

一种基于局部差分隐私的数据流高效在线直方图发布算法

陶陶^{1,2,3}, 张福南^{1,2,3}, 王修君^{1,2}, 郑啸^{1,2}, 赵欣⁴

¹安徽工业大学计算机科学与技术学院, 中国马鞍山市, 243032

²安徽省工业互联网智能应用与安全工程实验室, 中国马鞍山市, 243032

³安徽工业大学工程研究院, 中国马鞍山市, 243032

⁴东营市胜利第一中学, 中国东营市, 257000

摘要: 目前各领域都在产生包含用户敏感信息的实时数据流。尽管这些数据的共享和发布具有巨大商业价值, 但如果直接发布数据, 将会泄露数据中的用户隐私信息。因此, 如何基于滑动数据流窗口持续生成满足隐私保护要求的可发布直方图已成为一个关键问题, 尤其是在将数据发送给不受信任的第三方时。现有直方图发布方法在时间和存储成本方面的表现并不令人满意, 因为它们必须缓存当前滑动窗口 (SW) 中的所有元素。为解决这一问题, 我们为本地差分隐私数据流提出一种高效的在线直方图发布算法 (EOHP)。具体来说, 在 EOHP 算法中, 数据收集器首先使用数据流的近似计数方法实现在线处理数据获得初步直方图。其次, 提出了优化隐私预算分配策略减少隐私预算的消耗, 在近似直方图中添加适当噪声, 使其在保持较好数据可用性的同时发布直方图。经两个不同真实数据集上的大量实验结果表明, 与其他现有算法相比, EOHP 算法显著降低了时间和存储成本, 提高数据实用性。

关键词: 数据流; 差分隐私; 滑动窗口; 近似计数

<https://doi.org/10.1631/FITEE.2300368>