群智授权隐私保护的移动群智感知数据聚合



|  |  |
| --- | --- |
| **摘要**  我们开发了一种用于移动群智感知中隐私保护数据聚合的拍卖框架，在该框架中，平台扮演拍卖师的角色，为感知任务招募工人。观察该基于拍卖的框架发现，其存在工人数据隐私的外部性。这是因为每个工人的数据隐私取决于其注入的噪声和聚合结果中的总噪声，而这与选择哪些工人来完成任务密切相关。为了以经济有效的方式获得理想的数据聚合准确度水平，我们明确地描述了外部性，即每个工人添加的噪声对数据隐私和聚合结果准确性的影响。进而，我们探索了问题结构、刻画了问题的隐性单调性，并确定了工人的临界报价，这使得设计一个真实、个体合理、计算效率高的激励机制成为可能。拟议的激励机制可以招募一组工人，以根据聚合结果的准确性要求尽量最小化从工人购买私人感知数据的成本。我们通过理论分析和大量模拟来验证该方案的有效性。  **关键词** 群智感知、激励机制、隐私保护、数据聚合    一个关键问题是，当工人报告噪声数据时，如何以经济有效的方式达到理想的数据聚合准确度水平。由于多个Nash Equilibria（例如，〔24〕）的存在，博弈理论模型不能保证数据聚合的理想准确度水平，因此在设计激励机制时，本文采用了考虑准确性要求的拍卖方法。然而，允许工人在其感知结果中加入噪声，这给激励机制设计带来了三大挑战：   * *战略性行为*。当工人报告噪声数据时，平台不知晓真实感知数据和附加噪声。因此，在数据聚合阶段，工人可以通过在其感知数据中添加更多噪声来增强其数据隐私性，从而实现战略性操作。因此，需要一种新的数据聚合方案，使平台能够在不知晓工人数据真实感知数据的情况下控制工人数据的噪声水平。此外，工人可以操纵他们的报价以最大化他们自身的利益，而这可能会导致实现理想数据聚合准确度水平的高成本。因此，需要一个真实的激励机制。 * *外部性。*现有的工作（例如，[15]）将噪声添加到工人的感知数据中，并且工人的数据隐私仅取决于平台添加的噪声，与此相比，本文中每个工人的数据隐私不仅取决于自身添加的噪声，还取决于聚合结果中的总噪声（见第2.3节）。换言之，每个工人的数据隐私取决于选择哪些工人来完成任务以及所选工人产生噪声量，而这将引入外部性。这使得本文中的激励机制设计更具挑战性。 * *计算复杂性.*为了以经济有效的方式达到理想的聚合结果准确度水平，平台需要找到一组最佳工人来完成感知任务。由于不同的工人对其数据隐私的评估不同，而工人的数据隐私又因外部性而相互依赖，因此寻找最佳工人群组最小化系统成本，同时达到理想的准确度水平具有组合性质。因此，需要一种高效的计算机制。   **1.2 主要结果汇总**  本文开发了一个移动群智感知中隐私保护数据聚合的拍卖框架。在该框架中，工人向平台提交他们的报价，平台扮演拍卖人的角色以招募工人完成感知任务。该平台在对来自工人的噪声数据进行聚合时，旨在最大限度地降低从工人那里购买隐私感知数据的成本，同时达到理想的聚合结果准确度水平。我们的主要结果总结如下： | **CCS概念** • **安全与隐私**→ **隐私保护**；• **网络** →**网络经济**； • **以人为中心的计算** → **移动计算**； • **计算理论** → **算法的设计和分析**；  **1简介 1.1 诱因**  移动群智感知作为一种很有前景的传感模式应运而生，它利用人类所携带移动设备的感知能力来执行各种感知任务（例如，医疗保健、环境监测、室内定位和智能交通）[22]。 通过将感知任务外包给公众群体，移动群智感知系统能够有效地收集精细信息。然而，任何参与感知任务的个人都必须授权任务代理访问其敏感的感知数据，这导致在数据发布给不可信一方时可引起隐私问题。这成为阻碍个人（工人）参与的一个主要挑战，而不仅仅是其移动设备有限系统资源（如电池和计算能力）的消耗。因此，移动群智感知的成功取决于设计有效的激励机制以刺激工人参与。  针对移动群智感知系统，已开发出许多激励机制（例如，[1, 4, 7, 10–12, 14, 15, 21, 25–30]）。大多数现有成果只考虑工人的感知成本，而只有最近的一些作品考虑到了工人的隐私成本。然而，在这些考虑工人隐私成本的成果中，要么工人无法控制他们的数据隐私（例如，[15]中平台被假定为可信并完全负责保护工人的隐私数据），要么平台通过博弈论模型与工人交互（例如，[24]），而这可能会导致低效率的平衡，即平台可能无法达到聚合结果的理想准确度水平。为解决上述问题，我们需要一种新的用于移动群智感知的数据聚合方法，使每个工人不仅能够保护自己的数据隐私，而且平台还可以根据其感知质量有选择性地选择工人，进而达到聚合结果的理想准确度水平。解决这些问题的一个可能办法是允许工人通过报告噪声数据来保护他们的数据隐私[24]。显然，这种方法会对感知结果的可靠性产生负面影响。1为保证聚合结果的准确性，平台需要设计更有效、考虑工人隐私保护行为的激励机制，以便在工人的数据隐私和聚合结果的准确性之间实现良好的平衡。   * *差异隐私数据聚合。*为了应对工人战略性行为带来的挑战，我们利用著名的差异隐私概念，提出了一个差异隐私数据聚合方案。其核心思想是根据Laplace分布的可分割性，为每个工人精心设计噪声分布，使每个工人都能根据设计的噪声分布报告其数据的隐私保护版本，同时平台保证每个工人数据的差异隐私。通过利用该方案，平台可以在不了解工人真实感知数据的情况下，防止工人在感知数据中战略性地加入较大噪声，以及控制其数据的噪声水平。 * *外部性*。在所提议的差异隐私数据聚合方案下，将为不同工人群组的工人设计不同的噪声分布。换言之，如果平台选择不同的工人，则工人的隐私会发生变化，从而引入外部性。对于Laplace噪声分布，我们明确描述了工人之间的外部性以及每个工人的参与对其他工人隐私的影响，而这也纳入了激励机制设计考虑中。 * *隐私准确性权衡*。为了保证聚合结果的准确性，如果工人报告的数据准确性较高（即，添加的噪音更少），则平台将给工人更多的奖励。显然，在（隐私）成本和准确性之间需要权衡。基于差异隐私的概念，我们描述了工人数据隐私和聚合结果准确性之间的权衡。由于工人所添加噪声的影响，聚合结果的准确性则表现为失真。 * *差异隐私数据拍卖。*根据所提出的差异隐私数据聚合，激励机制的设计归结为解决隐私拍卖问题，即在聚合结果准确性的约束下，将感知任务分配给工人总报酬最小化的工人群组。结果表明，求解该问题的最优解为NP-hard。通过探究问题结构，我们发现了问题的*隐单调*性，并确定了工人的临界报价。根据上述发现，尽管问题具有组合特性，我们提出了一个计算效率高的差异隐私数据拍卖方案。此外，我们还证明了所提出的差异隐私数据拍卖方案是真实、个体合理的，并且接近最优解。通过大量模拟，对该方案的性能进行了评估。   **1.3 相关工作**  移动群智感知系统的激励机制设计最近受到了广泛的关注（例如， [1, 4, 7, 11, 12, 14, 15, 21, 25–30]）。不同的模型（例如，拍卖[7,14,15,25-30]和博弈论模型[1,4,11,12,21]）被用来设计具有不同目标的激励机制，包括社会福利最大化（例如，[1,8,14]）、成本或报酬最小化（例如，[7,15]）和平台利润最大化（例如，[11,27]）。大多数现有的作品（例如，[ 7, 14, 25，26, 28 - 30 ]）只考虑参与者的感知成本。 |

|  |  |
| --- | --- |
| 近年来，人们对数据隐私给予了极大的关注（例如，[2,3,9,10,15,24,27]）。这些作品（例如，[2，3，9，15，27]）大多认为到平台（即数据采集器）是可信任的，并将真实的数据报告给平台，但在平台上工人无法控制其数据隐私。 近期的作品[23，24]允许工人通过报告噪声数据来保护他们的数据隐私，并研究如何在博弈论模型中交换隐私数据，然而，这可能会导致低效率的均衡，即无法保证聚合结果的准确性。为了解决这些问题，本文提出了一种移动群智感知新的拍卖框架。其中，工人可以根据提议数据聚合方案确定的噪声分布，通过添加噪声来保护其数据隐私，并且平台可以选择工人，以最小化数据聚合达到准确度水平所需的成本。这使得我们的方案有别于现有的关于移动群智感知数据聚合的作品。  下文安排如下。在第二部分，我们描述了移动群智感知系统中的群组授权隐私保护数据聚合。在第三部分，我们提出了激励机制，并对其性质进行了分析。在第四部分中，我们评估了所提出的激励机制的性能。第五部分对本文进行了总结。  **2移动群智感知中的隐私保护数据聚合**  **2.1系统概述**  设想一个由一个集中平台a、一个任务代理T和一组参与的工人N{1、…、N}组成的移动群智感知系统，如图1所示。这项任务要求工人向平台报告其特定物体或现象的本体感知数据（例如光谱感知和环境监测）。由于每个工人感知数据的可靠性可能因传感器质量不同而有所差异，故为了提高结果的可靠性，平台将聚合感知数据[15]。*与移动群智感知系统现有的拍卖方面作品（例如，[ 1, 4, 7，10 - 12, 14, 15，21, 25 - 30 ]）不同，我们允许每个人报告其隐私保护版本的数据，以保护其自身的数据隐私[24 ]。*    **3.应支付额**  **1.任务**  **2. 报价**  **工人**  **每个中标者的感知报告=感知数据+噪声**  **奖励＜成本**  **奖励≥成本**  **6. 隐私保护感知数据**  **5. 数据报告需要**  **激励机制**  **数据聚合**  **4. 中标者集合**  **群智感知平台**  **7. 发布结果**  **3. 应支付额**  **1.发布任务**  **任务代理**  **图1：群组授权的隐私保护数据聚合框架。** | * 首先，任务代理在群智感知平台中发布一个任务，然后将该任务发布给一个包含N个工人的群组，表示为N（步骤1）。 * **激励机制.**然后，工人向平台提交他们的报价（步骤2），其中报价反映了每个工人隐私损失的评估（见第2.2.1节）。基于该激励机制，平台确定中标者，即被选中去完成任务的工人，以及中标者相应的报酬（步骤3和4）。 * **数据聚合.**接下来，平台将数据报告要求发送给中标者，并允许中标者报告其感知数据的隐私保护版本（步骤5和6）。 * 最后，平台将聚合结果发布给任务代理（步骤7）。   **2.2 群智感知拍卖模型**  在群智感知系统中，平台扮演拍卖师的角色，负责招募工人完成感知任务并对感知数据进行聚合。  首先，平台（拍卖商）将从工人那里获得单位隐私成本的报价。工人的隐私成本通过差异隐私进行量化（见第2.2.1节）。设b=（b1，…，bN）表示工人提交的报价向量，b-i表示没有工人i报价的报价向量。在不失去一般性的前提下，我们假设工人的报价是按递增顺序排列的，即*b*1 ≤ *b*2 ≤ · · · ≤ *bN* 。为了防止可能导致系统性能下降（如高成本）的报价操纵，需要一个真实的激励机制（将在第3节中讨论）。此外，由于平台不知道工人添加的噪声，因此在数据聚合阶段，工人可通过在感知数据中添加更多噪声来增强数据的隐私性，从而进行战略性的操作。因此，一种新的数据聚合方案有待研发，这将在第2.3节中讨论。  然后，拍卖输出分配结果（x，p），其中x=（x1，…，xN）表示参与者，p=（p1，…，pN）表示支付给参与者的金额。具体地说，*xi* ∈ {0, 1}表示工人*i*是否选择来执行任务：*xi* = 1意味着被选择，*xi* = 0则表示未选中。因此，我们定义S为具有S个工人的中标者群组。对于每个工人，平台将向工人i∈N支付pi奖励金额以从工人i收集隐私数据，并在数据聚合后以差异隐私方式使用数据（见第2.3节）。  *2.2.1隐私成本。*根据差异隐私的效用理论特征[9]，隐私成本可以建模为具有真实数据向量的效用和具有扰动数据向量的效用之间的差异，它亦是工人隐私*ϵi*的线性函数。设*vi* >0表示工人*i*对单位隐私成本的内在估价。 工人*i*的隐私成本为：  *ci* = *viϵi*（**x**）. （1）  直观地说，*vi*值越大，则表明工人*i*通过暴露其感知数据对隐私损失的评估越高。我们假设所有单位隐私成本*vi*对于平台或其他工人都是未知的。*此成本函数已被发现用于许多现有作品中（例如，[ 3, 9, 15，27 ]）。然而，本文（1）中工人隐私ϵi是x的函数，它不仅取决于工人自己添加的噪声，也取决。* |

|  |  |
| --- | --- |
| *于聚合结果中的总噪声，这就入了外部性（见第2.3节）。这是这项作品与移动群智感知（例如，[15]）中其他相关作品的一个主要区别，即工人的隐私成本完全取决于其自身的参与*具体而言，提议群组授权隐私保护数据聚合的工作流程（见图1）如下：  *2.2.2 工人的效用。*在群智感知系统中，假定工人自私且有战略性，以最大化自己的效用。基于隐私成本（1），我可以将工人*i*的效用*ui*设为，  *ui*（*bi*, **b**-*i*） = *pi*（*bi*, **b**-*i*） - *ci* = *pi*（*bi*, **b**-*i*） - *viϵi*(**x**), (2)  其中*ui*和*pi*是b、给定*ϵi*和x的函数。此处，对于非参与者*i* ∈ N （i.e., *xi* = *ϵi* = *pi* = 0），其效用为零。注意，为了简化表示，我们没有明确地将执行任务的感知成本包括在效用函数（2）中。同时，本文的结果可以很容易地推广，以包含如[5，12]中的感知成本。例如，类似于[12]，设*si为*用户*i*的感知成本，我们可以将用户*i*的个体效用修改为*ui* = *pi* - *si* - *ϵivi*，并定义*pi* = *pi* -*si* 以将感知成本包含在报酬中。因此，我们的结果可以推广到这个案例。  *2.2.3 设计目标.* 我们旨在设计一种基于拍卖的分配机制，通过设计一种具有以下理想属性的激励机制（请参见第3部分），以在获得理想的数据聚合准确性的同时，最小化对工人的总报酬：   * 真实性：每个工人*i*都能通过真实地竞标其隐私权评估来最大化其效用，即任何b的*ui*（*vi*, **b**-*i*） ≥ *ui*（*bi*, **b**-*i*）。 * 个人理性：每个工人i都能获得非负效用，即*ui* = *pi* - *ci* ≥ 0。 * 成本最小化：该机制可以最小化对工人的总报酬。 * 计算效率：解（x，p）可以用多项式次数计算。   **2.3差异隐私数据聚合**  在聚合数据时，每个中标者i将通过添加随机噪声*ni*来报告其数据*di*的隐私保护版本dˆi。在不失去一般性的情况下，我们假设所有感知数据*di*都是在[0，1]范围内的标准化值。在本文中，我们设想了一个加权聚合运算*f*来计算基于工人数据的聚合结果*r*。设*d*为工人数据的向量。聚合结果*r*可以写成  *r* = *f* （**d**） = *i* ∈N *wi*（*di* + *ni*）*xi* = *i* ∈S *wi*（*di* + *ni*）, (3)  其中，*wi>*0是worker *i*的标准化权重，因此这些权重之和等于1。与[15，17，19]类似，加权聚合是为了捕捉工人不同技能水平对聚合结果计算的影响。直观地说，那些感知数据更可能接近真实情况的工人，将被分配更高的权重。这使得聚合结果更接近由更可靠工人提供的数据，而这也已被许多最先进的数据聚合方法 [15、17、19] 所使用。权重的选择可以如 [15]中基于工人的技能水平，这是平台和工人已知的*先验知识*。 | 本文基于著名的差异隐私概念对数据聚合中的隐私损失进行了量化[6]，并且提议的差异隐私数据聚合定义如下。  定义1（差异隐私数据聚合）。*聚合运算f：[0，1]S→R对于工人i是ϵi-差异的隐私，如果对于仅在第i个工人的数据中不同的任何一对相邻向量d和d（i）和任何一组聚集结果O⊆ Ranдe*（*f* ）*，则以下不等式成立：*  *Pr*[*f* （**d**） ∈ *O*] ≤ exp（*ϵi*）*Pr*[*f* （**d**（*i*）） ∈ *O*], （4）  *其中 ϵi 是正参数。*  因此，在运算*f*下，工人i的数据以*ϵi*差异隐私方式使用。这一定义与 [6]中的定义略有不同，后者是以最坏情况下的隐私（即ϵ-差异隐私，其中ϵ=sup*iϵi*）来表述的。  考虑到聚合操作*f*，一种提供差异隐私的著名方法是将从Laplace分布提取的随机噪声添加到此函数中[6]。因为我们允许每个工人自行添加噪声，我们需要仔细设计每个工人的噪声分布，以使这些噪声的和等于从Laplace分布中提取的随机噪声，即聚合噪声*n=i∈S wini*遵循Laplace分布  命题1. *对于（3）中的聚集运算f，定义ϵi=si（f）/σ，其中si（f）=maxd，***d**（*i*） ∈[0,1]*S* |*f* （**d**） - *f* （**d**（*i*））|*是f对第i项di的灵敏度，σ是Laplace分布的参数。如果所有i∈S的ni=G1（S，σ/wi）-G2（S，σ/wi）是独立的，则聚合运算f相对于工人i是ϵi-差异隐私的，其中G1（S，σ/wi）和G2（S，σ/wi）是i.i.d.服从pdf* *的gamma分布的随机变量。*  证明. 为证明命题1，只要证明聚合噪声服从Laplace分布就足够了。基于Laplace分布的可除性[16]，Laplace分布是可除的，可以构建为i.i.d. gamma分布的和。根据gamma分布的标度律，*wini* = *G*1（*S*, *σ*） - *G*2（*S*, *σ*）。因此，我们有    其中，第二等式遵从Laplace分布的可除性[16]，从而得出该证明。  基于命题1，如果每个工人的噪声分布均是精心设计的，那么（3）中聚合运算*f*相对于工人*i*是*ϵi*-差异隐私的。因此，我们在算法1中提出了数据聚合机制。在算法1中，平台只需要通知工人*S*和*σ/wi*的值，每个工人根据该值产生随机噪声并将*dˆi*反馈给平台。  **评论：** |

|  |  |  |  |
| --- | --- | --- | --- |
| |  | | --- | | **算法 1** 差异隐私数据聚合 | | 1： **输入：** 工人集合 S，工人数量 *S，*每个工人的权重 *wi*, ∀*i* ∈ S，Laplace 分布参数 *σ。*  2： **输出：** 聚合结果 *r。*  3： 对于每个工人i∈S，平台告知参数S和*σ/*wi的值。  4： 每个工人根据*G*1（*S*, *σ*/*wi*）-*G*2（*S*, *σ*/*wi*）分布生成一个随机噪声*ni*，然后向平台报告*dˆi=di+ni*。  5： 平台使用（3）来聚合工人给出的数据，并将聚合结果r发布给任务代理。 |  * 请注意，在提出的数据聚合算法中，平台并不知道工人数据的真实值，而是知晓其数据的隐私保护版本，该版本是根据工人在提出的拍卖框架中赞同的噪声分布生成。通过如此，所提出的算法不仅可以让工人报告噪声数据以保护他们的隐私，而且还可以防止工人战略性地将大噪声添加到他们的感知数据中，因为检查每个工人报告的分布是否遵循指定噪声分布是很容易的。 * 请注意，对于不同的中标者集合，将给中标者我分配不同的噪声分布。换言之，每个中标者的隐私取决于中标者集合的选择，这就引入了外部性。这使得本文的激励机制的设计不同于移动集群感测系统中现有的拍卖作品。   **2.4隐私与准确性**  当允许工人报告噪声数据时，添加到聚合结果中的噪声必然会降低结果的准确性。从命题1，我们观察到*ϵi*依赖于*σ*的值。*σ*值越高，*ϵi*越小，因此，隐私保证越好。但是，*σ*值越高，聚合结果的准确性就越低。*显然，工人的数据隐私和聚合结果的准确性之间存在着自然的权衡。*  为了描述准确性，我们在两个聚合函数（一个使用所有工人的无噪声数据，另一个使用选定工人的有噪声数据（即（3）中的聚合结果*r*））之间引入了失真概念。由于平台需要为工人的数据付费，因此获取所有工人的数据将非常昂贵，而且工人亦将添加噪音以保护他们的数据隐私。因此，我们可以将所有工人的无噪声数据的聚合作为基准。  定义2（失真）.*给定向量x，失真δ（x）定义为*    在定义2中，失真定义为工人报告的任何感知数据与真实结果之间的最大预期偏差值。显然，失真取决于完成任务的工人集合以及添加到数据中的噪声。 | 它们的依赖性通过以下命题来量化。  命题2（隐私与失真）.*给定所有工人的xi和wi，在聚合函数（3）下，每个工人的隐私和聚合结果的失真可以表示为*    证明。给定聚合函数（3）下的*xi*和*wi*，我们得到    因此我们得到。对于失真，我们得到    其中（a）遵从等式（3）得出*∑i∈Nwinixi = ∑i∈Swini*，而（b）遵从命题1得出*∑i∈Swini*是具有零均值和2*σ*2方差的Laplace随机变量。  从命题2可以清楚地看出，给定*σ*，完成任务的工人越多，失真越小；给定所选工人集合S，*σ*的值越高，则*ϵi*越小（即隐私更好）且失真越小。根据[3]，如果工人添加的Laplace噪声具有以下形式的参数，我们将（3）中的聚合操作*f*称为*规范的*    基于（9），每个工人的隐私和聚合结果的失真可以表示为    方程式（10）和（11）引入工人之间的*外部性*，使得工人*i*的数据隐私依赖于其他工人的参与。具体地说，参与者越多，失真越小，但*ϵi*越大（即，隐私越差）。直观地说，由于所有的工人完成相同的感知任务，因此参与者越多，就越容易找出真实的数据（即，隐私损失越大）。此外，因为工人有不同的技能水平（即*wi*），会造成不同程度的失真，因此我们需要仔细选择工人。此外，选择不同工人的成本也不同。因此，寻找合适的工人集合来完成感知任务是一项具有挑战性的任务。  **3激励机制**  **3.1数学公式**  群智感知拍卖的目标是最小化对工人的总付款，从而使聚合结果的准确性高于特定的预定阈值（换句话说，失真低于阈值Δ）。具体来说，这个问题可以表示为 |

|  |  |
| --- | --- |
| 最小化 *i* ∈N *pi*  服从 *pi* ≥ *biϵi*（x）, ∀*i* ∈ N, （个体合理性）  *δ*（x） ≤ Δ, （准确度要求） 12  *xi* ∈ {0, 1}, ∀*i* ∈ N.  在问题（12）中，个体合理性的限制保证了每个工人都能获得非负效用。对于准确性要求的限制，阈值通常会决定工人的总付款和隐私保护水平。在低阈值（即高准确度）的情况下，平台将向工人支付更多费用以获得更少噪声的数据（即工人的隐私更差）。请注意，与大多数关于群智感知的作品不同，问题（12）考虑了工人之间的外部性，使得工人的数据隐私相互依赖。这已在第2.3节和第2.4节中进行了讨论。由于外部性，设计解决问题（12）的激励机制是一个具有挑战性的任务。定理1表明问题（12）为NP-hard。  定理 1. *群智感知拍卖问题（12）为NP-hard。*  为了证明定理1，我们首先建立问题（12）与下列问题之间的等式：  最小化  服从  其中  引理1。*问题（12）的最优分配x\*与问题（13）的相同。*  证明：注意，为了最小化（12），*pi*始终等于*biϵi（x∗）*。因此，个体合理性的不平等是严密的。换句话说，最小化*i∈N pi*等价于最小化*i∈N biϵi（x）*。接下来，我们可以在某些代数之后重写限制*δ*（x） ≤ Δ as *i* ∈N *wixi* ≥ *W*，从而得出该证明。  很容易证明问题（13）可以归结为NP-hard的反向二进制背包问题。根据引理1，定理1成立。  **3.2机制设计**  根据定理1，当N的基数较大时，问题（12）在计算上是困难的。为了应对该挑战，我们提出了一种高计算效率的机制（见算法2和算法3），即差异隐私数据拍卖（DPDA）。如第3.3节所述，该机制真实且个体合理性，并能够找到问题（12）接近最优分配x\*的中标者集合。  在算法2中，其目的是首先找到问题（13）的分数松弛的解C，即    其被用以作为目标成本。基于目标成本C，可以通过选择总成本大于或等于C的最小工人集合来确定中标者集合。因为问题（14）比问题（13）约束更小，因此C是问题（13）的解的下界。为了找到最小的工人集合，我们探究了问题（14）的解决结构。基于问题（14）与问题（13）间的关系，我们发现了单调性的性质（见附录定理4的证明）。在此基础上，通过逐步将工人加入中标者集合，直到总成本大于或等于目标成本（见算法2中的主循环（第6-10行）），最终找到中标者集合。本 | 质上，我们想找到最小k之，以使∑*i* ≤*k biwi*/（∑*i* ≥*k*+1 *wi*） ≥ *C*，即*k* = min{*j* ： ∑*i* ≤*j biwi*/（∑*i* ≥*j*+1 *wi*） ≥ *C*, ∀*j* ∈ N }，所有i≤k的工人都在中标者集合中。注意，由于外部性，这个单调性隐藏在问题（13）中，这使得我们的问题在技术上比现有的移动群智感知拍卖更具挑战性。  在算法3中，我们利用拍卖理论中的临界值方法[20]。这样做的目的是确定主要报价bc，致使工人的出价大于或等于bc时就不会被选中。具体来说，我们首先从worker集合N中移除工人i，并确定使工人不是中标者的最小报价。这在每个工人上均会实施（算法3中的第5行）。注意，报价是递增排列的。临界报价是根据所有报价的最高价确定的（算法3的第6行）。利用这一临界报价，我们根据每个中标者的权重确定其报酬（算法3中的第8行）。从第3.3节对DPDA的分析可以看出，算法2和算法3给出的解是可行的，并且接近问题（12）的最优解。  对于算法2的复杂性，我们需要对线性分数程序问题（14）求解C。为了有效地求解C，我们可以基于以下引理将问题（14）转化为线性程序。  引理 2.*问题（14）等同于下列线性程序：*    证明。为了证明等价性，我们将证明问题（14）中的任何可行点在目标值相同的问题（15）中也是可行的，反之亦然。我们注意到如果x在问题（14）中是可行的，那么 ，∀*i* ∈ N，并且z= 在问题（15）中是可行的，并具有相同的目标值∑*i* ∈N *biwiyi* = ∑*i* ∈N *biϵi*（x）。由此可知，问题（14）的最优值大于或等于问题（15）的最优值。相反，注意问题（15）中的z>0。如果yi和z在问题（15）中是可行的，则*xi* = *yi*/*z*具有相同的目标值∑*i* ∈N *biϵi*（x） = ∑*i* ∈N *biwiyi*，并在问题（14）中是可行的。因此，问题（14）的最优值小于或等于问题（15）的最优值。因此，问题（14）等价于问题（15）。 |

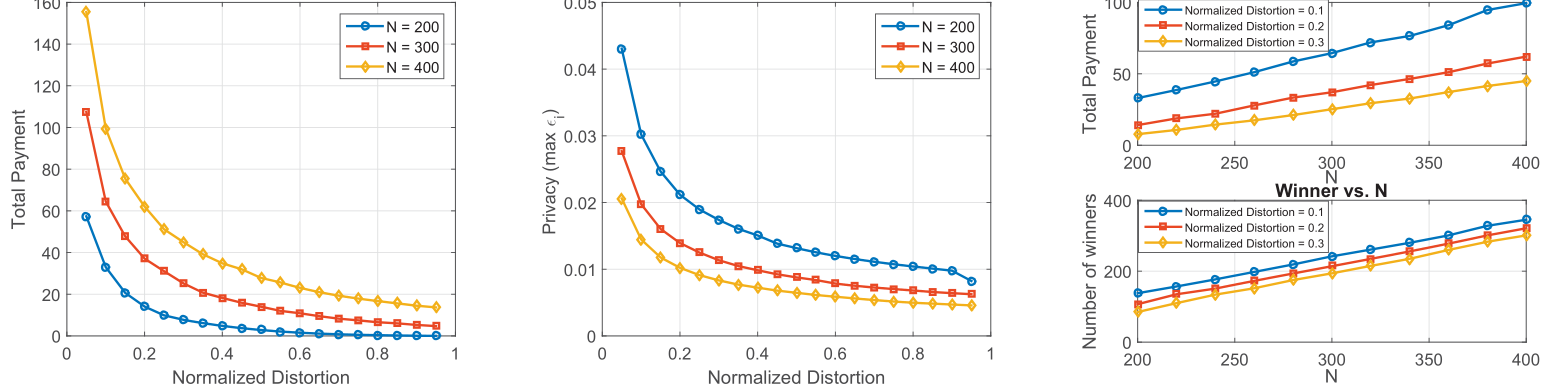
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 基于引理2，我们可以通过解一个线性程序（15）来解C。注意，算法2的计算复杂度由两部分组成：求解线性程序（15）（第3行）和找到一组中标者（第6-10行）。为了有效地求解（15），我们可以对线性程序使用多个解算器，例如CPLEX[13]，它可以在多项式次数内求解线性程序（15）[18]。要找到中标者，最坏的情况下最多需要*O*（*N*）次。因此，算法2可以在多项式次数内确定问题（12）的中标者集合。 对于算法3，它需要为每个中标者运行算法2，最坏的情况是运行N次，这意味着它也可以在多项式次数内求解。   |  | | --- | | **算法 2 差异隐私数据拍卖：确定中标者** | | 1： 输入： 工人集合N，每个工人的权重 wi，∀i ∈ N，每个工人的报价bi，∀i ∈ N。  2：输出： 中标者集合S。  3：通过解决问题（15）找到目标成本C。  4： 设k = 1，x1 = 1 且xi = 0，∀i = 2, . . . , N。  5： 设S = {1} 且计算C ‘= b1ϵ1（x）。  6： 当C’ < C时， 执行 \\ 查找中标者集合。  7. k = k + 1。  8. 设xk = 1 ，且S = S ∪k  9. C’=∑ki=1biϵi（x）  10： 结束  11： 返回 S. |  |  | | --- | | **算法 3差异隐私数据拍卖：确定付款额** | | 1： **输入：**工人集合N，每个工人的权重 wi，∀i ∈ N，每个工人的报价bi，∀i ∈ N，中标者集合S。  2： **输出：** 付款额 p。  3： 设 p = （0, . . . , 0）且*bc* = *bk*+1,其中k是S中报价最高的工人指数。  4：对于每个i∈S， 执行\\ 查找临界报价。  5： 在N \ {*i*}上运行算法2，以获得中标者集合S'，其中k’是报价最高的S'中的工人指数。  6： *bc* = min{*bc*,*bk*+1}。  7： **结束**  8： 对每个  9： 返回p |   **3.3 DPDA分析**  在这部分中，我们将证明DPDA是真实的、个体合理性的，并且相对于最优成本为*α*-近似。  首先，我们分析了DPDA的真实性。  定理 2. *DPDA 是真实的。*  证明。要证明DPDA是真实的，只需证明用户不能通过将其报价偏离真实估价来提高其效用。请注意，在DPDA中，中标者由其报价在集合N中的排名决定，排名越高，被选中的机会越低。此外，由算法3确定的临界报价不依赖于中标者的报价值。接下来，我们讨论工人i报价*b*˜*i*不真实的情况。 | * 报价过高 *b*˜*i* > *vi*。在这种情况下，工人i的排名可能会向后移。如果她能通过真实的报价*vi*赢得拍卖，并且她通过高报价仍然位列中标者集合中，那么她的效用将保持不变，因为由算法3确定的临界报价bc保持不变；如果她通过高报价输掉了拍卖，则她的效用将为零。 如果她以真实的报价输掉了拍卖，那么她仍然会因高报价而输掉。不管是哪种情况，工人i都不能提高她的效用。 * 报价过低 *b*˜*i* < *vi*。在这种情况下，工人i的排名可能会向后移。如果她能够通过真实的报价vi赢得拍卖，那么她的效用则无法提高，因为她必须继续保持在中标者集合中，并且临界报价仍保持不变。如果她因真实的报价而输掉了拍卖，但低报价有助于她成为中标者，她的效用将是。由于她最初不在中标者集合中，这意味着vi≥bc，这将导致她的效用ui≤0。   因此，DPDA 是真实的。  接下来，我们分析了DPDA的个体合理性。  定理 3。*DPDA属于个体合理。*  证明。 对于中标者集合中的每个工人，我们有    由于*bc* ≥ *bi*，∀*i* ∈ S。对于所有输掉拍卖的工人，*pi-ci*=0。因此，所有工人的*pi-ci*≥0，即DPDA是个体合理的。  然后，我们分析了DPDA的近似比。该想法是首先描述问题（14）的最优解，然而由于外部性，该问题仍然具有挑战性。为了解决这一难题，我们研究了问题（14）的结构，并在将问题（14）转化为等价问题后，发现了问题（14）的隐单调性。基于这一发现，我们证明了DPDA满足问题（12）的准确度要求，并利用DPDA输出与问题（12）最优解之间的关系推导出DPDA的近似比。该结果在下面的定理中进行了总结。    定理4的证明见附录。  **4性能评估**  **4.1 模拟装置**  在我们的模拟中，报价是从区间[1，20]均匀随机产生的，而权重首先是从区间[1，10]均匀随机产生，然后进行标准化。工人数量N从200到400不等。失真被由一些最大的失真Δmax规范化，以使W在不同的失真下总是正的。利用CPLEX优化求解器，基于对分算法来计算问题（13）的最优解[13]。据我们所知，由于没有允许工人在考虑外部性的同时报告噪声数据的移动群智感知拍卖机制，因此我们只分析第3部分中分析的DPDA算法的性能。 |

支付额与准确性 隐私与准确性 支付额与工人数量

标准化失真度=0.1

标准化失真度=0.2

标准化失真度=0.3



工人数量

标准化失真度=0.1

标准化失真度=0.2

标准化失真度=0.3

中标者与工人数量

中标者数量

总应付额

标准化失真度

隐私（max *ϵi*）

标准化失真度

总支付额

图2：不同准确度要求下的支付额 图3：数据隐私和准确性之间的关系 图4：外部性的影响

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **4.2 结果和讨论**  **支付额与准确性.** 图2中，我们说明了在不同工人总数和不同准确度要求下的支付额。我们观察到，随着失真度的增加，总支付额减少。这仅仅是因为W随着Δ的增加而减少，即平台不需要向工人购买太多隐私。同时，对于相同的失真度，总支付额随着工人数量的增加而增加。因为根据（13），对于相同的失真度，W随着工人数量的增加而增加。这就要求平台选择更多的工人，因此总支付额增加。  **隐私与准确性.** 图3中，我们阐述了数据隐私和准确性之间的关系。由于每个工人的隐私都是不同的，我们使用所有工人的最大*ϵi*（*ϵ* = max*i* ∈S *ϵi*）来表示在给定失真度下的隐私保护水平。正如我们所预期的，随着失真度的增加，隐私保护水平增加（*ϵ*越小，隐私保护水平越高），这与我们在第2.4节中的分析一致。  **外部性.** 图4中，我们说明了外部性所产生的影响。 如第3.1节所述，每个工人的数据隐私取决于其他工人的参与，工人数量发生的变化将改变每个工人的隐私。随着工人数量的增加，平台需要雇佣更多的工人来维持相同的失真度。因此，我们观察到总支付额和中标者人数的增加。此外，失真度越高，总支付额越低，中标者数量越少。  **近似比.** 在表1中，我们通过比较DPDA的总支付额和最优支付额，来说明所提出DPDA算法的性能。对于每个N，我们运行100个实验。在每个实验中，我们随机生成第4.1节中提到的参数。在不同的设置下，我们观察到，由DPDA算法产生的总支付额非常接近最优支付额，并且每种情况下的最大近似比都在2左右。 | **表1：DPDA算法的近似值，其中我们选择标准化失真等于0.2**   |  |  |  |  | | --- | --- | --- | --- | | 工人数量N | 200 | 300 | 400 | | 平均近似比值 | 1.88 | 1.85 | 1.85 | | 最小近似比值 | 1.71 | 1.69 | 1.70 | | 最大近似比值 | 2.15 | 2.07 | 1.99 |   **真实性。**图5中，我们验证了所提出的DPDA算法的真实性。我们在拍卖会上随机选出一个中标者和一个落选者。我们固定其他工人的报价，并操纵所选工人的报价来评估效用。图5展示了所择工人的效用如何随其报价而改变。我们可以看到，无论报价如何变化，中标者或落选者都无法提高自己的效用。并且，对一个工人来说，最好的报价战略是如实报价。  操纵中标者的报价    效用  临界报价  真实评估  报价  操纵落选者的报价    效用  真实评估  临界报价  报价  图5：DPDA算法的真实性  **计算复杂性.** 图6中，我们阐述了所提出的DPDA算法的计算复杂性。对于每个N，我们通过运行100个实验来检查算法的平均运行时间，其中参数是按照第4.1节所述随机生成的。这些实验是在一台拥有2.7 GHz英特尔酷睿i7处理器和16GB内存的电脑上进行的。在不同的设置（即，不同的失真度、报价和权重）下，我们观察到，所提出的DPDA算法的计算时间是短的，并且与问题大小近似线性。 |

|  |  |
| --- | --- |
| 工人数量  时间（秒）  标准化失真度=0.1  标准化失真度=0.2  标准化失真度=0.3  图6：不同设置下，DPDA算法的计算时间  **5结论与未来工作**  我们在一个拍卖框架下研究了移动群智感知中的隐私保护数据聚合，其中平台扮演拍卖者的角色来招募工人完成感知任务。 在该模型中，我们利用差异隐私概念设计了一个新的移动群智感知系统。具体来说，我们设计了一种数据聚合，它允许每个工人报告有噪声的数据，并且可以保证以差异隐私的方式使用每个工人的数据。然后，我们设计了一个真实的、个体合理的、计算效率高的激励机制。它可以找到一组工人，以根据聚合结果的准确度要求，尽可能地减少从工人那里购买隐私感知数据的成本。通过理论分析和大量模拟，我们验证了该方案的有效性。  我们知道，虽然这项工作通过允许工人自身进行数据干扰成功地解决了不可信平台的问题，但工人仍然无法完全控制其数据隐私。在这个模型中，只要满足个体的合理性要求（她的隐私损失得到了补偿），每个工人都愿意参与。每个工人的噪声参数由工人参与产生的外部性决定，因此由平台间接决定。实际上，如果只允许工人在数据中添加少量噪声，即使他们可以得到足够多的金钱补偿，他们也可能拒绝参与。  在未来的工作中，我们将采用另一模式来解决这一问题。这种模式赋予工人更多决定他们隐私保护水平的权力。一种可能的选择是拍卖模型，其中除了单位隐私成本报价，潜在的工人将竞标他们期望的隐私水平。在此基础上，平台为感知任务选择一组工人。我们需要评估工人报告隐私水平的战略性行为，并保证所设计机制的真实性。  **6 致谢**  这项工作部分得到了美国国家科学基金会CNS-1559696、IIA-1301726、CNS-1566634、ECCS-1711991、ECCS-1408409和SaTC-1618768的资助，部分得到了国家自然科学基金委员会61731004和工业控制技术国家重点实验室开放项目ICT1800373的资助。 | **附录：定理4的证明**  为了证明定理4，我们将首先描述问题（14）的最优分配xR∗，在此基础上我们可以建立DPDA输出与问题（12）最优解之间的关系。  首先，我们需要建立问题（14）和以下问题间的等式：    引理 3.*问题（14）的最优分配xR\*与问题（16）的最优分配相同。*  证明. 由于问题（14）和问题（16）具有相同的可行集，因此问题（14）的任何可行x必须对（16）可行。注意，C是问题（14）的最优值，即。因此，对于任何可行的x，我们都有。在一些代数中，我们有，在x*R*∗上保持相等。换言之，xR\*是问题（16）的最佳解决方案。因此，引理成立。  在引理3的基础上，为了描述最优分配xR∗，我们可以利用单调性来描述问题（16）的最优解，这是由以下引理得出的。  引理 4. *定义* ≤0, ∀*j* = 1, . . . , *N* }*.问题（16）的最优解xR\*如下：*    证明.首先，我们证明了l和*xiR*∗的存在性。定义 *p*（*k*） = ，其中*k* = 0, 1, . . . , *N*。如果*k* = 0，*p*（0） = ＜0； 如果*k* = *N*，*p*（*N* ） = > 0。注意*p*（*k*）严格地随k而增加。因此，存在*l* 使得 *p*（*l*） ≤ 0 和 *p*（*l* + 1） > 0。进而，定义*q*（*xl*+1） = *p*（*l*） + （*bl*+1*wl*+1 + *Cwl*+1）*xl*+1 ，且0 ≤ *xl*+1 ≤ 1,，其中 *q*（0） = *p*（*l*） ≤ 0 和 *q*（1） = *p*（*l* + 1） > 0。当 *q*（*xl*+1） 是连续且严格随 *xl*+1增加时，存在一个唯一的 *xlR*+∗1 ∈ [0, 1] ，以使*q*（*xlR*+∗1） = 0。接下来，为了验证（17）中的最优解xR∗，我们可以利用问题（16）的KKT条件。具体地说，问题（16）的Lagrangian方程是    其中，*λ、μ*和*ν*是Lagrangian乘数。 服从 （18），对于所有*xi，*我们都有*biwi* + *Cwi* - *λwi* + *μi* - *νi* = 0 。很容易验证引理4给出的xR\*i满足KKT条件，*λ μi*∗ = 1{*i* ≤*l* }（*bl*+1 - *bi*）*wi*，且*νi*∗ = 1{*i* >*l*+1}（*bi* - *bl*+1*wi*）。其中1条件表示指标函数，即条件成立时条件为1条件= 1，否则1条件= 0。因此，引理成立。  基于引理4，很容易证明DPDA选择了l+1个工人，并满足准确度要求。 |

|  |  |
| --- | --- |
| 引理 5.*设l为引理4中的定义，k为DPDA选择的中标者数。然后，我们得到k=l+1，DPDA满足准确度要求（即δ（x）≤Δ）。*  证明. 通过构建，DPDA将选择最小的k，以使≥0，即来自引理4的工人1，2，. . ，l+1。 此外，因为*xR*∗ *l*+1 ∈ [0, 1]， ≥ +wl+1*xlR*+∗1 ≥W，进而得出此证明。  现在，我们将基于引理1和引理3，证明DPDA相对于最佳总付款额为α近似值，这等效于表明DPDA相对于问题（16）的最佳值为α近似值。注意，（16）中的目标函数包含一个常数*i*，删除该常数不会影响问题（16）的优化结果。换句话说，我们可以把焦点放在函数*h*（x） = 中的近似值上。 设x*DPDA* 为 DPDA的分配 ，x∗ 为 OPT = *h*（x∗） 的最优分配。由于问题 （16） 是问题 （13）的松弛，因此 *h*（x*R*∗） ≤ OPT。接下来，我们将证明h（xdpa）≤*α*OPT。根据引理5，我们有    其中：（a）服从*h*（x*R*∗） ≥ （*bl*+1 + *C*）*wl*+1*x1lR*+∗。由于 *xR*∗ *l*+1 ≤ 1，我们有*α* ≥ 1。因此， *h*（x*DPDA*） ≤ *αh*（x*R*∗） ≤ *α*OPT，从而得出该证明。 |  |