

群智感知网络中数据质量保证方法分析

张志敏, 纪雷, 王心怡, 赵国生

哈尔滨师范大学计算机科学与信息工程学院, 黑龙江 哈尔滨

收稿日期: 2023年10月9日; 录用日期: 2023年11月29日; 发布日期: 2023年12月13日

摘要

移动智能终端设备和无线技术的快速发展, 使得移动群智感知网络作为一种全新的物联网感知范式得到了广泛关注。然而在实际应用中, 移动群智感知系统的大规模应用仍面临一些挑战。感知数据作为移动群智感知网络的重要组成部分, 其质量的高低直接影响着移动群智感知系统的服务质量, 进而对移动群智感知系统能否大规模普及起着关键作用, 因此确保高质量数据是移动群智感知网络研究的重点之一。本文首先从感知数据生命周期的角度出发, 对数据感知、数据上传、数据交易阶段中影响感知数据质量的因素进行了详细说明。其次, 从激励机制、数据可靠性评估、隐私保护三个方面对现有的感知数据质量保证方法进行了总结; 最后, 对移动群智感知网络的应用前景以及未来的研究方向进行了总结与展望, 以期移动群智感知数据质量保证方法研究提供一些参考。

关键词

移动群智感知网络, 数据质量, 隐私保护

Analysis of Data Quality Assurance Methods in Mobile Crowdsensing Networks

Zhimin Zhang, Lei Ji, Xinyi Wang, Guosheng Zhao

College of Computer Science and Information Engineering, Harbin Normal University, Harbin Heilongjiang

Received: Oct. 9th, 2023; accepted: Nov. 29th, 2023; published: Dec. 13th, 2023

Abstract

The rapid development of mobile intelligent terminal devices and wireless technology has made mobile crowdsensing networks as a new paradigm of IoT attract widespread attention. However, in practical applications, the large-scale application of mobile crowdsensing systems still faces some challenges. As an important component of mobile crowdsensing networks, the quality of crowdsensing data directly affects the service quality of mobile crowdsensing systems, and thus

plays a crucial role in the large-scale popularization of mobile crowdsensing systems. Therefore, ensuring high-quality data is one of the focuses of research on mobile crowdsensing networks. This article first provides a detailed explanation of the factors that affect the quality of crowdsensing data during the stages of data perception, data upload, and data trading from the perspective of the crowdsensing data lifecycle. Secondly, summarizing the existing methods for ensuring perceived data quality from three aspects: incentive mechanism, data reliability evaluation, and privacy protection. Finally, a summary and outlook were made on the application prospects and future research directions of mobile crowdsensing networks in order to provide some reference for the research on data quality assurance methods for mobile crowdsensing.

Keywords

Mobile Crowdsensing Networks, Data Quality, Privacy Preservation

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

移动群智感知源于众包(crowdsensing)思想, 2006年杂志主编 Howe 在美国《连线》中提出了这一概念。与传统的传感器网络相比, 移动群智感知网络中以携带感知设备的人作为感知节点, 具有数量大、类型多、范围广的优势。近年来, 随着集成多种传感器的移动智能终端设备的快速发展, 移动群智感知网络作为一种全新的物联网感知范式已经成为了研究热点之一。移动群智感知网络以大量普通用户作为感知源, 可以完成那些仅依靠个体很难实现的大规模、复杂的社会感知任务, 进而为城市及社会管理提供智能辅助支持。现已经应用于医疗、教育、社区服务等领域。

移动群智感知系统的服务质量是决定移动群智感知网络能否大规模应用的关键因素。感知数据是移动群智感知网络的基础, 感知平台正是依托大量移动终端用户贡献的感知数据, 才能对数据进行分析 and 处理, 提取出有效信息, 进而为服务请求者提供服务。移动群智感知系统服务质量与感知数据质量的高低息息相关。然而在实际应用中, 一方面, 受用户不专业行为、设备差异、环境噪音等因素影响, 感知数据的质量往往是参差不齐的。另一方面, 潜在的隐私泄露风险使得感知用户选择上传虚假数据甚至拒绝参与感知任务。这些低质量数据的存在, 使得感知平台后续数据分析的准确性和有效性大大降低, 不仅浪费了感知用户和平台的资源, 更是损害了群智感知系统的服务质量[1]。因此感知数据质量保证是一个完备、成熟的移动群智感知系统必须面对的问题。本文首先对移动群智感知网络中影响数据质量的因素进行了概述与分析。其次, 对现有的移动群智感知数据质量保证方法进行归纳总结。最后, 对移动群智感知网络的未来研究方向进行了探讨。

2. 移动群智感知网络中的数据质量问题

传感器网络的质量概念通常被定义为网络中传统的服务质量的衍生物, 一般根据带宽、延迟、数据包等服务质量指标来进行评估。由于群智感知系统的独特性, 其质量问题中也引入了新的范畴。为了便于理解群智感知中的质量问题, 本文首先展示群智感知系统的典型框架。典型的移动群智感知系统主要由感知用户、感知平台和服务请求者三部分组成, 如图 1 所示。感知平台负责管理群智感知的整个工作流程, 感知用户在接受感知任务后, 利用自身的移动智能终端和移动传感设备来收集感知数据, 并上传

至感知平台以获得相应奖励；感知平台会对收到的数据进行分析与处理，进一步与服务请求者之间进行数据交易。从感知数据角度来看，感知数据的生命周期可以大致分为三个阶段：数据感知、数据上传和数据交易[2]，在各阶段均存在影响感知数据质量的问题。

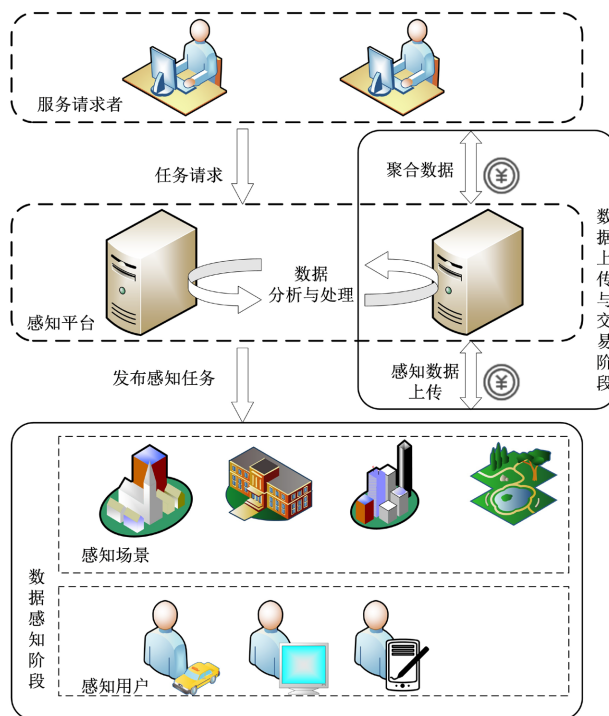


Figure 1. The architecture of mobile crowdsensing network
图 1. 移动群智感知网络架构

在数据感知阶段，感知用户在确定参与感知任务之后，需要利用其携带的移动智能终端设备和内置的传感设备，结合感知任务的具体场景来采集对应的感知数据。但由于用户不专业的行为、设备故障、环境噪声等因素的存在，不同用户对同一感知任务收集到的感知数据往往是参差不齐的。例如，在监测道路噪声时，处在室内的用户与处在室外的用户相比，其采集的数据与真实值相比，会存在较大误差，即用户采集数据的质量会受到参与用户的主观行为影响。此外，设备故障、传感器老化等客观因素也会使收集到的感知数据质量受到影响。因此，在这一阶段不仅需要提高用户提供高质量数据的意愿，同时需要感知平台具备可靠的数据聚合机制来降低低质量感知数据对最终结果的消极影响。

数据上传阶段，发生在感知用户与感知平台之间、感知平台与服务请求者之间。在此阶段，数据的隐私安全性是首要考虑的问题。用户采集到的感知数据中，往往包含一些位置、身份等敏感信息，敏感信息的泄露，对用户带来了极大安全威胁。例如，攻击者通过获取用户上传的医疗相关感知数据来推测用户的健康状况，从而进行一系列恶意攻击[3]。首先，感知数据质量也依赖于数据的数量，而对隐私泄露的担忧会降低用户参与感知任务的意愿，导致参与用户数量不足；其次，用户出于对自身数据隐私保护的考虑，可能会对自身数据进行匿名化、加噪等处理，这虽然一定程度上减少了隐私泄露的概率，但也降低了数据的可用性；最后，用户作为追求自身利益最大化的理性实体，系统中难免存在为了获取报酬而上传虚假数据的恶意用户。因此，一个完备的移动群智感知机制不仅需要提供有效的隐私保护机制，降低用户对隐私泄露的担忧，同时，需要设计相应的恶意用户识别机制，选择可靠的用户参与感知任务。

感知用户与感知平台、感知平台与服务请求者之间都存在数据交易。在数据交易阶段，感知用户有

可能会在明知自身数据不可靠的前提下,为了从感知平台获得奖励而上传该虚假数据。感知平台有必要设计相应的数据可靠性评估机制来保证数据质量。此外,感知平台和服务请求者都是半可信实体,有可能会向敌手提供与感知用户相关的背景知识,帮助敌手挖掘用户的隐私信息,因此在数据交易阶段同样要考虑隐私泄露问题对用户参与感知任务积极性的影响。

3. 数据质量保证技术研究

3.1. 基于数据质量的激励机制

激励被认为是鼓励感知用户参与感知任务的有效手段。在移动群智感知系统中,激励机制可以用如下的模型[4]表示:

$$I : M \rightarrow \max(B(P), B(U)) \quad (1)$$

该模型表示群智感知激励机制 I (Incentive),通过某种激励方式 M (Mechanism),使得感知平台 P (Crowdsensing Platform)和参与用户 U (Users)获得的收益 B (Bonus)最大。

在实际应用中,感知用户的行为具有不确定性,这意味着感知用户即使具有参与感知任务的意愿,也会存在提交低质量数据的可能。此外,系统中不可避免存在恶意用户,为了获得额外收益或是破坏群智感知系统服务质量,恶意用户会上传虚假数据。基于数据质量的激励机制以用户提交的感知数据质量作为奖励分发的依据,可以有效解决上述问题。Zhao 等人[5]提出了一种支持数据质量感知的隐私保护群智感知激励机制,该机制首先设计了一种基于零知识证明的数据质量评估模型,从位置、时间以及数据自身范围三个维度对用户的提供的感知数据进行异常数据滤掉;随后根据数据与真实值之间的偏差来对数据质量进行量化,最后以数据量化结果向参与用户分配不同的金钱奖励,但是该机制需要预设数据的可靠性要求。Peng 等人[6]提出的数据质量感知激励机制,采用期望最大化算法来评估每个参与用户的努力程度以计算感知数据的质量,在此基础上,运用信息论的知识来量化数据的有效贡献,最后根据贡献给与参与用户相应的报酬。但是该机制只针对数值型数据,对于其他类型的数据难以度量质量。应等人[7]从解决不可信第三方问题出发,设计了一种区块链群智感知激励机制,通过真值发现算法计算用户权重,并结合用户的竞价来确定用户应得的报酬。该方案在真值估计模块和激励模块采用 CKKS 同态加密方法来保护隐私,但也面临计算开销较大的问题。综上所述,基于数据质量的激励机制中,最关键的是要设计公平合理的多模态数量质量量化算法,既要避免提供低质量数据的用户获得高报酬,也要保证提供高质量数据的用户获得高报酬。

3.2. 数据可靠性评估

数据可靠性评估也是解决移动群智感知数据质量问题的重要方法。参与用户不专业的行为、感知设备异构性和环境噪声等因素对移动群智感知服务可靠性的影响,最终都会在用户上传的感知数据中体现出来,因此设计合理的数据可靠性评估模型对移动群智感知系统至关重要。在现有研究中,通常由感知平台承担数据可靠性评估的任务。Zhao 等人[8]提出一种面向异构数据的可靠性评估方案,可在保护参与用户隐私的同时,支持对数值数据、图片数据和文本数据的可靠性评估。An 等人[9]采用区块链技术和模糊理论来实现去中心化的移动群智感知系统,并完成数据可靠性评估。此外,真值发现作为从多个用户提供的感知数据中发现真值的有效方法,也被广泛应用于数据可靠性评估之中。真值发现算法的基本思想是为每个数据源即感知用户赋予不同的可靠性,基于数据源的可靠性程度对数据进行加权聚合,以得到最符合真实值的数据。目前,研究者们已经提出了多种真值发现方案,如 TrustFinder、LTM、CATD、CRH 等。而考虑到移动群智感知过程中的数据隐私问题,融合隐私保护技术的真值发现方案显然更适用

于移动群智感知系统。Xue 等人[10]提出了一个名为 InPPTD 的隐私保护真值发现方案, 该方案采用同态加密技术解决真值发现过程中的隐私泄露问题, 但同态加密技术带来了一定的计算开销。徐等人[11]将真值发现与激励机制结合, 使用对称加密技术实现隐私保护真值发现, 并基于真值发现模型对数据进行可靠性评估以计算奖励, 但该方案不适于存在异构数据的感知任务。综上所述, 移动群智感知网络中的数据可靠性评估不仅要考虑数据隐私问题, 保证数据评估过程中用户的隐私信息不被泄露, 同时考虑到不同类型数据的评估标准并不完全相同, 数据可靠性评估机制还要满足异构数据的评估需求和评估准确性。

3.3. 隐私保护技术

合理的隐私保护技术, 一方面可以提高感知数据上传过程中的安全性, 避免攻击者对数据的截获、篡改带来的感知数据质量低下问题。另一方面, 也可以降低感知用户对于自身数据隐私泄露的担忧, 提高其参与感知任务积极性。因此对于移动群智感知中参与用户位置、身份、数据的隐私保护研究也是移动群智感知领域中的研究热点。现有的移动群智感知隐私保护技术主要分为三类: 密码学、匿名化、数据扰动。高等人[12]针对现有方案中依赖可信第三方的问题, 基于区块链技术, 提出了一种群智感知位置隐私保护参与者选择方法, 采用保序加密和 Merkle 树实现参与用户的个性化位置隐私保护, 但还需要进一步提高效率并对影响参与者选择的评估指标进行更细粒度的分析。金等人[13]提出了一种基于移动节点的数据隐私保护算法 DPPMN, 构建数据传输的匿名路径, 并由源节点使用服务器与路径节点的公钥依次对感知数据进行加密, 保证数据隐私的同时可以低于共谋攻击。李等人[14]在引入边缘计算的移动群智感知场景下, 采用本地化差分隐私技术来实现对数据的隐私保护, 基于属性关系构建用户数据的隐私模型和数据可用性模型, 针对属性联合和属性独立两种情况提出了不同的本地化差分隐私方案。但该方案只考虑了数据提交阶段的隐私保护困难。综上所述, 群智感知领域中的隐私保护大多是采用已有的较为成熟的隐私保护技术, 从实用性的角度出发, 隐私保护的研究应考虑与移动群智感知网络中的任务分配、参与者选择、数据聚合等具体场景相结合。

4. 移动群智感知网络的未来研究方向

物联网的根本发展需求是对物理世界进行更加全面和透彻的感知。传统的无线传感器网络, 更适用于小范围内的感知任务, 难以适应大规模的城市感知, 存在组网成本高、系统维护难、服务不灵活等问题。移动群智感知网络具有“以人为中心”的特点, 它以现实世界中的人作为数据源, 利用人们携带的内置丰富传感器的移动终端设备(如手机、可穿戴设备、车载传感器等)来完成感知数据的采集工作, 是一种全新的物联网感知范式, 在智能交通、环境监测、健康医疗等大规模社会感知任务上具有广阔的发展前景。移动群智感知为解决复杂多样的感知问题提供了新方法, 但也面临着诸多挑战, 需要进一步研究。未来的移动群智感知研究工作可从以下几个角度展开。

(1) 个性化的群智感知隐私保护

在移动群智感知系统中, 用户对于自身感知数据的隐私需求并不相同, 集中式隐私保护对所有数据采用同一种隐私保护方式或分配同样的隐私预算, 存在隐私保护开销大、数据可用性降低的问题。个性化的隐私保护是从用户属性以及数据可用性的角度出发, 对于具有不同强度隐私保护需求的数据分配不同的隐私预算或采用不同的隐私保护方法。需要设计合理的隐私保护水平度量算法, 为具有不同隐私保护需求参与者匹配合适的隐私保护强度, 降低隐私开销的同时减少数据失真, 达到隐私保护与数据可用性之间的平衡。

(2) 更完备的数据可靠性评估机制

数据可靠性评估设计要考虑群智感知的实际应用场景, 如任务分配、数据聚合、激励分发等。首先,

感知任务是不断变化的, 如何在不需要预先设定可靠性要求的前提下, 实现实时的针对不同感知任务的数据可靠性评估值得我们更多思考。其次, 不同的感知任务需要的数据类型存在差异, 只针对数值型数据的评估方案难以适用于当下的移动群智感知应用, 同时, 数值型数据与图片、音频等数据类型的评估标准也不完全一样, 因此, 需要设计面向异构数据的多元数据可靠性评估方法, 确保数据评估的完整性和可靠性。最后, 在数据可靠性评估的过程中, 直接对明文数据进行可靠性评估不可避免地存在隐私泄露的风险, 攻击者截获、篡改数据不仅会损害群智感知系统的服务质量, 也会给感知用户带来安全风险, 因此设计具有隐私保护的数据可靠性评估机制具有重要研究意义。

(3) 去中心化的群智感知系统

在移动群智感知网络中, 感知平台往往充当群智感知系统的中心角色, 而感知平台实际是一个半可信实体。当前的移动群智感知任务分配、数据聚合、激励机制等场景大多将感知平台看作可信第三方, 没有考虑感知平台不可信的问题。因此, 实现去中心化的移动群智感知系统极具意义。区块链技术的去中心化特性, 可以有效解决群智感知系统中的第三方依赖问题, 同时区块信息的不易篡改性和可追溯性也为数据安全提供了保障。因此, 将区块链技术应用用于群智感知系统也是一个有价值的研究点。

5. 结语

移动群智感知网络作为一种新型物联网感知范式, 具有广阔的应用前景。但是移动群智感知网络的大规模推广仍面临一定挑战, 感知数据作为移动群智感知的重要组成部分, 是影响移动群智感知广泛普及的关键因素。本文首先对移动群智感知网络的概念及基本架构做了详细介绍, 将感知数据的生命周期划分为数据感知、数据上传、数据交易三个阶段, 在此基础上, 对不同阶段中影响感知数据质量的因素进行了阐述。随后, 从激励机制、数据可靠性评估、隐私保护技术三个方面总结了现有的感知数据质量保护技术的研究现状和存在问题。最后, 对移动群智感知网络的应用前景和未来研究方向进行了探讨。

基金项目

黑龙江省高等教育教学改革研究一般项目(SJGY20220351)和 2023 年度省规划办重点课题(GJB1423438)。

参考文献

- [1] 杨朔. 移动群智感知中的数据质量分析和基于质量的用户管理[D]: [硕士学位论文]. 上海: 上海交通大学, 2019.
- [2] 熊金波, 毕仁万, 田有亮, 等. 移动群智感知安全与隐私: 模型, 进展与趋势[J]. 计算机学报, 2021, 44(9): 1949-1966.
- [3] 王涛春, 金鑫, 吕成梅, 等. 移动群智感知中融合数据的隐私保护方法[J]. 计算机研究与发展, 2020, 57(11): 2337-2347.
- [4] 吴垚, 曾菊儒, 彭辉, 等. 群智感知激励机制研究综述[J]. 软件学报, 2016, 27(8): 2025-2047.
- [5] Zhao, B., Tang, S., Liu, X., et al. (2020) PACE: Privacy-Preserving and Quality-Aware Incentive Mechanism for Mobile Crowdsensing. *IEEE Transactions on Mobile Computing*, **20**, 1924-1939. <https://doi.org/10.1109/TMC.2020.2973980>
- [6] Peng, D., Wu, F. and Chen, G. (2017) Data Quality Guided Incentive Mechanism Design for Crowdsensing. *IEEE Transactions on Mobile Computing*, **17**, 307-319. <https://doi.org/10.1109/TMC.2017.2714668>
- [7] 应臣浩, 夏福源, 李颀, 等. 区块链群智感知中基于隐私数据真值估计的激励机制[J]. 计算机研究与发展, 2022, 59(10): 2212-2232.
- [8] Zhao, B., Tang, S., Liu, X., et al. (2020) IronM: Privacy-Preserving Reliability Estimation of Heterogeneous Data for Mobile Crowdsensing. *IEEE Internet of Things Journal*, **7**, 5159-5 170. <https://doi.org/10.1109/JIOT.2020.2975546>
- [9] An, J., Liang, D., Gui, X., et al. (2019) Crowdsensing Quality Control and Grading Evaluation Based on a Two-Consensus

-
- Blockchain. *IEEE Internet of Things Journal*, **6**, 4711-4718. <https://doi.org/10.1109/JIOT.2018.2883835>
- [10] Xue, K., Zhu, B., Yang, Q., *et al.* (2020) InPPTD: A Lightweight Incentive-Based Privacy-Preserving Truth Discovery for Crowdsensing Systems. *IEEE Internet of Things Journal*, **8**, 4305-4316. <https://doi.org/10.1109/JIOT.2020.3029294>
- [11] 徐苗苗, 陈珍萍. 基于对称加密和双层真值发现的连续群智感知激励机制[J]. 计算机科学, 2023, 50(1): 294-301.
- [12] 高胜, 陈秀华, 朱建明, 等. 基于区块链的群智感知中任务预算约束的位置隐私保护参与者选择方法[J]. 计算机学报, 2022, 45(5): 1052-1067.
- [13] 金鑫, 王涛春, 吕成梅, 等. 移动群智感知中原始数据隐私保护算法[J]. 计算机应用, 2020, 40(11): 3249-3254.
- [14] 李卓, 宋子晖, 沈鑫, 等. 边缘计算支持下的移动群智感知本地差分隐私保护机制[J]. 计算机应用, 2021, 41(9): 2678-2686.