Adaptive Personalized Randomized Response Method Based on Local Differential Privacy

Dongyan Zhang, Henan University of Science and Technology, China

Lili Zhang, Henan University of Science and Technology, China*

Zhiyong Zhang, Henan University of Science and Technology, China

 https://orcid.org/0000-0003-3061-7768

Zhongya Zhang, Henan University of Science and Technology, China

ABSTRACT

Aiming at the problem of adopting the same level of privacy protection for sensitive data in the process of data collection and ignoring the difference in privacy protection requirements, the authors propose an adaptive personalized randomized response method based on local differential privacy (LDP-APRR). LDP-APRR determines the sensitive level through the user scoring strategy, introduces the concept of sensitive weights for adaptive allocation of privacy budget, and realizes the personalized privacy protection of sensitive attributes and attribute values. To verify the distorted data availability, LDP-APRR is applied to frequent items mining scenarios and compared with mining associations with secrecy konstraints (MASK), and grouping-based randomization for privacy-preserving frequent pattern mining (GR-PPFM). Results show that the LDP-APRR achieves personalized protection of sensitive attributes and attribute values with user participation, and the maxPrivacy and avgPrivacy are improved by 1.2% and 4.3%, respectively, while the availability of distorted data is guaranteed.

KEYWORDS

Frequent items mining, Local differential privacy, Personalized privacy, Privacy budget, Randomized response

1. INTRODUCTION

In recent years, big data technologies have emerged in the scenario of booming network and information technology. People can discover the laws and knowledge hidden in the data from the huge amount of data through data mining algorithms, which is important for industrial development, social services, and many other fields (Chen et al., 2023). If the data is directly provided to a third party, it will lead to the leakage of personal privacy information, which will bring a great threat to personal safety and property security. In addition, if the data miner cannot provide sufficient privacy protection, it will lead to some users refusing to provide data due to a lack of trust, and the data miner will not be able

to mine more accurate information due to a lack of data. Therefore, it is necessary to design a secure privacy protection scheme.

There are more existing privacy protection schemes, such as data anonymization, data encryption, data perturbation, etc. (Cormode et al., 2021), Differential Privacy is a privacy protection technique based on data perturbation, which has become a hot research topic today due to its rigorous mathematical proofs and possession of quantitative privacy protection capabilities (Zhang et el., 2023; Duan et al., 2022; Ren et al., 2022; Qian et al., 2022). In the big data environment, to prevent privacy attacks by untrustworthy third parties and attackers with arbitrary background knowledge, sensitive information needs to be more comprehensively protected, and the Local Differential Privacy (LDP) (Duchi et al., 2013) technique has emerged. Locality refers to the random perturbation of user data before it leaves a smart device, such as a cell phone, and subsequently sent to a third-party data collector, i.e., the data collector only gets a part of the true data, and the data still retains a certain utility. Since it was formally proposed in 2013, LDP technology has been greatly developed and improved, and widely deployed in practical applications, such as Microsoft, Google, Apple, other companies have embedded LDP in their applications (Arcolezi et al., 2023).

With the frequent occurrence of privacy leakage incidents, users' awareness of privacy protection is increasing, and the demand for personalized privacy protection is also growing, for which scholars have proposed many personalized privacy protection methods (Niu et el., 2021; Ma et al., 2022; Li et al., 2022; Qian et al., 2022). Among the existing solutions, GR-PPFM (Guo et al., 2021) is more relevant as it can guarantee the availability of perturbed data while providing personalized privacy protection for users. However, it ignores the fact that there are also different privacy protection needs between the user's data attributes and attribute values. For example, home address requires a higher level of privacy protection compared to gender, and infectious disease (HIV) requires a higher level of privacy protection compared to common class of diseases (flu, fever), so GR-PPFM has some limitations. In order to solve the problem of adopting the same level of privacy protection for sensitive data in the process of data collection, ignoring the fact that different users have personalized privacy protection needs for data security and usability, as well as personalized differences in the attributes and attribute values of the data itself, this paper designs a personalized random response algorithm based on local differential privacy, which determines the sensitive level of user data by a scoring strategy, introduces the concept of sensitive weight for adaptive allocation of privacy budget, realizes the personalized privacy protection of sensitive attributes and attribute values, and ensures the availability of data while meeting the user's personalized needs. The main contributions of this paper are as follows:

1. In order to solve the problem that the existing personalized privacy protection schemes only divide the privacy level based on experience and lacks the user's personalized needs, this paper designs a user scoring strategy to reasonably set the privacy level of the sensitive data based on the user scoring; in order to satisfy the different privacy needs of different users for different sensitive attributes and attribute values, this paper introduces the concept of sensitive weighting and puts forward the method of personalized privacy budget allocation, which allocates a reasonable privacy budget for different attributes and attribute values to satisfy the users' personalized needs.
2. In order to check the availability of distorted data, the proposed algorithm is applied to the frequent items mining scenario, and aiming at the problem of low accuracy of direct mining of distorted data in this scenario, the support reconstruction method is proposed, which theoretically deduces the estimation process of the true support by establishing the mathematical relationship between distorted data and true data in order to improve the accuracy of mining.
3. In order to evaluate the privacy of the proposed algorithm, it is strictly proved that the algorithm satisfies LDP theoretically. To verify data availability, the LDP-APRR method is tested on real

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/adaptive-personalized-randomized-response-method-based-on-local-differential-privacy/335225

Related Content

Performance Evaluation of Web Server's Request Queue against AL-DDoS Attacks in NS-2
Manish Kumarand Abhinav Bhandari (2017). *International Journal of Information Security and Privacy (pp. 29-46).*
www.irma-international.org/article/performance-evaluation-of-web-servers-request-queue-against-al-ddos-attacks-in-ns-2/187075

Risk and Security of Information Systems in the Portuguese Financial Sector: Model and Proof of Concept in Portuguese Regulator
Pedro Fernandes da Anunciaçãoand Alexandre Miguel Barão Rodrigues (2019). *International Journal of Risk and Contingency Management (pp. 18-38).*
www.irma-international.org/article/risk-and-security-of-information-systems-in-the-portuguese-financial-sector/234432

Exploring Secure Computing for the Internet of Things, Internet of Everything, Web of Things, and Hyperconnectivity
Maurice Dawson (2017). *Security Solutions for Hyperconnectivity and the Internet of Things (pp. 1-12).*
www.irma-international.org/chapter/exploring-secure-computing-for-the-internet-of-things-internet-of-everything-web-of-things-and-hyperconnectivity/164690

Security Flaws and Design Issues in Cloud Infrastructure
 (2019). *Detection and Mitigation of Insider Attacks in a Cloud Infrastructure: Emerging Research and Opportunities (pp. 52-61).*
www.irma-international.org/chapter/security-flaws-and-design-issues-in-cloud-infrastructure/221682

Privacy Preserving Fuzzy Association Rule Mining in Data Clusters Using Particle Swarm Optimization

Sathiyapriya Krishnamoorthy, G. Sudha Sadasivam, M. Rajalakshmi, K. Kowsalyaaand M. Dhivya (2021). *Research Anthology on Privatizing and Securing Data (pp. 1096-1116).*