ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 9(38), DOI: 10.17485/ijst/2016/v9i38/101593, October 2016* ISSN (Online) : 0974-5645

Frequent Pattern Technique using Federation Rule Mining

T. Nusrat Jabeen[1] and M. Chidambaram[2]

[1]Department of Computer Science, Bharathiyar University, Coimbatore, Tamil Nadu, India; nusratjabeent@gmail.com
[2]Computer Science Department, Rajah Serfoji Government College, Thanjavur, Tamil Nadu, India; childsuba@gmail.com

Abstract

Objectives: To improve the security, as well as privacy while sharing data/information to third parties. From the database Duplicate data were eliminated and extracted the original database **Methods/ Statistical Analysis:** FP tree based algorithm was proposed in this paper. It is used to generate the frequent item data sets. Those frequent data Item sets are extracted by using inverse data item set. It must achieve good security and privacy. **Findings:** The main problem in existing system is information leakage. In frequent pattern technique, federation rule mining process which tries to find some correlations and associations among the various types of data items in a dataset. It finds more privacy preserving techniques related to the data mining process. **Applications/Improvement:** To compare and evaluate the proposed of many algorithms, federation rule mining should able to maintain the data privacy in a proper manner.

Keywords: Federation Rule Mining, Frequent Pattern, Crypt Analysis, Data Contortion, Data Sensation

1. Introduction

1.1 Federation Rule Mining

It is the process of finding combinations or casual structures, interesting relationships and frequent patterns that exists among the set of items in a business database. Federation rule mining is widely used in areas such as telecomm networking, market-research, inventory stock control, risk management, etc. Different areas where alliance standard mining assumes a fundamental part and are discovering layouts in organic information investigation, breaking down library information, protein synthesis study, investigation of populace and monetary evaluation, and so on. Subsequently alliance standard mining is one of the center ideas in learning disclosure process[1].

Performing federation rule mining is a difficult task and the whole process is dissolved into two sub tasks. The first task in federation rule mining is to find frequent item sets (whose occurrences exceeds a predefined limit in the transaction database). The second task is to extract federation rules from large item sets that satisfy the minimum confidence. The first task is to generate the candidate large item sets and frequent item sets. Providing safety measure in federation rule mining process is decided by practical application of different privacy protection requirements[2]. The traditional methods include data contortion, data encryption and data released. More number of algorithms for privacy preservation was refined mainly based on encryption method. Less number of algorithms was described on data streams. In future this paper reviews several algorithms relating to privacy preservation during federation rule mining process.

Privacy is important for federation rule mining process[3]. It is utilized to remove appropriate learning from enormous measures of information while securing in the meantime touchy data. An essential perspective in the configuration of these calculations is the distinguishing proof of important assessment criteria and the

* *Author for correspondence*

advancement of related benchmarks. Recent research in the area has devoted much effort to determine a trade-off between the right to privacy and the need of knowledge discovery[4]. The existing System surveyed about lot of privacy protection techniques such as, data contortion, Disruption method, Block-board federation rule mining, Remodeling based federation rule mining, Crypt analysis based federation rule mining[5].

1.2 Reconstruction Algorithm

It is an Innovative and protected reconstruction scheme with privacy measure. Reconstruction scheme initially adds the noise element to original data fields in the data set and updates the same with set of associated key pairs. When the third party desire accessing the sensitive data, in order to access those data, user admin requires a key and legitimate can perform modification or alteration of data[6].

1.3 Data Contortion

The first step is to classify the primary concretion function of the original series and to estimate the parameters of this concretion function. The second step is to create a progression of information from the anticipated thickness capacity. Last stride is to arrange and recover the produced arrangement for the first one. Since it is recovered by the contorted information set, likelihood bending protects the security of an individual having a place with the first information set[7]. In the meantime, the likelihood of contorted arrangement gives demographic the same diagnostic properties as those of the first gathering, since both are under the same dissemination. Not at all like routine point contortion, anticipated twisting is hard to trade off by reduplicated inquiries, and contribute a greatest bearing for demographic analysis[8]. By surveyed about test perturbation technique individual can adapt their privacy protection. In which every individual can report their original information to the minor data. But it is possible to the Intruder can enter and modify their original data.

1.4 Disruption Method

It is used to filter their original dataset. It follows some of the rules like primitive rules and non-primitive rules. Primitive rules are kept from original database. No primitive rules are removed from the database[9].

1.5 Block-Board Federation Rule Mining

Data sharing information hiding is a major problem. It provides an efficient method for federation rule mining method[10]. It is mainly to improve desirable side effects and reduce the undesirable side effects. It is used to handle large volumes of data. Confidentiality of the data will maintain for entire large data.

1.6 Remodeling based Federation Rule Mining

It is used to perform the perturbation of data after that reconstruct their distributions. Remodeling process is used to reconstruct their original data. Most of the remodeling process is using an OB (Outlook boost) algorithm, which can estimate their original distribution whenever a huge amount of data is obtained[11]. OB algorithm is used to keep the original data in somewhere else. Mining process is start from sensation knowledge discover data and the new data will reconstructed from sensation knowledge discover data.

1.7 Crypt Analysis based Federation Rule Mining

It is used to share the data or information with different users who they are entering into the federation rule mining process[12]. This is based on crypto security analysis. It is same as Alice and babu security cryptosystems process. First Alice can encrypt their original dataset for their federation rule mining. Babu will decrypt their original content. By applying Feature subset selection method given data's are encrypted. It is an efficient technique to encrypt their data. But it suffers from leakage problem while encrypting data or information.

1.8 Problem Identification

In Data contortion method, main problem is easy to modify the original records to handle the largest records. Issue of the disruption method can be done should not alter the correct data by adding noise to the original data. Confidentiality should not be maintained Block-based federation rule mining.

2. Proposed Methodology

Implementation of Frequent Pattern Techniques more

thrust one. It is more efficient than other algorithms; it uses large databases of business environment. Uses Divide and conquers methodology[13]. It is based on top-down approach for designing algorithms that consists of dividing the problem into smaller sub problems. First find out the solution for smaller problem then find out the solution for original problem. It is more efficient compared to existing techniques. The main objective is to compute on large data efficiently. It is an efficient and scalable method for mining the complete set of frequent patterns by using frequent pattern Tree format.

A user's transaction database is a ordering of transactions (k=k1…KN), where each transaction is an item set (ki⊆M). An item set with n elements is called an n-item set. The frequent item set problem is to find all frequent items set in a given transaction database. It follows FP-Produce algorithm.

2.1 FP-Produce Algorithm

FP-produce utilizes a blend of the vertical and level database design to store the database in primary memory. Rather than putting away the spread for each thing the database, it stores the genuine exchanges from the database in a tree structure and each thing has a connected rundown experiencing all exchanges that contain that thing. This new information structure is indicated by FP-tree (Frequent-Pattern tree). FP-produce is a key incessant thing set mining calculation, which depends on the example development worldview. It contains, Item set-It is an accumulation of one or all the more thing set in a given bunch. Bolster Frequency of event of a set. Bolster Count-division of event of a thing set in a given bunch.

2.2 Split Algorithm

It is the base of efficient algorithms for any kind of such problems (e.g. quick sort, merge sort, and computing discrete Fourier cosine transforms). It is used to reduce the Complex problems or larger problems into smaller problems. By using this algorithm optimized outputs released. It will naturally suitable for multiple process executions[14]. In Tower Of Hanoi Problem fathoms scientific diversion settling technique. It comprises of three bars, and various plates of various sizes which can slide onto any bar. The riddle begins with the circles in a perfect stack in rising request of size on one bar, the littlest at the top, hence making a funnel shaped shape.

2.3 Data Sensation Algorithm

Input: Ordinary Data

Output: Sanitized Data

Step 1: Create a collection of a data

Step 2: Collect the Original Database from their data

Step 3: Modify or remove the duplicate items in a database.

Step 4: Reduce frequently used data items.

Step 5: Generate filtered frequent item datasets.

Step 6: Extract from sensation data.

3. Statistical Analysis

3.1 Overview

It is implemented by WEKA Tool. WEKA is tool used for developing expert Systems (ES) techniques and their application to natural world data mining problems. It will convey you step by step through the separation of a simple problem using WEKA pioneer pre possessing, categorization, clustering, combination, feature selection, and visualization tools. WEKA tool is the specific development is currently possible has information programming system been used. Be that as it may, there happen to be different strategies inside the past for institutionalization to look at their information and use it further bolstering their good fortune. Past taking opinion, or making use of store scanners, product codes and bar codes, folks happen to be in a position to collect info, valuate it and utilize it with their advantage. Yet it cannot be discard that the convenience to better technology has incomparably upgraded the facility to store or gather info, make prevision about outcomes and rehearse client trend reports to greater positive aspects.

3.2 Machine Learning Algorithm

Machine learning algorithm implemented by WEKA Tool it consists of classification and regressions, clustering, finding associations, attributes selection, data visualization, etc. It can be classify into Traditional Programming and machine learning. Traditional programming is used to run on the computer and to produce the output[15]. Machine learning is used to run on the computer to create a program which is used in traditional programming. Variety of applications based on machine learning process such as, e-Commerce, Robotics, Debugging, Social networks, information extraction, finance etc.

4. Results and Discussion

In this paper, frequent pattern technique is to find frequent item sets in the transaction database which is ensure about privacy and security of information sharing, considering frequent pattern technique such as, it must produce good privacy with accuracy of the given dataset. Federation rule mining techniques should also be designed for calculating efficiency. It is an excessive growth in dimensional and geological applications. Compared to other techniques sanitized data's are extracted in a proper manner. Data sharing to third parties is a confidential one.

5. Conclusion

This paper concluded about the privacy protection techniques, but Frequent Pattern technique is an effective one in federation rule mining process[16]. The algorithm initially introduces split algorithm where complex problems are subdivided into smaller problems. Second, FP–produce algorithm is used to find the entire frequent item datasets in a database. Once all the data are found, the complete set of frequent item sets can be obtained. FP–produce algorithm also eliminates a repeated dataset in a given database. It performs in different application areas including market research, risk assessment, commercial environments, crime prevention, etc.

6. References

1. Dragos N, Trinca T. Fast and Cost-Effective Algorithms for Information Extraction in some Computational Domains. International Journal of Dissertation Abstracts. 2008; 70:01.
2. Bertino E, Lin D, Jiang W. Springer Publication: A Survey of Quantification of Privacy Preserving Data Mining Algorithms. 2008; 34:183-205.
3. Hemalatha R, Elamparithi M. Privacy Preserving Data Mining Using Sanitizing Algorithm. (IJCSIT) International Journal of Computer Science and Information Technologies. 2015; 6(5):4174-79.
4. Vaidya J. Clifton C, Kantarcioglu M, Patterson S. Privacy-preserving decision trees over vertically partitioned data. 2008; 2(3):1-4.
5. Aruna Kumara T, Gunasekar. A Reconstruction Algorithm using Binary Transform for Privacy–Preserving Data Mining. Indian Journal of Science and Technology. 2016 May; 9(17):1-5.
6. Han J, Cheng H, Xin D, Yan X. Springer Publication: Frequent pattern mining: current status and future directions. 2007; 15(1):55-86.
7. Shuting XU, Zhang J, Han D, Wang J. Springer publication: Data Distortion for Privacy Protection in a Terrorist Analysis System. 2005; 3495:459–64.
8. Kamakshi K, Babu V. Preserving Privacy and Sharing the Data in Distributed Environment using Cryptographic Technique on Perturbed data. Journal of Computing. 2010; 2(4):115-19.
9. Liu L, Kantarcioglu M, Thuraisingham B. The Applicability of the Perturbation Model-based Privacy Preserving Data Mining for Real-world Data. Dallas, TX: Sixth IEEE International Conference on Data Mining-Workshops (ICDMW'06). 2006.
10. Dass R. Indian Institute of Management Ahmedabad: An Efficient Algorithm for Frequent Pattern Mining for Real-Time Business Intelligence Analytics. 2005; p. 1-16.
11. Paul R, Groza T, Hunter J, Zankl A. Semantic interestingness measures for discovering association rules in the skeletal dysplasia domain. Journal of Biomedical Semantics. 2014; 5(8):1-4.
12. Shana J. A Cryptography Based Privacy Preserving Association Rule Mining in Academic Analytics. World Engineering & Applied Sciences Journal. 2015; 6(1):41-44.
13. Usha D, Rameshkumar K. A Complete Survey on application of Frequent Pattern Mining and Association Rule Mining on Crime Pattern Mining. International Journal of Advances in Computer Science and Technology. 2014; 3(4):1-12.
14. Aalst MP W V. A General Divide and Conquer Approach for Process Mining Architecture of Information Systems. Federated conference on computer science and information systems. 2013; p. 2.
15. Lindell Y, Pinkas B. Secure Multiparty Computation for Privacy-Preserving Data Mining. Journal of Privacy and Confidentiality. 2009; 1(1):59-98.