隐私保护的群体感知数据交易算法

张勇，李丹丹，韩璐，黄小红
（北京邮电大学计算机学院（国家示范性软件学院），北京 100876）

摘 要：为解决群体感知数据交易模式下参与者数据隐私泄露的问题，提出了一种隐私保护的群体感知数据交易算法。首先，为实现对参与者的隐私保护，设计了基于差分隐私的聚合方案，参与者不再需要上传原始数据，而是按照任务需求对收集的数据进行分析和计算，将任务结果按照平台分配的隐私预算添加噪声后发送给平台；其次，为确保参与者的可信性，构建了参与者的信誉模型；最后，为激励消费者和参与者参与交易，在考虑消费者对结果偏差的容忍约束和参与者的隐私泄露补偿的基础上构建了交易优化模型以优化平台的收益，并给出了基于遗传算法的收益优化算法（POA）来求解该模型。仿真结果表明，POA 不仅保护了参与者的隐私，而且在平台的收益方面相比于 VENUS 和 DPDT 分别提高了 29.27%和 20.45%。

关键词：群体感知；数据交易；差分隐私；信誉模型
中图分类号：TP399
文献标志码：A
DOI: 10.11959/j.issn.1000−436x.2022082

Privacy-protected crowd-sensed data trading algorithm

ZHANG Yong, LI Dandan, HAN Lu, HUANG Xiaohong
School of Computer Science (National Pilot Software Engineering School), Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract: To solve the problem that data privacy leakage of participants under the crowd-sensed data trading model, a privacy-protected crowd-sensed data trading algorithm was proposed. Firstly, to achieve the privacy protection of participants, an aggregation scheme based on differential privacy was designed. Participants were no longer needed to upload raw data, but analyzed and calculated the collected data according to the task requirements, and then sent the analysis results to the platform after adding noise in accordance with the privacy budget allocated by the platform to protect their privacy. Secondly, in order to ensure the credibility of participants, a reputation model of participants was proposed. Finally, in order to encourage consumers and participants to participate in transactions, a data trading optimization model was constructed by considering the consumer's constraint on the result deviation，the participant's privacy leakage compensation and platform profit, and a POA based on genetic algorithm was proposed to solve the model. The simulation results show that the POA not only protects the privacy of participants, but also increases the profit of the platform by 29.27% and 20.45% compared to VENUS and DPDT, respectively.

Keywords: crowd sensing, data trading, differential privacy, reputation model

0 引言

为了发掘数据中蕴含的潜在价值，打破“数据孤岛”，一些数据交易市场应运而生，如 Infochimps、Datacoup、Microsoft Azure Market-place 等。数据消费者可以在这些平台上搜索和购买他们所需要的数据[1-2]。然而，当前这些平台的数据大部分是机构或组织出于自身的业务需求而保存下来的数据，无

收稿日期：2021−12−21；**修回日期：**2022−03−02
通信作者：黄小红，huangxh@bupt.edu.cn
基金项目：国家重点研发计划基金资助项目（No.2020YFE0200500）；北京邮电大学优秀博士生创新基金资助项目（No.CX2019212）
Foundation Items: The National Key Research and Development Program of China (No.2020YFE0200500), The BUPT Excellent Ph.D. Students Foundation (No.CX2019212)

法满足数据消费者多样化的数据需求，从而使数据的可用性大打折扣[3-4]。因此，从多样性的角度来看，当前交易市场中的数据还十分有限，无法满足日益增长的数据交易需求。为解决这个问题，学者们提出了一种新的数据交易模式，即群体感知数据交易（CDT, crowd-sensed data trading）。CDT 采用移动群体感知技术为数据消费者提供数据资源，即数据提供者使用手机、平板电脑等智能设备按照数据消费者的需求为其收集数据[5-6]。

典型的 CDT 系统（例如 Thingful、Thingspeak）包括数据交易平台（以下简称为平台）、数据消费者（以下简称为消费者）和感知任务参与者（以下简称为参与者）。平台根据消费者的任务需求聘请大量的参与者，然后将参与者感知到的数据出售给消费者。当前，已有一些关于 CDT 系统设计的工作。An 等[7]提出了一个基于逆向拍卖的群体感知数据交易系统，采用贪心策略来招募参与者，并保证拍卖过程的真实性。Zheng 等[8]提出了一个收益驱动的数据收集框架，采用拍卖机制来最小化数据收集的成本。Jiang 等[9]提出了一个基于质量感知的数据共享市场模型，从博弈论的角度分析了消费者行为，并提出了最优响应迭代算法来提高社会福利。

当前，大部分群体感知系统将平台建模为半可信的，即平台能够诚实地执行预定程序，但同时会分析参与者的数据进而窥探其隐私[10-11]。少数文献关注了群体感知数据交易市场中的隐私问题。其中，Gao 等[12]和 Niu 等[13]使用同态加密和签名验证机制保护了数据交易过程中消费者的出价隐私和身份隐私。Zhao 等[14]提出了一种基于区块链的数据交易模型，采用环签名和相似性学习来保护参与者的身份隐私。当前的隐私保护方案主要针对消费者和参与者的身份及报价等信息进行保护，而如果将这些方案用于参与者的数据隐私保护，不仅效率低下，还会导致数据的可用性急剧下降。还有一些学者基于本地差分隐私提出了感知方案，允许参与者通过报告噪声数据来保护自己的数据隐私。Wang 等[15]基于本地化差分隐私提出了一种参与者位置保护方案。Xue 等[16]为保护参与者的隐私，基于本地化差分隐私提出一种个性化的隐私保护方案。然而，这些方案往往需要大量的参与者，在参与者数量较少的情况下，数据的实用性较差，而如果选取大量的参与者，则会增加平台的招募成本。基于此，本文提出了一种新颖的基于差分隐私的群体感知方案，参与者不再需要提交原始数据，而是提交对原始数据进行分析或计算得到的任务结果，同时通过对结果添加噪声来保护他们的数据隐私，噪声的分布是由平台分配的隐私预算来确定的。在这种情况下，平台可以通过对参与者隐私预算的动态调节来均衡结果的精度水平和相应的招募成本。

为激励用户参与群体感知数据交易，价格机制的设计也是需要考虑的问题。当前针对数据交易的价格机制研究主要考虑了数据量、数据质量等数据本身的因素[17-18]。例如，Yu 等[19]通过考虑数据质量和数据多版本发布的策略，提出了一个双层数学规划模型来优化数据交易。Jiao 等[20]考虑了大数据的“无限供给”性，提出了一种基于拍卖的大数据市场模型，通过拍卖机制得到最优的数据交易价格和交易量。黄小红等[21]基于数据质量、数据属性等多维因素构建了数据交易的价格机制。

本文提出的隐私保护的群体感知方案中，参与者上传的数据不再是原始数据，而是由数据分析得出的结果，并不具有上述方案中所考虑的数据属性，如数据量、数据质量等，因此不适用于上述价格机制。受到隐私数据发布的启发[22-23]，在满足消费者对结果偏差容忍的条件下，本文考虑了对参与者的隐私泄露补偿，以及平台的收益来优化数据交易。平台根据收到的消费者的偏差阈值、报价以及参与者的成本等信息，通过优化招募方案，在满足消费者偏差的容忍约束的基础上，最小化招募参与者的成本，最大化自身的收益。

因此，为了保护参与者的数据隐私，同时构建适用的价格机制，本文提出了隐私保护的群体感知数据交易算法。通过差分隐私聚合算法，平台聚合参与者的结果。对于结果的计算，可以采用参与者独立计算或者在平台的组织下所有参与者协同计算的方式进行。同时，本文提出了信誉模型，以确保参与者的可信性，在考虑消费者的偏差约束和参与者隐私泄露补偿的基础上，构建了交易优化模型，以最大化平台的收益。本文的主要贡献如下。

1) 为实现对参与者的隐私保护，以及保障数据的可用性，提出了基于差分隐私的数据聚合方案。平台为招募到的参与者动态地分配隐私预算。参与者根据收到的隐私预算，在计算得到的结果中添加噪声。这样既保护了参与者的隐私，又保证了聚合后结果的可用性。此外，为确保参与者的可信性，

提出了信誉模型。

2) 为激励消费者和参与者进行数据交易，通过考虑消费者对数据偏差的容忍约束，参与者的隐私泄露补偿构建了交易优化模型以优化平台的收益。通过优化参与者的招募来提高平台的收益，并提出了基于遗传算法的收益优化算法（POA, profit optimization algorithm）来求解该模型。

3) 使用北京的天气和空气质量数据进行了实验，以评估模型的性能。结果表明，该模型不仅实现了对参与者的隐私保护，同时提升了平台的收益。

1　系统模型

1.1　群体感知数据交易框架

图 1 展示了本文构建的群体感知数据交易框架，其中包括消费者、平台和参与者。消费者将任务发送给平台，平台按照任务的需求招募参与者，同时为每个参与者分配隐私预算。参与者收集数据并将分析或计算出的结果在添加噪声后发送给平台。平台聚合参与者的结果以完成消费者的任务。

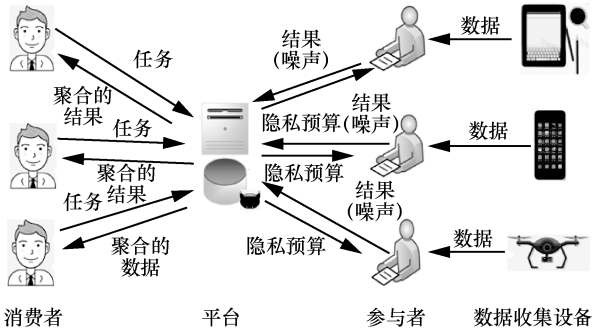


图 1　群体感知数据交易框架

群体感知数据交易框架中包含 3 种实体：消费者、平台和参与者。

1) 消费者是向平台发布任务的数据消费者。消费者会将自己的任务及相应的描述发送给平台，同时，以优化自身的满意度为目标确定报价和期望的偏差阈值。

2) 平台是群体感知数据交易的组织者。平台在收到消费者的任务后，会根据任务描述分析出完成该任务所需要的数据属性等，并招募相应的参与者以完成任务。在招募过程中，平台会以最大化自身的收益为目标，同时也需要为招募的参与者分配隐私预算以满足消费者的偏差要求。根据任务需求，平台也可以组织参与者进行联邦学习等协同计算以得到任务需求的结果。

3) 参与者是搜集数据的数据提供者。在收到平台所公布的任务要求后，参与者将自己所能满足的要求及相应的报价发送给平台，报价包含硬件成本（硬件损耗、能源消耗等）和隐私泄露成本。参与者收集数据，通过统计分析、数据挖掘或协同计算等方式得出消费者所需要的结果，并按照平台给出的隐私预算添加噪声后发送给平台。

1.2　数据交易流程

数据交易的过程如图 2 所示，其具体流程如下。

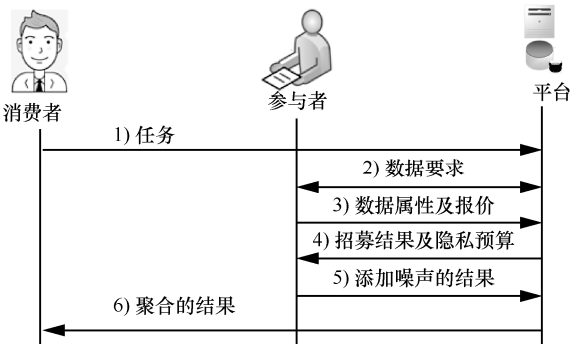


图 2　数据交易流程

1) 消费者将想要查询的任务发送到平台，其中包括对任务的详细描述、所能支付的报酬及期望的偏差阈值。

2) 平台对消费者的任务描述进行分析，提取出所需收集的数据属性以及需要计算的标签，然后将这些要求发送给参与者。

3) 参与者将自己所能收集的数据属性及相应的报价发送给平台。

4) 平台查询交易记录，以计算参与者的信誉值。平台以最大化自身的收益为目标确定招募的参与者，并为每个参与者分配隐私预算，随后将招募结果及隐私预算发送给相应的参与者。

5) 参与者根据任务的需求收集数据，并利用所收集的数据进行统计分析、数据挖掘或在平台的组织下进行协同计算以得出消费者想要的结果（即相应的标签），再根据平台分配的参数添加噪声后发送给平台。

6) 平台聚合所有参与者的结果，并将聚合的结果回复给消费者。消费者将报酬支付给平台，平台按照参与者的报价将酬金支付给参与者。

2　群体感知数据交易模型

本节将描述基于差分隐私的聚合方案，并构建数据交易优化模型。

2.1 问题描述

本节在描述问题之前，给出系统参数和变量的含义，如表 1 所示。

表 1 系统参数和变量的含义

参数和变量	含义
P	数据交易平台
U	消费者的集合
u_j	U 中的第 j 位消费者
γ_j	u_j 向 P 提交的任务
$\theta_j^{\max}$	u_j 可容忍的最大偏差
W	参与者的集合
w_i	W 中的第 i 位参与者
B_i	w_i 的报价
$r_{i,j}$	w_i 对任务 γ_j 计算的结果
$\varphi_{i,j}$	w_i 对 $r_{i,j}$ 所添加的噪声
r_j	P 对任务 γ_j 聚合得到的结果
$q_{i,j}$	w_i 针对任务 γ_j 的归一化权重
θ_j^*	u_j 的最优偏差
c_j^*	u_j 的报价
X	参与者的招募方案
$\varepsilon_{i,j}$	针对任务 γ_j，w_i 的隐私预算
θ_j	针对任务 γ_j，聚合后结果的偏差程度

在当前的 CDT 方案中，平台招募参与者收集数据，然后将数据出售给消费者，但是过程中可能泄露参与者的隐私。为保护参与者的隐私，本文基于差分隐私提出了聚合方案。参与者在计算出任务的结果后，将按照平台分配的隐私预算添加噪声，随后将结果发送给平台。同时，为确保参与者的可信性，避免招募恶意的参与者，本文提出了参与者的信誉模型。最后，为了激励用户参与交易，通过考虑消费者对结果偏差的容忍约束和对参与者的隐私泄露补偿，本文建立了以平台收益最大化为目标的交易优化模型并进行了求解。因此，本节将描述提出的差分隐私聚合方案、信誉模型和交易优化模型的构建及求解。

2.2 差分隐私聚合方案

将图 1 所示平台记作 P。一段时间内，P 收集消费者的任务请求。令 U 表示消费者集合，U 中包含 J 位消费者，$u_j \in U$，$1 \leqslant j \leqslant J$，$u_j$ 向 P 提交任务 γ_j，所有任务的集合为 H，$H \triangleq \{\gamma_j | 1 \leqslant j \leqslant J\}$。

设参与者的集合为 W，$w_i \in W$，$1 \leqslant i \leqslant I$。$w_i$ 对任务 γ_j 计算得到 $r_{i,j}$，$r_{i,j} \in [0,1]$，添加的噪声为 $\varphi_{i,j}$，则 P 对于任务 γ_j 聚合得到的结果为

$$r_j = f_j(r_{1,j}, r_{2,j}, ..., r_{|W_j|}) = \sum_{w_i \in W} q_{i,j}(r_{i,j} + \varphi_{i,j}) x_{i,j} = \sum_{w_i \in W} q_{i,j}(r_{i,j} + \varphi_{i,j}) \tag{1}$$

其中，$f_j(\cdot)$ 表示聚合函数，为参与者结果的加权求和；$q_{i,j}$ 表示 w_i 针对任务 γ_j 的归一化权重；$x_{i,j}$ 是一个二值函数，$x_{i,j} \in \{0,1\}$，$x_{i,j}=1$ 表示 w_i 被选中来参与任务 γ_j，$x_{i,j}=0$ 表示 w_i 未被选中；W_j 表示被选中参与任务 γ_j 的参与者的集合，设 W_j 中参与者的数量为 K，$0 \leqslant K \leqslant I$。

聚合后结果偏差可以定义为

$$\theta_j(X_j) = \max \mathrm{E}\left(\sum_{w_i \in W} q_{i,j} r_{i,j} x_{i,j} - \sum_{w_i \in W} q_{i,j}(r_{i,j} + \varphi_{i,j}) x_{i,j} \right)^2 \tag{2}$$

其中，$X_j \triangleq \{x_{i,j}, 1 \leqslant i \leqslant I\}$。

CDT 中需要考虑的是单个参与者的隐私泄露问题，主要应对差分攻击。差分隐私是一种应对差分攻击的有效方法。首先定义差分隐私。

定义 1 $\varepsilon_{i,j}$−差分隐私[24]。任务 γ_j 的聚合函数 $f_j : [0,1]^K \to R$ 满足 $\varepsilon_{i,j}$−差分隐私，则 f_j 需要满足：若存在 2 个只在第 i 位参与者的结果处存在区别的相邻向量 $\boldsymbol{d}_j \triangleq (r_{1,j}, \cdots, r_{i,j}, \cdots, r_{I,j})$ 和 $\boldsymbol{d}'_j \triangleq (r_{1,j}, \cdots, r'_{i,j}, \cdots, r_{I,j})$，$f_j$ 对于任何一组聚合结果 $O \subseteq \mathrm{Range}(f_j)$ 都满足式(3)，则 f_j 满足 $\varepsilon_{i,j}$−差分隐私。

$$\Pr[f_j(\boldsymbol{d}_j) \in O] \leqslant \exp(\varepsilon_{i,j}) \Pr[f_j(\boldsymbol{d}'_j) \in O] \tag{3}$$

其中，$\varepsilon_{i,j}$ 是正的参数，用来表征隐私保护的强度，$\varepsilon_{i,j}$ 越小，隐私保护强度越大，隐私泄露就越少。因此，隐私预算 $\varepsilon_{i,j}$ 可以用来表征泄露的隐私量。

对于聚合函数 f_j，提供差分隐私的一个著名方法是将拉普拉斯分布得出的随机噪声添加到聚合函数[25]。由于本文方案允许每个参与者自己增加噪声，因此需要仔细设计噪声，使这些噪声的和等于从拉普拉斯分布中得到的随机噪声，即聚合噪声 $\varphi_j = \sum_{w_i \in W_j} q_{i,j} \varphi_{i,j}$ 服从拉普拉斯分布。

根据拉普拉斯分布的可分性，可以将拉普拉斯分布构造为多个独立同分布的伽马分布的和[26]。

因此，对于任意的$w_i \in W_j$，设$\varphi_{i,j} = G_1\left(K, \frac{\sigma_j}{q_{i,j}}\right) - G_2\left(K, \frac{\sigma_j}{q_{i,j}}\right)$，其中$G_1$和$G_2$是 2 个独立同分布的服从伽马分布的随机变量，并且具有概率密度函数$g\left(y: K, \frac{\sigma_j}{q_{i,j}}\right) = \frac{1}{\Gamma\left(\frac{1}{K}\right)}\left(\frac{q_{i,j}}{\sigma_j}\right)^{\frac{1}{K}} y^{\frac{1}{K}-1} \mathrm{e}^{\frac{q_{i,j}y}{\sigma_j}}$。这样可以保证聚合之后的噪声$\varphi_j$满足拉普拉斯分布$L(\sigma_j)$，其中$\sigma_j$是拉普拉斯分布的参数。

结论　已知所有参与者的$x_{i,j}$和$q_{i,j}$，在聚合函数f_j下，针对任务γ_j每个参与者的隐私预算和聚合结果的偏差可表示为

$$\varepsilon_{i,j} = \frac{q_{i,j}x_{i,j}}{\sigma_j}, \forall w_i \in W \tag{4}$$

$$\theta_j = \left(\sum_{w_i \in W} q_{i,j}(1 - x_{i,j})\right)^2 + 2\sigma_j^2 \tag{5}$$

证明　已知所有参与者的$x_{i,j}$和$q_{i,j}$，在聚合函数f_j下，敏感度为

$$s(f_j) = \max_{\boldsymbol{d}_j, \boldsymbol{d}'_j \in [0,1]^K} | q_{i,j}(r_{i,j} - r'_{i,j})x_{i,j} | = q_{i,j}x_{i,j} \tag{6}$$

因此，有$\varepsilon_{i.j} = \frac{s(f_j)}{\sigma_j} = \frac{q_{i,j}x_{i,j}}{\sigma_j}$。

对于偏差，有

$$\theta_j(X_j) = \max_{r \in [0,1]^I} \mathrm{E}\left[\left(\sum_{w_i \in W} q_{i,j}r_{i,j}x_{i,j} - \sum_{w_i \in W} q_{i,j}(r_{i,j} + \varphi_{i,j})x_{i,j}\right)^2\right] =$$

$$\max_{r \in [0,1]^I} \mathrm{E}\left[\left(\sum_{w_i \in W} q_{i,j}r_{i,j}(1 - x_{i,j}) - \sum_{w_i \in W_j} q_{i,j}\varphi_{i,j}\right)^2\right] \overset{(a)}{=}$$

$$\max_{r \in [0,1]^I}\left(\sum_{w_i \in W} q_{i,j}r_{i,j}(1 - x_{i,j})\right)^2 + 2\sigma^2 =$$

$$\left(\sum_{w_i \in W} q_{i,j}r_{i,j}(1 - x_{i,j})\right)^2 + 2\sigma^2 \tag{7}$$

等式变换 (a) 是由拉普拉斯分布的可分性得到的，φ_j是一个均值为 0、方差为$2\sigma_j^2$的拉普拉斯随机变量。

根据文献[27]，如果添加的拉普拉斯噪声包含参数$\sigma_j = \sigma(X_j) = \sum_{w_i \in W} q_{i,j}(1 - x_{i,j})$，则称式(1)中的聚合函数$f_j(\cdot)$是规范的。此时，隐私预算和结果偏差可以表示为

$$\varepsilon_{i,j}(X_j) = \frac{q_{i,j}x_{i,j}}{\sum_{w_i \in W} q_{i,j}(1 - x_{i,j})}, \forall w_i \in W \tag{8}$$

$$\theta_j(X_j) = 3\left(\sum_{w_i \in W} q_{i,j}(1 - x_{i,j})\right)^2 \tag{9}$$

如此，构建了参与者的隐私预算和结果偏差与参与者选取的关系。从式(8)和式(9)可以看出，参与者越多，结果偏差越小，但是隐私预算越大。从直观上说，对于相同的任务，参与者越多，结果越精确，但每个参与者的隐私损失越大。此外，还需要仔细选择参与者，因为他们有不同的信誉值（即$q_{i,j}$），这也会导致不同的偏差。同时，不同参与者的成本也不同。因此，需要寻找一组合适的参与者来完成感知任务。

在得到X_j后，P会为每个参与者分配隐私预算，并聚合他们的结果，具体算法如算法 1 所示。

算法 1　差分隐私聚合算法

输入　X_j

输出　r_j

1) 根据X_j，P计算每个参与者的隐私预算$\varepsilon_{i,j}$（式(8)），并将$x_{i,j}$、K和$\frac{\sigma_j}{q_{i,j}}$发送给参与者
2) for $i = 1 : I$
3) w_i按照分布$G_1\left(K, \frac{\sigma_j}{q_{i,j}}\right) - G_2\left(K, \frac{\sigma_j}{q_{i,j}}\right)$向$r_{i,j}$中添加随机噪声$\varphi_{i,j}$，并发送给$P$
4) end for
5) P聚合参与者的结果（式(1)），得到回复给消费者的任务结果r_j

需要注意的是，由于添加的噪声服从拉普拉斯分布，因此算法 1 适用于计算结果是连续型的情况。证毕。

2.3 信誉模型

当前已经有一些研究通过度量参与者的工作能力来确定其信誉值，但平台需要知道参与者的一些信息如计算能力、存储能力、所处位置、轨迹信息等，这些信息中往往蕴含着用户的隐私[28-29]。因此，本文方案通过参与者在任务中的表现来设计信誉模型，基于参与者所提供的结果的准确程度来度量参与者的可信性，不再需要参与者提供上述隐私信息。参与者所提供的结果越精确，参与者的可信性越好，相应的信誉越高。不失一般性地，假设大部分参与者都是诚实的，会如实上报自己的结果。同时，用户在注册成为参与者时，平台一般会进行一些测试以确保用户有满足需求的工作能力[30]。

信誉模型是通过平台对参与者在任务中的评价而建立的。针对当前任务 γ_j，P 对 w_i 的信誉评价为

$$e_{i,j}=\begin{cases}1, & \left|r_j-(r_{i,j}+\varphi_{i,j})\right|\leqslant\tau\\ \dfrac{1}{\mathrm{e}^{\left|r_j-(r_{i,j}+\varphi_{i,j})\right|-\theta_j}}, & \left|r_j-(r_{i,j}+\varphi_{i,j})\right|>\tau\end{cases}\tag{10}$$

其中，τ 是固定的阈值，表示对误差的容忍程度。

P 会更倾向于选择近期进行过交易且信誉较高的 w_i，因此 w_i 会存在累积信誉，并且交易时间越接近，对当前任务信誉值的影响就越大。设 w_i 在一段时间内参与过 Ψ 次 P 发布的任务，$1\leqslant\psi\leqslant\Psi$，影响参数 $\vartheta_\psi=z^{\Psi-\psi}$，$z\in(0,1)$，$\psi$ 越大表示距离当前任务越近，影响参数也越大。考虑时间因素，信誉模型可以更新为

$$e_{i,j}=\frac{\sum_{\psi=1}^{\Psi}\vartheta_\psi e_i^\psi}{\sum_{\psi=1}^{\Psi}\vartheta_\psi}\tag{11}$$

其中，e_i^ψ 表示第 ψ 次任务中，w_i 所得到的信誉评价。

同时，任务的相似程度也是需要考虑的问题，与 w_i 当前所参与的任务越相似，给出的评价越具有价值。采用任务所需要收集数据属性的杰卡德相似系数来度量 2 个任务的相似程度[21]。因此，进一步考虑任务的相似程度，信誉模型可以更新为

$$e_{i,j}=\frac{\sum_{\psi=1}^{\Psi}\vartheta_\psi\chi_\psi e_i^\psi}{\sum_{\psi=1}^{\Psi}\vartheta_\psi\chi_\psi}\tag{12}$$

$$\chi_\psi=\frac{\left|M_\psi\cap M_j\right|}{\left|M_\psi\cup M_j\right|}\tag{13}$$

其中，M_ψ 是第 ψ 次任务所需要收集的数据属性，M_j 是 P 经过分析后认为完成任务 γ_j 需要收集的数据属性。式(1)中采用的 $q_{i,j}$ 是通过计算 $e_{i,j}$ 并进行归一化得到的，即 $q_{i,j}=\dfrac{e_{i,j}}{\sum_{w_i\in W}e_{i,j}}$。

2.4 交易优化模型的构建及求解

交易优化模型的目标是在满足消费者对结果偏差的约束及对参与者的隐私泄露补偿的基础上优化平台的收益。平台招募参与者以完成消费者的任务，因此平台的收益与消费者支付的报酬以及需要支付给参与者的酬金相关。

2.4.1 消费者支付的报酬

对于收到的结果，偏差越小，消费者从结果中得到的收益越大，但同时需要付出的成本也会越多，因此消费者会在收益和成本之间进行权衡。基于结果的偏差，构建消费者的满意度函数，它包含 2 个部分：消费者从结果中所获取的收益以及需要支付的报酬。消费者通过优化自身的满意程度来确定最优的偏差，以及所支付的报酬。

设 u_j 收到的结果的偏差为 θ_j，u_j 可容忍的最大偏差为 $\theta_j^{\max}$，则 u_j 的满意度函数可以表示为

$$\mathrm{ST}(u_j)=S(u_j)-C(u_j)\tag{14}$$

其中，$S(u_j)$ 表示 u_j 的收益，它是 θ_j 的单调递减函数，即偏差越大所能获得的满意度越低。因此，构建 $S(u_j)$ 为

$$S(u_j)=\alpha_j\ln(\theta_j^{\max}-\theta_j+1)\tag{15}$$

其中，α_j 表示固定的参数，由 u_j 的特性决定，用来调整消费者对满意度函数和成本的侧重程度。

$C(u_j)$ 表示消费者 u_j 的成本，也就是消费者支付给平台的报酬，结果是偏差越小，需要支付的报酬就越大，理论上，θ_j 越小，$C(u_j)$ 的增长程度越大，因为精度越高，继续提高精度需要的成本越大[31]。因此，定义

$$C(u_j)=\mathrm{e}^{\beta_j\left(\theta_j^{\max}-\theta_j\right)}\tag{16}$$

其中，β_j 表示固定的参数，由消费者对成本的容忍

程度来决定。因此，$\mathrm{ST}(u_j)$ 的完整形式为

$$\mathrm{ST}(u_j)=a_j\ln(\theta_j^{\max}-\theta_j+1)-\mathrm{e}^{\beta_j\left(\theta_j^{\max}-\theta_j\right)}$$
$$0\leqslant\theta_j\leqslant\theta_j^{\max} \tag{17}$$

$\mathrm{ST}(u_j)$ 对 θ_j 的一阶导数为

$$\frac{\partial\left(\mathrm{ST}(u_j)\right)}{\partial\theta_j}=-\frac{\alpha_j}{\theta_j^{\max}-\theta_j+1}+\beta_j\mathrm{e}^{\beta_j\left(\theta_j^{\max}-\theta_j\right)} \tag{18}$$

$\mathrm{ST}(u_j)$ 对 θ_j 的二阶导数为

$$\frac{\partial^2\left(\mathrm{ST}(u_j)\right)}{\partial{\theta_j}^2}=-\frac{\alpha_j}{\left(1+\theta_j^{\max}-\theta_j\right)^2}-\beta_j^2\mathrm{e}^{\beta_j\left(\theta_j^{\max}-\theta_j\right)} \tag{19}$$

很明显，$\frac{\partial^2\left(\mathrm{ST}(u_j)\right)}{\partial{\theta_j}^2}<0$，$\mathrm{ST}(u_j)$ 是凸函数且存在唯一的最优解。

消费者的目标是最大化其满意度 $\mathrm{ST}(u_j)$，也就是最小化 $\left(-\mathrm{ST}(u_j)\right)$。

$$\min_{\theta_j}\left(-\mathrm{ST}(u_j)\right)=$$
$$\min_{\theta_j}\left(-a_j\ln(\theta_j^{\max}-\theta_j+1)+\mathrm{e}^{\beta_j\left(\theta_j^{\max}-\theta_j\right)}\right),$$
$$0\leqslant\theta_j\leqslant\theta_j^{\max} \tag{20}$$

由于式(20)所示函数是严格凹的，且具有凸的约束，因此在 KKT 条件下存在唯一的最优解。放宽约束，得到拉格朗日函数为

$$L(\theta_j,\lambda_j,\eta_j)=-a_j\ln(\theta_j^{\max}-\theta_j+1)+$$
$$\mathrm{e}^{\beta_j\left(\theta_j^{\max}-\theta_j\right)}-\lambda_j\theta_j+\eta_j\theta_j-\theta_j^{\max} \tag{21}$$

其中，λ_j 和 η_j 是拉格朗日参数。

最优解满足以下条件

$$\frac{\partial L}{\partial\theta_j}=\frac{\alpha_j}{\theta_j^{\max}-\theta_j+1}-\beta_j\mathrm{e}^{\beta_j\left(\theta_j^{\max}-\theta_j\right)}-\lambda_j+\eta_j=0 \tag{22}$$

通过以上方式，u_j 可以计算出最优的偏差 θ_j^* 以及需要支付的相应报酬 c_j^*，c_j^* 为偏差取 θ_j^* 时成本函数 $C(u_j)$ 的值，即 $c_j^*=e^{\beta_j\left(\theta_j^{\max}-\theta_j^*\right)}$。

2.4.2　支付给参与者的酬金

w_i 在收到 P 发布的任务后，将针对 γ_j 能够收集的数据属性 $M_{i,j}$ 及报价发送给平台，设 w_i 的报价为 B_i，$B_i\triangleq\{b_{i,j},1\leqslant j\leqslant J\}$。$b_{i,j}$ 由 2 个部分构成。一部分是 w_i 完成任务 γ_j 所需的固定成本，用 $h_{i,j}$ 表示，包含能源消耗、计算成本、数据传输成本等[32]。另一部分是 w_i 的隐私成本，用 $p_{i,j}$ 表示，它与 w_i 在任务中泄露的隐私成正比[22-23]。此时，参与者只需要上报单位隐私的价格，隐私的量可由平台给出的隐私预算得出。当 w_j 放弃任务 γ_j 时将回复 $b_{i,j}=0$。

假设 w_i 被分配任务 γ_j，对于任务 γ_j，w_i 的隐私成本为

$$p_{i,j}=\omega_{i,j}\varepsilon_{i,j} \tag{23}$$

其中，$\omega_{i,j}$ 表示单位隐私预算的成本；$\varepsilon_{i,j}$ 表示泄露的隐私量，添加的噪声越多，隐私泄露越少，相应的成本越低。由于执行的任务都有时间限制，在划分的时间间隔内消费者只会被分配一个任务。平台支付给参与者的酬金为

$$B_{i,j}=h_{i,j}+p_{i,j} \tag{24}$$

2.4.3　交易优化模型

交易模型的目标是最大化平台的收益，也就是从消费者处收到的报酬及支付给参与者的酬金之间的差值，因此可以将它的目标函数表示为

$$\max_X\mathrm{Pro}=$$
$$\max_X\left(\sum_{\gamma_j\in H}R(\gamma_j)-\sum_{w_i\in W}B_{i,j}x_{i,j}^{'}\right) \tag{25}$$

s.t.

$$R(\gamma_j)=\begin{cases}c_j^*, & \theta_j(X_j)\leqslant\theta_j^*\\ 0, & \theta_j(X_j)>\theta_j^*\end{cases} \tag{26}$$

$$X_j^{'}=\begin{cases}X_j, & \theta_j(X_j)\leqslant\theta_j^*\\ \{0\}, & \theta_j(X_j)>\theta_j^*\end{cases} \tag{27}$$

$$x_{i,j}\in\{0,1\},\forall w_i\in W \tag{28}$$

$$0\leqslant\sum_{j=1}^{J}x_{i,j}\leqslant 1,\forall w_i\in W \tag{29}$$

其中，$X\triangleq\{X_j\,|\,1\leqslant j\leqslant J\}$，$X_j^{'}=\{0\}$ 表示 $X_j^{'}$ 是一个全 0 的集合，同时 $X_j^{'}\triangleq\{x_{i,j}^{'}|1\leqslant i\leqslant I\}$。

约束式(26)保证消费者得到的任务结果满足预期的偏差。约束式(27)表明当结果不满足消费者的预期偏差时，不会有参与者再进行此任务。约束式(28)说明参与者只有参与任务或不参与任务 2 种状态。约束式(29)保证每个参与者最多只能参与一个任务。

2.4.4　收益优化算法

交易优化模型具有大量的变量且约束函数包含二值函数，很难采用数值方法进行求解，而采用遍历的方法计算复杂度较高，为J^I，其中I和J分别表示参与者和消费者的数量。因此，本文基于遗传算法提出了POA，如算法2所示。

算法2　POA

输入　$U,\gamma_j,\theta_j^{\max},\alpha_j,\beta_j,W,\omega_{i,j},h_{i,j},p_{i,j},\nu$

输出　r_j

1) for　$j=1:J$
2)　　u_j最大化自身的满意度，计算出最优的偏差θ_j^*以及相应的成本c_j^*
3)　　u_j将任务γ_j、θ_j^*和c_j^*发送给P
4) end for
5) P对任务进行分析，找出想要解决γ_j所需要收集的数据属性M_j以及计算的标签l_j。分析完成之后，会将任务需求发送给参与者
6) for　$i=1:I$
7)　　w_i将针对每个任务能够收集的数据属性及报价发送给P
8) end for
9) 设定　Gen=1
10) P生成初始种群$\mathrm{CS}_{\mathrm{Gen}}$，种群中的每个个体为参与者的招募方案$X$，计算种群中每个个体的适应度函数（式(25)）
11) 采用轮盘选对种群进行遗传、交叉和变异操作，生成种群　$\mathrm{CS}_{\mathrm{Gen+1}}$
12) 计算$\mathrm{CS}_{\mathrm{Gen+1}}$中个体的适应度函数，判定$\mathrm{CS}_{\mathrm{Gen+1}}$和$\mathrm{CS}_{\mathrm{Gen}}$中具有最高适应度函数的个体的函数值之差是否小于阈值ν
13) 如果是，输出$\mathrm{CS}_{\mathrm{Gen+1}}$中具有最高适应度函数的个体
14) 如果否，Gen=Gen+1，返回步骤11)

POA的计算复杂度为$O(|\mathrm{CS}|IJ)$，其中$|\mathrm{CS}|$表示种群中个体的数量。POA属于标准遗传算法，在计算过程中会保留每次迭代的最优个体，因此理论上可以收敛到全局最优解[33]。

3　实验结果

本文实验使用北京的天气和空气质量数据来提供空气质量预测服务。该数据集包含12个空气质量监测站点的空气污染物数据。空气质量数据来自北京市环境监测中心。每个空气质量监测站点的气象数据均与最近的气象站相匹配。时间段是从2013年3月1日到2017年2月28日。因此每个数据集都包含35 064条数据样本。每条样本包含时间、PM2.5、PM10、SO_2含量等17种属性。实验中，将每个空气质量监测站的数据分成10份，并将每一份看成一个参与者收集的数据，因此实验中共引入了120位参与者。为了使这120位参与者的数据产生显著差异，实验中删除了某些参与者的一些数据属性。在实验过程中，参与者使用经典的机器学习算法（如线性回归算法等）提供空气质量查询、预测等服务。实验虚拟了10位消费者，即$J=10$。实验平台为一台PC主机，CPU为i7-8700，主频为3.20 GHz，内存为32 GB，频率为2 666 MHz。

实验主要展示了消费者满意度函数随偏差的变化，POA的收敛性、可扩展性及有效性验证，偏差和隐私预算随参与者选择的变化，消费者与参与者的变化对平台收益的影响及信誉模型的有效性验证。

3.1　消费者满意度函数随偏差的变化

消费者的满意度函数考虑了获取的收益和需要支付的报酬，为了展示最终的满意度、获取的收益及需要支付的报酬随偏差的变化，给出了消费者的满意度、获取的收益及需要支付的报酬随偏差的变化，如图3所示。

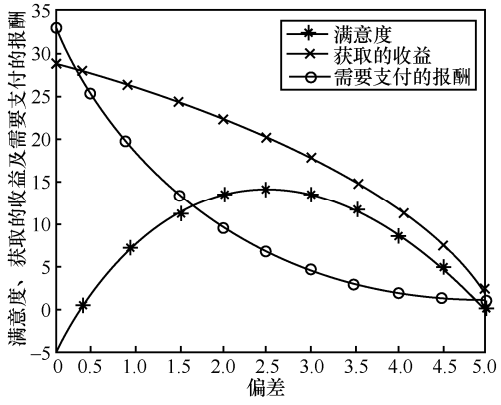


图3　消费者的满意度、获取的收益及需要支付的报酬随偏差的变化

消费者是从10位消费者中随机选取一位。由图3可知，随着偏差的增大，获取的收益和需要支付的报酬都逐渐降低，但获取的收益降低的速度逐渐加快，而需要支付的报酬降低的速度逐渐放缓。因此，消费者的满意度呈现先增长后降低的趋势。当偏差取2.4时，消费者的满意度达到最高值14.32，此时消费者需要付出的报酬为6.17。

3.2　POA 的收敛性、可扩展性及有效性验证

为展示算法的收敛情况，给出了平台的收益随迭代次数的变化，如图 4 所示。

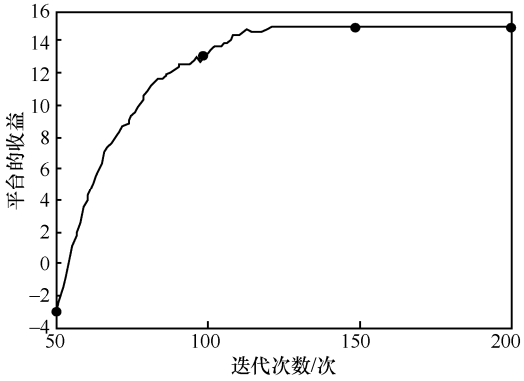


图 4　平台的收益随迭代次数的变化

从图 4 可以看出，随着迭代次数的增多，平台的收益逐渐增长。当达到一定的迭代次数后，收益的增长速度逐渐放缓，并最终趋于稳定。算法在 120 次之内可以达到稳定状态。

为了验证算法的可扩展性，给出了迭代次数随消费者和参与者数量增长的变化，如图 5 所示。

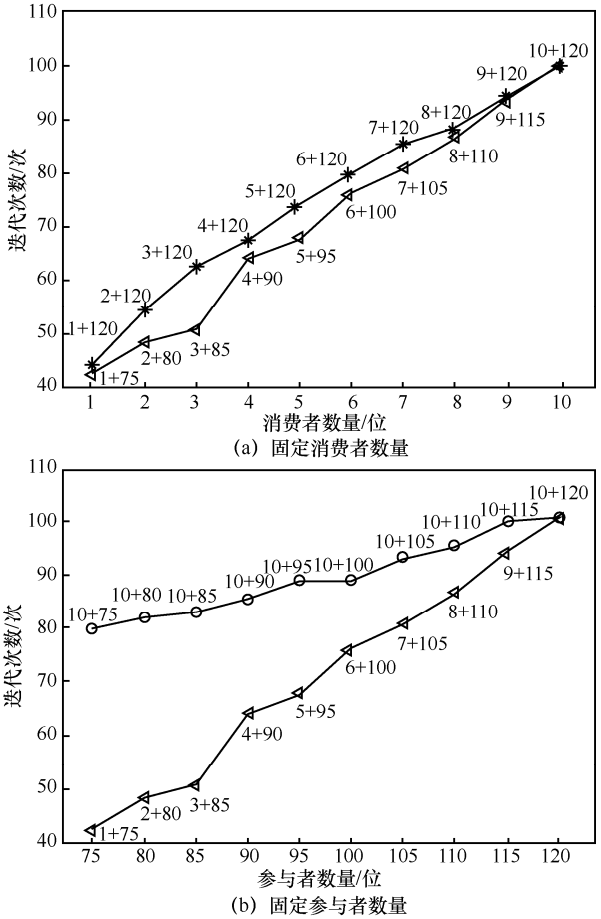


图 5　迭代次数随消费者和参与者数量增长的变化

图 5 中标注的数字表示“消费者数量+参与者数量”，如“5+120”表示此次实验中包含 5 位消费者和 120 位参与者。从图 5 可以看出，随着消费者和参与者数量的增加，算法的迭代次数呈线性增加。这一结果验证了 POA 对大规模数据交易系统具有较高的可扩展性。

为进一步展示 POA 的有效性，将其与 VENUS、DPDT 以及穷举法的实验结果进行了对比，实验结果如图 6 和图 7 所示。其中，VENUS 是 Zheng 等[8]提出的一种基于贪婪策略的群体感知数据交易方案，以联合优化利润最大化和支付最小化问题；DPDT 是 Gao 等[34]提出的一种满足差分隐私的群体感知数据交易机制，采用差分隐私拍卖的方法来实现数据定价和数据收集。

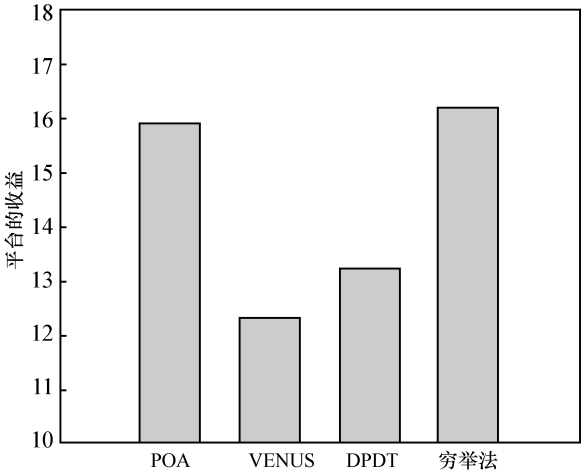


图 6　几种算法在平台的收益方面的对比

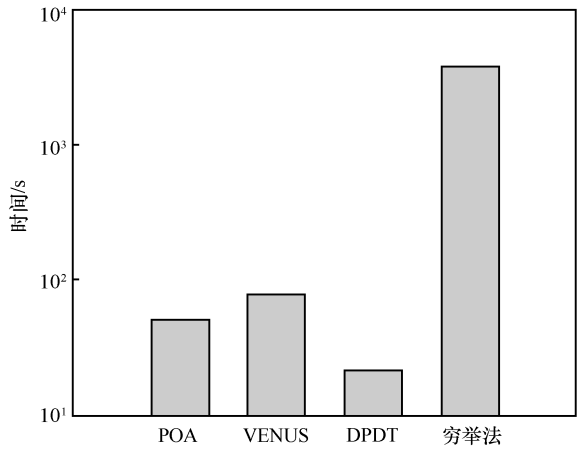


图 7　几种算法在运行时间上的对比

从图 6 和图 7 可以看出，穷举法具有最高的收益，但它的运行时间远超其他 3 种算法。相比于 VENUS 和 DPDT，POA 在平台的收益方面分别提

高了 29.27%和 20.45%，达到了穷举法的 98.15%，运行时间也仅略高于 DPDT。从平台的收益和运行时间综合来看，本文方案取得了最好的效果。

3.3　偏差和隐私预算随参与者选择的变化

按照本文方案中构建的参与者选择与隐私预算和偏差的关系，对于相同的感知任务，被选中的参与者越多，偏差越低，但同时会导致每个参与者有更高的隐私预算，从而泄露更多的隐私，为此，给出了偏差和隐私预算随被选中的参与者数量的变化，如图 8 所示。

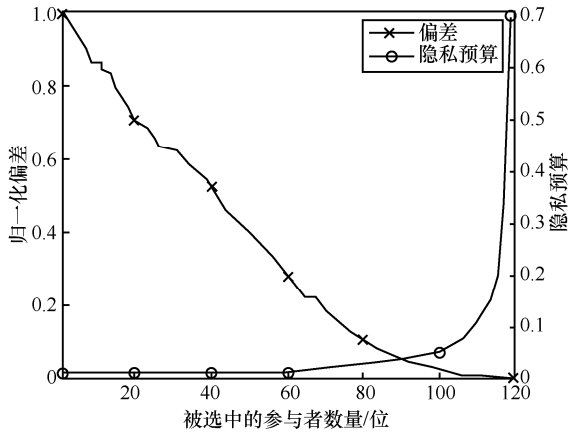


图 8　偏差和隐私预算随被选中的参与者数量的变化

从图 8 可以看出，随着被选中的参与者的数量增加，偏差逐渐减少，但每个参与者的隐私预算逐渐增多，且随着参与者数量的增长，隐私预算的增长速度越来越快，这与前文的分析是一致的。参与者的信誉值也会对参与者的选取造成影响，为展示这种影响，给出了不同信誉值下偏差和隐私预算随被选中的参与者数量的变化，如图 9 所示。

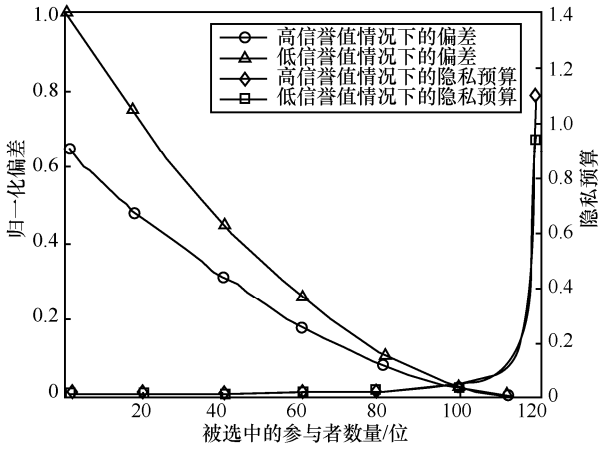


图 9　不同信誉值下偏差和隐私预算随被选中的参与者数量的变化

图 9 中，高信誉值情况下，参与者的平均信誉值为 0.9；低信誉值情况下，参与者的平均信誉值为 0.7。在选取相同参与者数量的情况下，高信誉值的参与者具有更低的偏差，同时他们的隐私预算情况基本重合，都出现了随着被选中的参与者数量增长隐私预算增长越来越快的现象。

为了展示隐私预算与数据效用之间的关系，给出了几种算法数据效用随隐私预算变化的对比，如图 10 所示。

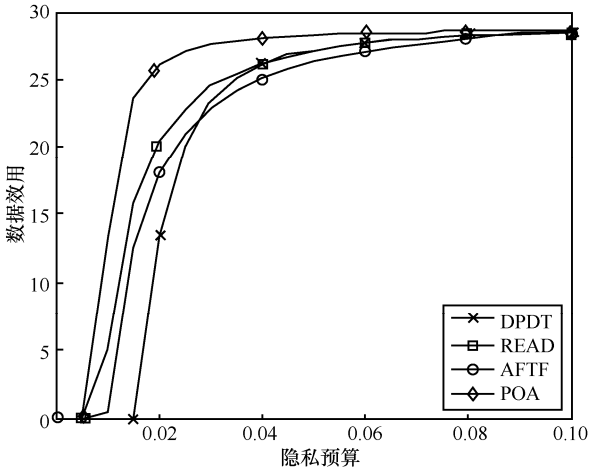


图 10　几种算法数据效用随隐私预算变化的对比

图 10 中，READ 和 AFTF（partial arbitrage free trading framework）分别是 Cai 等[35]和 Zheng 等[36]提出的基于差分隐私的数据交易算法。隐私预算可以用来表征隐私保护力度，两者成反比，隐私预算越小表示隐私保护力度越强。数据效用是利用数据为消费者带来的收益来计算的。从图 10 中可以看出，随着隐私预算的逐渐增长，即隐私保护力度的逐渐减小，数据效用逐渐增加，并最终稳定到最高值。从几种算法的对比来看，在相同的隐私保护程度下，POA 可以带来更高的数据效用。这是因为 POA 采用的差分隐私聚合算法考虑了拉普拉斯分布的可分性，将拉普拉斯分布构造为多个独立同分布的伽马分布的和，使每个参与者所添加的噪声之和等于从拉普拉斯分布中得到的随机噪声。相比于取噪声最大值的传统拉普拉斯机制，对于每位参与者来说，可以使用更少的噪声添加来提供更高的隐私保护力度。因此，相比于其余几种算法，POA 在相同隐私保护力度情况下的参与者所需要添加的噪声更少，从而带来了更高的数据效用。

图 11 展示了平台的收益随隐私预算的变化。当隐私预算很小时，任务的偏差无法满足消费者的需求，此时平台无法承接消费者的任务，因此收益为 0。

随着隐私预算逐渐增大，可以达到的任务偏差会逐渐减少，当满足任务需求时，平台开始承接消费者的任务。随着所承接任务数量的增长，平台的收益急剧增长。此时，由于支付给参与者的隐私成本最小，平台也可以获取最大的利润。随着隐私预算继续增大，平台从消费者处获取的任务报酬没有变化，但支付给参与者的酬金逐渐增长，平台的总收益与每笔任务的平均收益会逐渐减小，最终归于零收益。因此，在收到消费者提交的偏差需求及报酬后，平台会仔细挑选参与者，在满足消费者偏差需求的基础上，尽可能地减少隐私预算，以最优化自身的收益。而减小隐私预算，则意味着对参与者隐私保护的增强。

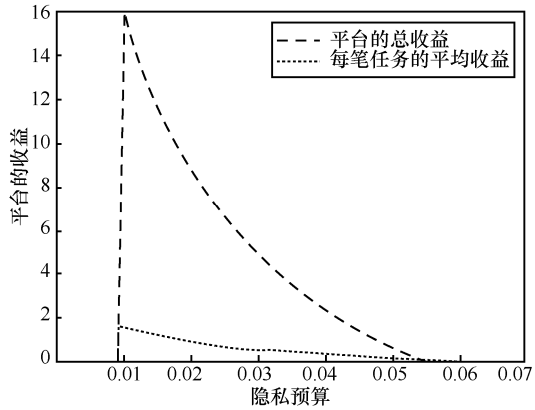


图 11　平台的收益随隐私预算的变化

图 12 展示了高信誉值、低信誉值和不考虑信誉值的情况下平台收益的变化。从图 12 可以看出，随着迭代次数的增长，3 种情况下平台的收益都逐渐增长，并最终达到稳定状态。当达到稳定状态时，高信誉值情况下的平台取得了最高的收益，信誉值的考虑确实提升了平台的收益，而且参与者的信誉值越高，提升的幅度越大。这也说明了平台会更倾向于选择具有高信誉值的参与者，从而督促参与者在任务中诚实表现以提升信誉值。

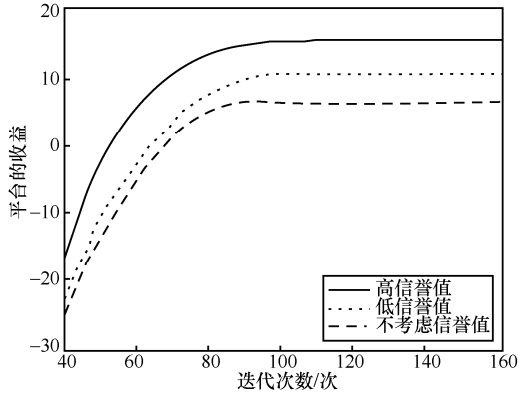


图 12　信誉对平台的收益的影响

3.4　消费者与参与者的变化对平台收益的影响

为展示消费者和参与者的数量变化对平台的收益的影响，给出了平台的收益及被选中的参与者数量随消费者数量的变化和平台的收益及完成的任务数随参与者数量的变化，分别如图 13 和图 14 所示。

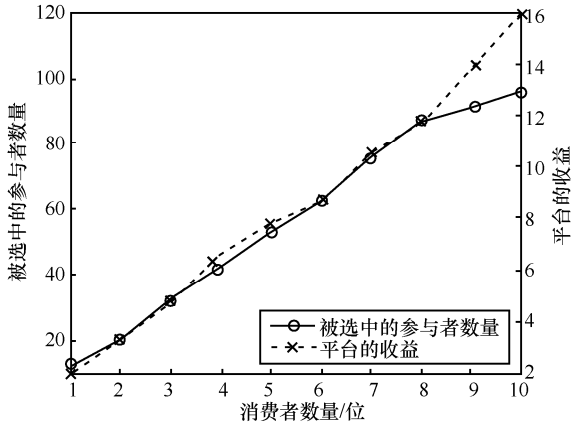


图 13　平台的收益及被选中的参与者数量随消费者数量的变化

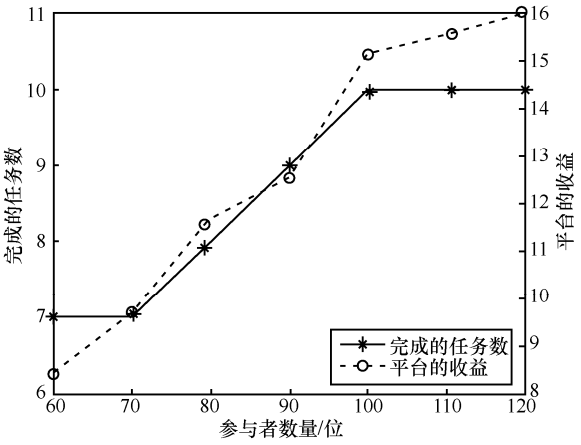


图 14　平台的收益及完成的任务数随参与者数量的变化

从图 13 可以看出，随着消费者数量的增长，被选中的参与者数量逐渐增加，所能完成的任务数随之增加，最终导致平台的收益也逐渐增加。从图 14 可以看出，随着参与者的数量增加，完成的任务数逐渐增多，当达到任务上限后，平台会对选取的参与者进行进一步优化，从而提高自身的收益。这说明消费者和参与者的增加都有利于提高平台的收益，这成为平台接受此模式并推动其发展的动机。

3.5　信誉模型的有效性验证

为了验证设计的信誉模型是否能够反映参与者在任务中的表现，本节设计了 2 种参与者，一种是普通的参与者，他会正常完成任务，并且在任务中逐步提升结果的精确度；另一种是恶意的参与

者，他开始也像普通的参与者那样表现，但达到一定次数后会发动攻击，即提供极差的结果，随后，恢复正常。两者的信誉值变化如图 15 所示。

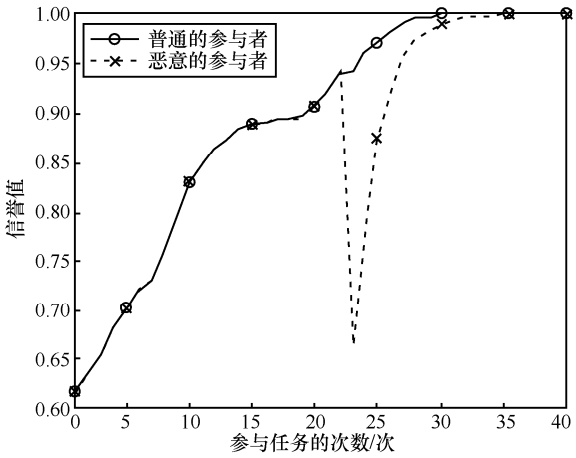


图 15 信誉值变化对比

从图 15 可以看出，随着结果精确度的逐渐提高，普通的参与者的信誉值随参与任务的次数的增多而逐渐增长，最终达到稳定。恶意的参与者的信誉值在他发起攻击时急剧下降，当他恢复正常后，他的信誉值也恢复正常的增长。这表明信誉模型对参与者的行为是敏感的。参与者“坏”的行为会体现在其信誉值上，从而促使参与者采取“好”的行为来提高信誉。

4 结束语

为了保护参与者的隐私，本文提出了一种隐私保护的群体感知数据交易框架。在此框架下，本文提出了基于差分隐私的聚合方案。参与者收集数据并进行计算，随后在计算结果中按照平台分配的隐私预算添加噪声，最后将分析结果发送给平台。这既保证了对参与者的隐私保护，又保障了聚合结果的可用性。此外，本文还提出了参与者的信誉模型，以保障参与者的可信性。通过考虑消费者的偏差约束和对参与者的隐私补偿，构建了以平台收益最大化为目标的交易优化模型，并提出了 POA 进行求解，从而激励用户参与数据交易。基于北京的天气和空气质量数据的实验表明，本文方案在实现参与者隐私保护的基础上提升了平台的收益。

参考文献：

[1] DAI W Q, DAI C K, CHOO K K R, et al. SDTE: a secure blockchain-based data trading ecosystem[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 725-737.

[2] HUANG Y D, ZENG Y M, YE F, et al. Fair and protected profit sharing for data trading in pervasive edge computing environments[C]//Proceedings of IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2020: 1718-1727.

[3] WANG X D, YING C H, LUO Y. Privacy-friendly decentralized data aggregation for mobile crowdsensing[C]//Proceedings of 2020 IEEE Global Communications Conference. Piscataway: IEEE Press, 2020: 1-6.

[4] YU J L, CHEUNG M H, HUANG J W, et al. Mobile data trading: behavioral economics analysis and algorithm design[J]. IEEE Journal on Selected Areas in Communications, 2017, 35(4): 994-1005.

[5] YANG G, HE S B, SHI Z G, et al. Promoting cooperation by the social incentive mechanism in mobile crowdsensing[J]. IEEE Communications Magazine, 2017, 55(3): 86-92.

[6] 杜小勇, 陈峻, 陈跃国. 大数据探索式搜索研究[J]. 通信学报, 2015, 36(12): 77-88.
DU X Y, CHEN J, CHEN Y G. Exploratory search on big data[J]. Journal on Communications, 2015, 36(12): 77-88.

[7] AN B Y, XIAO M J, LIU A, et al. Truthful crowdsensed data trading based on reverse auction and blockchain[C]//Database Systems for Advanced Applications, Berlin: Springer, 2019: 292-309.

[8] ZHENG Z Z, PENG Y Q, WU F, et al. Trading data in the crowd: profit-driven data acquisition for mobile crowdsensing[J]. IEEE Journal on Selected Areas in Communications, 2017, 35(2): 486-501.

[9] JIANG C K, GAO L, DUAN L J, et al. Scalable mobile crowdsensing via peer-to-peer data sharing[J]. IEEE Transactions on Mobile Computing, 2018, 17(4): 898-912.

[10] ZHENG Y F, DUAN H Y, YUAN X L, et al. Privacy-aware and efficient mobile crowdsensing with truth discovery[J]. IEEE Transactions on Dependable and Secure Computing, 2020, 17(1): 121-133.

[11] ZHANG L, LI Y N, XIAO X, et al. CrowdBuy: privacy-friendly image dataset purchasing via crowdsourcing[C]//Proceedings of IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2018: 2735-2743.

[12] GAO W C, YU W, LIANG F, et al. Privacy-preserving auction for big data trading using homomorphic encryption[J]. IEEE Transactions on Network Science and Engineering, 2020, 7(2): 776-791.

[13] NIU C Y, ZHENG Z Z, WU F, et al. Achieving data truthfulness and privacy preservation in data markets[J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 31(1): 105-119.

[14] ZHAO Y Q, YU Y, LI Y N, et al. Machine learning based privacy-preserving fair data trading in big data market[J]. Information Sciences, 2019, 478: 449-460.

[15] WANG J, WANG Y L, ZHAO G S, et al. Location protection method for mobile crowd sensing based on local differential privacy preference[J]. Peer-to-Peer Networking and Applications, 2019, 12(5): 1097-1109.

[16] XUE Q, ZHU Y W, WANG J. Mean estimation over numeric data with personalized local differential privacy[J]. Frontiers of Computer Science, 2021, 16(3): 1-10.

[17] 付钰, 俞艺涵, 吴晓平. 大数据环境下差分隐私保护技术及应用[J]. 通信学报, 2019, 40(10): 157-168.
FU Y, YU Y H, WU X P. Differential privacy protection technology

and its application in big data environment[J]. Journal on Communications, 2019, 40(10): 157-168.

[18] 郭艺, 叶剑, 张鹏. 基于偏差约减的大数据交易模型分析与修复方法[J]. 电子学报, 2018, 46(7): 1754-1761.
GUO Y, YE J, ZHANG P. Analysis and repair of big data transaction model based on deviation reduction[J]. Acta Electronica Sinica, 2018, 46(7): 1754-1761.

[19] YU H F, ZHANG M X. Data pricing strategy based on data quality[J]. Computers & Industrial Engineering, 2017, 112: 1-10.

[20] JIAO Y T, WANG P, NIYATO D, et al. Profit maximization auction and data management in big data markets[C]//Proceedings of 2017 IEEE Wireless Communications and Networking Conference. Piscataway: IEEE Press, 2017: 1-6.

[21] 黄小红, 张勇, 闪德胜, 等. 基于多目标效用优化的分布式数据交易算法[J]. 通信学报, 2021, 42(2): 52-63.
HUANG X H, ZHANG Y, SHAN D S, et al. Distributed data trading algorithm based on multi-objective utility optimization[J]. Journal on Communications, 2021, 42(2): 52-63.

[22] ZHANG J X, SUN J C, ZHANG R, et al. Privacy-preserving social media data outsourcing[C]//Proceedings of IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2018: 1106-1114.

[23] CAI H, YE F, YANG Y Y, et al. Towards privacy-preserving data trading for web browsing history[C]//Proceedings of the International Symposium on Quality of Service. Piscataway: IEEE Press, 2019: 1-10.

[24] YANG L, ZHANG M Y, HE S B, et al. Crowd-empowered privacy-preserving data aggregation for mobile crowdsensing[C]//Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing. New York: ACM Press, 2018: 151-160.

[25] DWORK C. Differential privacy[C]//Automata, Languages and Programming. Berlin: Springer, 2006: 1-12.

[26] KEMP F. The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance[J]. Journal of the Royal Statistical Society: Series D (the Statistician), 2003, 52(4): 698-699.

[27] DANDEKAR P, FAWAZ N, IOANNIDIS S. Privacy auctions for recommender systems[C]//Internet and Network Economics. Berlin: Springer, 2012: 309-322.

[28] JIN H M, SU L, XIAO H P, et al. INCEPTION: incentivizing privacy-preserving data aggregation for mobile crowd sensing systems[C]//Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing. New York: ACM Press, 2016: 341-350.

[29] MENG C S, JIANG W J, LI Y L, et al. Truth discovery on crowd sensing of correlated entities[C]//Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems. New York: ACM Press, 2015: 169-182.

[30] KAZAI G, KAMPS J, MILIC-FRAYLING N. An analysis of human factors and label accuracy in crowdsourcing relevance judgments[J]. Information Retrieval, 2013, 16(2): 138-178.

[31] STRUBELL E, GANESH A, MCCALLUM A. Energy and policy considerations for modern deep learning research[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(9): 13693-13696.

[32] WANG L Y, ZHANG D Q, YAN Z X, et al. effSense: a novel mobile crowd-sensing framework for energy-efficient and cost-effective data uploading[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2015, 45(12): 1549-1563.

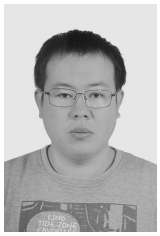
[33] 汪民乐. 遗传算法的收敛性研究[J]. 计算技术与自动化, 2015, 34(1): 58-62.
WANG M L. Research on convergence of genetic algorithm[J]. Computing Technology and Automation, 2015, 34(1): 58-62.

[34] GAO G J, XIAO M J, WU J, et al. DPDT: a differentially private crowd-sensed data trading mechanism[J]. IEEE Internet of Things Journal, 2020, 7(1): 751-762.

[35] CAI H, ZHU Y M, LI J, et al. A profit-maximizing mechanism for query-based data trading with personalized differential privacy[J]. The Computer Journal, 2020, 64(2): 264-280.

[36] ZHENG S Y, CAO Y, YOSHIKAWA M. Trading data with personalized differential privacy and partial arbitrage freeness[J]. arXiv Preprint, arXiv:2105.01651, 2021.

[作者简介]



张勇（1990− ），男，河北衡水人，北京邮电大学博士生，主要研究方向为区块链、大数据交易和数据隐私保护等。

李丹丹（1987− ），女，河南平顶山人，博士，北京邮电大学讲师，主要研究方向为网络安全、密码学。

韩璐（1991− ），女，蒙古族，内蒙古赤峰人，北京邮电大学博士生，主要研究方向为安全多方计算、联邦学习等。

黄小红（1979− ），女，广东广州人，博士，北京邮电大学教授，主要研究方向为计算机网络应用、下一代互联网和网络安全等。