Chapter 45
Privacy Preserving Data Mining Using Time Series Data Aggregation

Sivaranjani Reddi
ANITS, Bheemunipatnam, India

ABSTRACT

This article proposes a mechanism to provide privacy to mined results by assuming that the data is distributed across many nodes. The first objective includes mining the query results by the node in a cluster, communicating it to the cluster head, aggregating the data collected from all the cluster nodes and then communicating it to the group controller. The second objective is to incorporate privacy at each level of the clusters node: cluster head and the group controller level. The final objective is to provide a dynamic network feature, where the nodes can join or leave the distributed network without disturbing the network functionality. The proposed algorithm was implemented and validated in Java for its performance in terms of communication costs computational complexity.

INTRODUCTION

Many real life applications of data mining is facing problems towards the privacy preservation of the data (Anderson, 2010; Acs & Castelluccia, 2011; Dansana, 2012; van Dijk et al., 2010; Chowdhuri, 2014; Sarkar et al., 2017). It includes, firstly, certain attributes of the data or attributes that might leak the personal recognizable information. Secondly, the data can be split across multiple nodes either horizontally or vertically, and may not allow the data transfer to another side. Finally, usage of data model might have restriction on rules, and few rules may lead to law violation in order to access individual profiling. Privacy preserving based data mining (PPDM) (Agrawal, 1994) has arisen to discuss the above-mentioned issues. Majority of the PPDM techniques are the modified versions of the standard data mining algorithms, where the modification includes the cryptographic mechanisms which guarantee the privacy for the application. In many cases, restraints PPDM are: preserving data accuracy and retaining

DOI: 10.4018/978-1-7998-8954-0.ch045

Copyright © 2021, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

mining process performance while maintaining the privacy restrictions. Copious methodologies used by PPDM can be summarized based on following dimensions:

- **Data Distribution:** This dimension concentrates on data distribution. The approaches adopt either centralized data distribution or decentralized data distribution. Generally, the data distribution can be categorized as horizontal and vertical data distribution. While horizontal data splitting is discussed in detail in the forthcoming sections, vertical distribution distributes all values for different attributes in different places.
- **Data Alteration:** It is used to change the actual data into other form before releasing to the public in order to accomplish the data privacy. Data modification mechanisms include perturbation, blocking, aggregation (Chen et al., 2014; Won et al., 2014), swapping and sampling.
- **Privacy Preservation:** Assures the delivery of data to the intended data mine by adapting data alteration before delivering. Distribution of data is done among more than one node without revealing the data at individual site. In classification phase, where the results will be given to designate node, which does the classification, it checks for the occurrence of certain rules without disclosing them.

Many authors have proposed techniques in order to provide the confidentiality in data mining (Aggarwal, 2010; Oliveira, 2004; Rawat, 2015; Fouad & Hassan, 2016), Elaine Shi et al. (2011) has proposed a time series based aggregation mechanism in order to attain data privacy, where group participants can occasionally upload encrypted data to the group aggregator(GA), who is responsible to do the summation on data in every time periodically. The authors suggested a mechanism which allows group users to submit encoded values to data aggregator, Afterwards aggregator will perform the summation on participants' values in every period, without prior knowledge. We achieve strong privacy using this technique.

Rongxing Lu et al. (2012) has advised an efficient privacy preserving aggregation method based on homomorphic Paillier cryptosystem technique, uses a super increasing sequence to structure multi-dimensional data. The Paillier Cryptosystem can achieve homomorphic properties, widely desirable in many privacy-preserving applications (Sang et al., 2009; Zhong, 2007). Concretely, it is comprised of three algorithms: key generation, encryption and decryption. RSA modulus is used in key generation, to produce both publickey pk = (n,g) and privatekey sk = (λ,μ). Encryption phase is used to convert the original message m into ciphertext c=E(m)= g^m,r^n mod n^2. Decryption algorithm uses cipher text in order to recover the plaintext m = D (c) = $L(c^{\lambda \bmod n^2}).\mu \bmod n.$ As this cryptosystem is provably secure against chosen plain text attack, hence privacy (Kantarcıoglu, 2004; Surekha et al., 2010, 2011, 2012, 2013, 2017; Dey et al., 2016, 2017; Rajeswari, 2017; Tyagi, 2017; Jauvart, 2016; Rao, 2016; Reddi, 2017; Dharavath et al., 2016) is achieved. Compared against traditional 1-Dimentional data aggregation methods, confirmed that proposed can significantly reduce computational cost and improve communication efficiency, satisfying the real-time high-frequency data collection requirements in smart grid communications. The authors also provided security analysis to demonstrating its security strength and performance analysis shows efficiency improvement.

Rastogi and Nath (2010) first applies differentially aggregation for distributed time-series data, offers good practical utility even without of trusted server. It has addressed two challenges in datamining applications, (i) users can publish temporally correlated time-series data such as location traces, web history and health data. (ii) An untrusted aggregator allows to run aggregate queries on the data. As in

14 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/chapter/privacy-preserving-data-mining-using-time-series-data-aggregation/280213

Related Content

Social Aid Fraud Detection System and Poverty Map Model Suggestion Based on Data Mining for Social Risk Mitigation
Ali Serhan Koyuncugiland Nermin Ozgulbas (2011). *Surveillance Technologies and Early Warning Systems: Data Mining Applications for Risk Detection (pp. 173-193).*
www.irma-international.org/chapter/social-aid-fraud-detection-system/46810

Goals and Practices in Maintaining Information Systems Security
Zippy Erlichand Moshe Zviran (2012). *Optimizing Information Security and Advancing Privacy Assurance: New Technologies (pp. 214-224).*
www.irma-international.org/chapter/goals-practices-maintaining-information-systems/62724

Mitigation of Juvenile Delinquency Risk Through a Person-Centered Approach: The Intervention of Juvenile Probation Services
Christina Antonia Moutsopoulouand Afroditi Mallouchou (2018). *International Journal of Risk and Contingency Management (pp. 73-83).*
www.irma-international.org/article/mitigation-of-juvenile-delinquency-risk-through-a-person-centered-approach/205634

Hazmat Transport Safety and Alternative Transport Modes: A Study of US Accidents between 1990 and 2010
Luca Zamparini, Genserik Reniersand Michael Ziolkowski (2017). *International Journal of Risk and Contingency Management (pp. 1-17).*
www.irma-international.org/article/hazmat-transport-safety-and-alternative-transport-modes/177837

A Chronicle of a Journey: An E-Mail Bounce Back System
Alex Kosachevand Hamid R. Nemati (2009). *International Journal of Information Security and Privacy (pp. 10-41).*
www.irma-international.org/article/chronicle-journey-mail-bounce-back/34056