, 引用的论文越多, 引用次数就越多。根据出版物的 H 指数排名, H 指数越大, 排名越高。参考 H 指数的定义, 用户等同于发布者, 用户的数据元组等于发表的论文, Yuncheng Shen 等人提出了一种数据引用指数。所谓数据引用指数指的是至少 r 个数据元组分别购买 r 次, 最大值称为用户数据引用指数, 被简称作 R 指数。采用 R 指数测量用户的购买能力, R 指数越大, 购买的数据元组越多, 购买的次数就越多。 R 指数同时根据数据元组的购买金额和购买时间来衡量用户的购买能力。根据用户的 R 指数排名, R 指数越大, 排名越高。

假设一个交易数据集中包含 m 个用户, 第 j 个用户的 R 指数是 r_j , 所有用户的 R 指数之和是 r 。

因此, 第 j 个用户的 R 指数权重是 $\frac{r_j}{r}$, 其满足约束:

$$\sum_{j=1}^m \frac{r_j}{r} = 1 \quad (10)$$

为所有数据类型创建通用模型是一项艰巨的任务, 数据源可能需要基于数据类型及其潜在用途的不同定价模型。模型的开发需要进一步探索可能与数据值有关的客观自变量, 其中一些已在上文概述。此外, 必须收集具有其价格和属性的适当数量的样本数据集作为模型的输入, 理想情况下从大到小, 跨越多种用途^[19]。根据 Yuncheng Shen 等人提出的正向定级、反向定价原理, 待交易数据集的质量越高, 用户要求提供数据的供应价格越高, 买家愿意为数据支付更高的需求价格^[20]。数据交换平台可以为买方分析和转换数据, 从而可以减少数据的不确定性, 从而提高数据质量。正如前文所述, 有许多因素会影响数据元组的价值。设计通用模型几乎不可能涵盖所有影响因素。有两个最重要的因素: 隐私含量、数据引用指数。假设 α 为隐私含量的参数, β 为数据引用指数的参数, 其满足如下约束:

$$\alpha + \beta = 1 \quad (11)$$

基于上述假设, 将数据包中的第 i 个数据元组的价格设为 p_i , 原始数据包的供给价格为 P_s 。可以

推导出第 i 个数据元组的价格计算公式：

$$p_i = P_S * \left(\frac{\theta_i}{\theta} * \alpha + \frac{r_j}{r} * \beta \right)$$

(12)

在公式 4-3 中， $i=1,\cdots,n$ ， $j=1,\cdots,m$ ， p_i 定义了第 i 个元组的价格， P_S 定义了待交易数据集的供给价格， n 定义了交易数据集中的元组个数， m 表征了数据集的购买用户数， θ_i 表示了第 i 个元组的隐私含量， θ 表示整个数据集的隐私含量， r_j 表示第 j 个用户的数据引用， r 是所有用户的数据引用指数之和。因此， p_i 满足如下情况：

$$\sum_{i=1}^n p_i = P_S$$

(13)

公式 4-4 表明各个数据元组的供给价格之和等于整个待交易数据集的供给价格。

本文使用了一个具体的例子来验证定价模型的合理性和有效性。假设数据交换平台中有一个描述对象为人的包含 10 个数据元组的数据包，每个元组有 4 个项目。该数据包如表 1 所示。

表 1 详细数据项和数据包中的相应概率

Tab.1 Detailed data items and corresponding probabilities in the data package

用户 ID	支出（元）	类别	隐私级别
2019001(40%)	30(20%)	购物(20%)	L7(20%)
2019001(40%)	80(30%)	娱乐(30%)	L6(30%)
2019001(40%)	50(10%)	交通(20%)	L0(10%)
2019001(40%)	200(10%)	医疗(10%)	L9(10%)
2019002(30%)	100(10%)	娱乐(30%)	L6(30%)
2019002(30%)	40(10%)	购物(20%)	L7(20%)
2019002(30%)	80(30%)	交通(20%)	L1(10%)
2019003(20%)	60(10%)	晚餐(10%)	L3(10%)
2019003(20%)	30(20%)	通讯(10%)	L5(10%)
2019004(10%)	80(30%)	娱乐(30%)	L6(30%)

假设数据集的供给价格为 100 元，隐私指数 θ 的参数 $\alpha=0.5$ ，数据引用指数 R 的参数 β 为 0.5。根据上述公式进行计算，所有计算过程均保留小数点后三位，可得出结果如表 2 所示。

表 2 每个元组的详细价格表

Tab.2 Detailed price per tuple

序号	隐私级别值	信息熵	信息熵占比	隐私含量	R 指数	R 指数比率	影响因素加权和	元组价格（元）
1	8	1.922	0.109	46.128	3	0.100	0.105	10.450
2	7	2.092	0.118	43.932	3	0.100	0.109	10.900
3	1	1.658	0.094	4.974	4	0.133	0.114	11.350
4	10	1.525	0.086	45.75	4	0.133	0.110	10.950
5	7	1.895	0.107	39.795	2	0.067	0.087	8.700
6	8	1.782	0.101	42.768	4	0.133	0.117	11.700
7	2	1.839	0.104	11.034	2	0.067	0.086	8.550
8	4	1.461	0.083	17.532	3	0.100	0.092	9.150
9	6	1.593	0.091	28.674	3	0.100	0.096	9.550
10	7	1.895	0.107	39.795	2	0.067	0.087	8.700
总计	6	17.662	1.000	317.916	30	1.000	1.000	100.000

如表 2 所示，所有的计算结果最后均满足之前提出的公式，并且各个元组的价格之和等于整个交易数据集的价格，这证明了当前提出的模型是正确，合理，有效的。

目前大多数数据定价算法采用平均价格来决定每个数据元组的价格。也就是说，数据集的价格被均等地分配给每个数据元组，而不考虑数据元组所包含的价值。这种定价方法并不反映数据定价的公平性和合理性。本文提出了一种定价方法，可以准确地控制每个数据元组的价格并反映其应有的价值。此外，该定价模型可以动态调整，能根据数据的不同而有所不同，可以为数据元组制定公平合理

的市场价格，反映数据的真实价值并形成良性数据市场^[21]。

4 结论

本文通过实际交易数据集实例验证了本文提出的基于信息熵的隐私度量方法的有效性。同时，通过上述数据集验证了以上述方法为核心的基于元组的数据定价模型的有效性。实验结果表明，本文提出的隐私度量方法和元组度量模型能在一定条件下准确衡量交易数据集的价格。但是，本文提出的方案仍具有一定的局限性，需要继续展开进一步的研究。

参考文献

- [1] Gkatzelis V, Aperjis C, Huberman B A. Pricing private data[J]. *Electronic Markets*, 2012, 25(2): 1-15.
- [2] 彭长根, 丁红发, 朱义杰, 等. 隐私保护的信息熵模型及其度量方法[J]. *软件学报*, 2016, 27(8): 1891-1903.
- [3] 连玉明. 重新定义大数据[M]. 北京: 机械工业出版社, 2017: 164-175.
- [4] 张晓玉. 基于讨价还价博弈的大数据商品交易价格研究[D]. 辽宁科技大学, 2016.
- [5] Riederer C, Erramilli V, Chaintreau A, et al. For sale: your data: by: you[C]//ACM Workshop on Hot Topics in Networks. ACM, 2011: 13.
- [6] Niyato D, Alsheikh M A, Wang P, et al. Market model and optimal pricing scheme of big data and Internet of Things (IoT)[C]//IEEE International Conference on Communications. IEEE, 2016.
- [7] Balazinska M, Howe B, Koutris P, et al. A Discussion on Pricing Relational Data[M]//In Search of Elegance in the Theory and Practice of Computation. Springer Berlin Heidelberg, 2013: 167-173.
- [8] Shen Y, Guo B, Shen Y, et al. A Pricing Model for Big Personal Data[J]. *清华大学学报: 自然科学英文版*, 2016, 21(5): 482-490.
- [9] 陈筱贞. 大数据交易定价模式的选择[J]. *港澳经济*, 2016(18): 3-4.
- [10] 刘朝阳. 大数据定价问题分析[J]. *图书情报知识*, 2016(1): 57-64.
- [11] Moiso C, Minerva R. Towards a user-centric personal data ecosystem The role of the bank of individuals' data[C]// International Conference on Intelligence in Next Generation Networks. IEEE, 2012: 202-209.
- [12] Muschalle A, Stahl F, Löser A, et al. Pricing Approaches for Data Markets[C]//Workshop Business Intelligence for the Real Time Enterprise. 2012: 129-144.
- [13] 赵子瑞. 浅析国内大数据交易定价[J]. *信息安全与通信保密*, 2017(5): 61-67.
- [14] 干春晖, 钮继新. 网络信息产品市场的定价模式[J]. *中国工业经济*, 2003(5): 34-41.
- [15] Tsai Y C, Cheng Y D, Wu C W, et al. Time-Dependent Smart Data Pricing Based on Machine Learning[M]// Advances in Artificial Intelligence. 2017: 103-108.
- [16] 熊金波, 王敏桑, 田有亮, 等. 面向云数据的隐私度量研究进展, *软件学报*, 2017, DOI: 10.13328/j.cnki.jos.005363.
- [17] Balazinska M, Howe B, Dan S. Data Markets in the Cloud: An Opportunity for the Database Community[J]. *Proc of the Vldb Endowment*, 2011, 4: 1482-1485.
- [18] 王文平. 大数据交易定价策略研究[J]. *软件*, 2016, 37(10): 94-97.
- [19] Li C, Li D Y, Miklau G, et al. A Theory of Pricing Private Data[J]. *ACM Transactions on Database Systems (TODS)*, 2014, 39(4): 34.
- [20] Koutris P, Upadhyaya P, Balazinska M, et al. Query-based data pricing[C]//Symposium on Principles of Database Systems. ACM, 2012: 167-178.
- [21] Tang R, Wu H, Bao Z, et al. The Price Is Right[C]// International Conference on Database and Expert Systems Applications. Springer Berlin Heidelberg, 2013: 380-394.
- [22] 彭云聪, 任心晴, 石浩森. 基于核加权KNN和多目标优化的众包平台定价系统设计[J]. *软件*, 2018, 39(6): 150-154.