Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7778443

PERFORMANCE EVALUATION OF HIDING ALGORITHMS USING ASSOCIATION RULE MINING

Yamini richhariya
MITS college Gwalior

Dr. Saumil Maheshwari
Asssistant Professor, Department of IT, MITS,Gwalior

ABSTRACT

Data privacy is of paramount importance in today's digital age, with a growing need to protect sensitive information while still enabling data analysis. In the era of data-driven decision-making, ensuring data privacy has become a paramount concern. In order to protect sensitive information while still making data usable for analysis, association rule concealing algorithms have emerged as a potential approach. Hiding algorithms, particularly that using association rule mining, have developed as powerful methods for striking this equilibrium. In this study, we show how FP-growth stacks up against Apriori in terms of performance. Results are broken down by execution time, instance count, and trust in the Supermarket data set to draw conclusions about performance. Both the algorithms and their experimental results are provided. Compared to the Apriori approach, the FP-growth technique is roughly an order of magnitude quicker and more scalable, as shown by our performance analysis.

Keywords: Association rule, Hiding, Data mining, Algorithm, Growth

I. INTRODUCTION

In the rapidly evolving landscape of data-driven decision-making and digital transformations, the paramount importance of data privacy has become increasingly evident. As organizations and individuals a like leverage data for insights and strategic advantages, the necessity to protect sensitive information from unauthorized access and malicious intent has become a pressing concern. This concern is compounded by the growing regulatory landscape surrounding data protection and the potential reputational and financial consequences of data breaches. Hiding algorithms, emerging as a crucial facet of data privacy preservation, offer a solution to this intricate challenge. These algorithms, often grounded in the principles of association rule mining, provide a means of ensuring the confidentiality of sensitive information while still enabling valuable data analysis. This research paper embarks on a comprehensive exploration of the performance evaluation of various hiding algorithms, employing association rule mining as the underlying framework. By investigating the effectiveness of these algorithms in safeguarding data privacy and maintaining data utility, this paper aims to contribute to a deeper understanding of the delicate balance between privacy and analysis in today's data-driven world.

Data privacy is a multifaceted concern that permeates industries, domains, and individuals alike. The digital age has ushered in an era where data serves as the lifeblood of decision-

making processes. Be it for predictive analytics, customer insights, supply chain optimization, or medical research, the potential benefits of data analysis are immense. However, this potential is accompanied by an inherent vulnerability – the risk of unauthorized exposure of sensitive information. With the advent of high-profile data breaches, the revelation of personal information, and the increasing sophistication of cyberattacks, the criticality of preserving data privacy has risen to the forefront of public consciousness. Association rule mining, a cornerstone of the field of data mining, has found practical applications in a wide range of domains. Its ability to uncover meaningful relationships, dependencies, and patterns within large datasets has been harnessed for decades. In recent years, the fusion of association rule mining with hiding algorithms has led to the development of innovative techniques aimed at enhancing data privacy. The fundamental principle underlying these algorithms is the transformation of the original data in a way that conceals sensitive attributes or information while maintaining the overall statistical characteristics and utility of the dataset.

II. ASSOCIATION RULE HIDING ALGORITHM

The term "Association Rule hiding algorithm" refers to a class of algorithms that utilize the principles of association rule mining to enhance data privacy by concealing sensitive information while preserving the statistical and functional properties of the data. These algorithms are designed to achieve a delicate balance between safeguarding privacy and maintaining data utility, which is vital in today's data-driven world. Association rule mining is a technique used in data mining that aims to discover interesting relationships or patterns in large datasets. It involves identifying associations, dependencies, and correlations between items or attributes within the data. This technique has been widely applied in various domains, including market basket analysis, recommendation systems, and healthcare research.

The concept of association rule hiding algorithms builds upon this foundation by addressing the growing concern for data privacy. In scenarios where datasets contain sensitive information such as personal identifiers, medical records, or financial details, traditional data sharing and analysis techniques may risk exposing this sensitive data. Association rule hiding algorithms step in to mitigate this risk by altering the data in a way that prevents unauthorized identification of individuals while still enabling valuable insights to be derived from the data. These algorithms typically operate by modifying the dataset to ensure that sensitive associations or patterns cannot be directly deduced from the data. For example, consider a dataset containing medical records with information about patients' diagnoses and treatments. An association rule hiding algorithm might alter the data so that certain associations, such as a specific diagnosis linked to a particular patient, are no longer easily discernible.

III. REVIEW OF LITERATURE

Mohan, S. & Angamuthu, Tamilarasi (2021) Keeping personal information secure when mining data is recognized as one of the field's most promising areas of study. Data mining techniques enable individuals uncover previously unknown connections and hidden meanings in large datasets. Knowledge acquired from most circumstances comprises private information about people and businesses. Furthermore, this private data can be exploited in ways that

violate the individual's privacy. Often, crucial aspects of a company's marketing strategy are set in stone by the regulations of an association. By efficiently summarizing the data and revealing any previously unknown relationships between entities in the data, significant association rules are a valuable source of insight for the data miner. When mining for associations, it's important to hide any potentially sensitive rules so that any insights gained from those rules remain safe. During association rule concealing, some sensitive association rules are removed from the original database without significantly altering the data or the non-sensitive rules. In this piece, we offer a novel method of concealment that uses a genetic algorithm (HGA), as well as another method that uses fake items (DIC). The sensitive association rules are hidden using a genetic algorithm-based hiding approach, and the updated sensitive items are hidden alongside their dummy counterparts using a dummy-items-creation technique. It has been demonstrated experimentally that the offered methods are effective.

Gayathiri, P. & Balakrishnan, Poorna (2017) Increases in data collecting, storage, and analysis pose significant threats to the privacy of individuals whose identities are associated with such datasets. In a number of scenarios, the retrieved knowledge is extremely sensitive, necessitating a cleansing process before it can be made public. The technology behind data mining can quickly and easily glean vast amounts of information. It's possible that the most private details about a person or company might be exposed by the information gleaned from an intelligent data mining system. The degree of confidentiality attached to a person's or a company's data varies. This information is restricted to authorized users only. Therefore, restricting access is not an adequate technique for protecting sensitive information. As a result, the 'Inference Problem' may arise, whereby the user is able to re-identify sensitive data items from non-sensitive data with the assistance of the information gained from the data mining process. By creating data mining methods that may be used on databases without diminishing the quality of data mining results, privacy preserving data mining aims to address the issue of protecting personally identifiable information. simultaneously without compromising people's personal space. This paper presents the results of a comprehensive research project on the subject of association rule hiding algorithms.

Gayathiri, P. & B, Dr. (2015) For large sets of transactional datasets, association rule mining is a useful data mining approach for identifying commonalities and associative rules from market basket analysis. Customers' product-buying patterns can be represented by an associative rule, which is presented by calculating the probability of the most often occurring data item from the transactional data pieces. Data mining's discovery of associative rules in a transactional database presents a potential risk of compromising personal and corporate data. The problem of privacy leaks in data mining can be remedied by Privacy Preserving Data Mining (PPDM). Association Rule Hiding (ARH) methods developed within the context of Privacy Preserving Data Mining (PPDM) are used to address this concern. This study of the Association Rule Hiding methodology in data mining employs a method of concealing the development of private association rules from transactional data. The sensitive rule hiding approach has fewer unintended consequences and more data value since it conceals just the rules and not the data.

Al-Janabi, Sufyan et al., (2014) The capacity to keep personal data about users and the complexity of data mining algorithms have both contributed to the growing urgency of the challenge of Privacy Preserving Data Mining (PPDM) in recent years. In recent years, a variety of methods have been proposed for carrying out PPDM. These methodologies are used to investigate various privacy-related transformation strategies. An approach for PPDDM of association rules is provided in this research. In order for this system to function, it relies on the widely held and plausible assumptions that STTPs (Semi-Trusted Third Parties) are involved and that databases are horizontally distributed across STTPs. A new method for concealing private rules is introduced here. When compared to the same technique in a centralized system and other current algorithms in a distributed database system, the experimental findings reveal that this approach has high concealment accuracy with an acceptable degree of side effects. Additionally, the proposed system use SSL's commutative encryption to back certificates and protect the system's many moving parts.

Sreenivasa Rao, Kuncham et al., (2014) Research into privacy-preserving data mining focuses on protecting individuals' anonymity and privacy in the context of data collection and analysis. This study solves the privacy issue by taking both privacy concerns and algorithmic needs into account. The work presented in this article is an attempt to create an association rule concealing algorithm for privacy preserving data mining that would be effective in ensuring secrecy and would boost speed when dealing with large amounts of data stored and retrieved from a database. In this study, we evaluate the suggested method in comparison to the two state-of-the-art alternatives, ISL and DSR.

Jain, Yogendra et al., (2011) When data is shared across a network, protecting the big database that holds sensitive information becomes a major concern. Research on privacy data for data mining and statistical databases has recently taken a turn toward privacy-preserving data mining. Association analysis is an effective method for unearthing previously undetected connections within a dataset. When it comes to hiding sensitive or vital information, association rule hiding algorithms work exceptionally well. One of the most crucial methods for protecting information is changing and disguising the rules that govern it. The suggested Association rule concealment algorithm for private data mining is meant to conceal data from being mined using an association rule algorithm. To conceal part of the created association rules, most methods for concealing them work by either increasing the rules' support or decreasing their confidence. Left-hand side (LHS) and right-hand side (RHS) association rule elements that cannot be inferred using association rule mining techniques. The idea behind the Increase Support of Left Hand Side (ISL) method is to lower the confidence in the rule by raising the degree of support for the LHS. The rule can only be modified on the left hand side; it does not apply to the right. In the DSR algorithm, rule confidence is lowered by reducing the support value on the right hand side. As a means of RHS alteration, it is effective. To address this issue, we suggested a novel method. More rule can be hidden with fewer changes if the support for the LHS and RHS item of the rule is adjusted accordingly. Using actual databases, we compare the proposed algorithm's efficiency to that of ISL methods and DSR algorithms in terms of the number of rules hidden, CPU time, and the number of modified entries.

Gkoulalas-Divanis, Aris & Verykios, Vassilios (2010) Privacy protecting data mining is a relatively new area of study that focuses on the potential privacy and security hazards associated with applying various data mining techniques to big institutional data repositories. The emerging field of data mining known as "association rule hiding" investigates the challenge of concealing private association rules inside datasets. In this paper, we offer a number of heuristic solutions to the problem of "hiding" sensitive association rules in data mining. In this article, we present several recently proposed exact solutions with increased time complexity, as well as a number of computationally efficient (parallel) approaches that alleviate time complexity problems, and we discuss closely related problems (inverse frequent item set mining, data reconstruction approaches, etc.) in great detail. Important parts of this difficult topic are studied, assimilated, and appreciated by the reader, along with unsolved difficulties, future directions, and particular instances. Researchers, academics, and advanced CS students interested in privacy-preserving data mining, association rule mining, and data mining will find Association Rule Hiding for Data Mining to be an invaluable resource. Professionals in the field can also benefit from reading this book.

IV. RESEARCH METHODOLOGY

WEKA version 3.6.1 was used to evaluate the two association rule mining techniques. The WEKA program is a free and open-source compilation of several data mining and machine learning techniques, such as those used for data pre-processing, classification, clustering, and the extraction of associations. Both Apriori and FP-growth were scored on how quickly they could complete a task. Time to execute is analyzed on the Supermarket dataset with varying instance counts and confidence levels.

Both methods have been examined by us using the supermarket dataset. There are a total of 3626 occurrences and 215 characteristics in this data collection. Our experiment makes use of data imported from an ARFF file. We have utilized the graphical user interface (GUI) version of WEKA to do the efficiency analysis. In the preprocess tab, we use OpenFile to import the database. To evaluate the speed of these algorithms, we have chosen APriori and FP-growth under the Associate menu.

V. DATA ANALYSIS AND INTERPRETATION

Here, we compare the efficiency of various ARM-based algorithms. The tables below show the results of running tests with varying numbers of instances and levels of Confidence using Apriori and FP-growth.

Table 1 Execution Time for Different Instances

Instances	**Execution Time (in Secs)**	
	Apriori	**FP-growth**
3626	46	5

1687	25	3
940	8	1

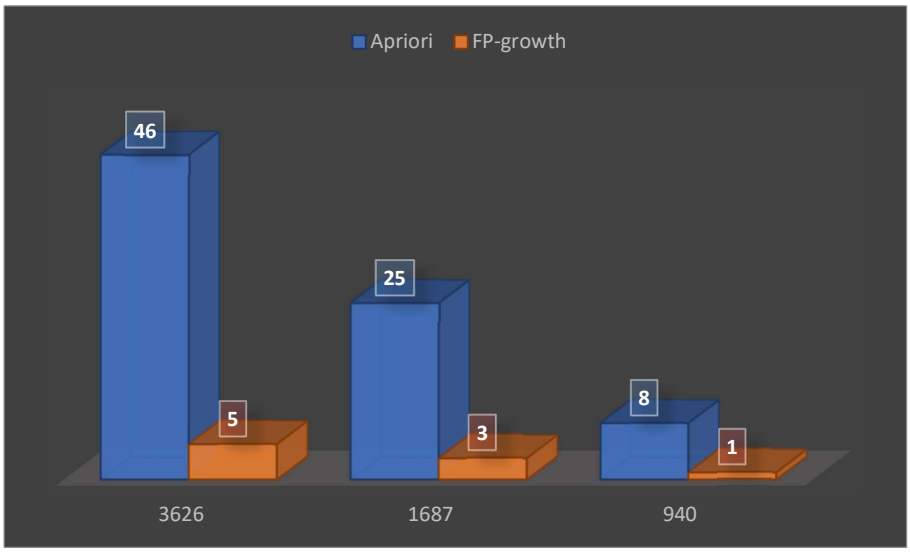

Figure 1: Comparison between Apriori and FP-growth with respect to Execution Time for Different Instances

Thus, both algorithms' execution times dropped as the number of instances was reduced. APriori takes 46 seconds to generate the association rules for the 3626 occurrences of the supermarket data set, whereas FP-growth only takes 5 seconds.

When pitted against Apriori, FP-Growth always finishes faster, regardless of the number of instances being processed. Thus, for many different case counts, FP-growth surpasses Apriori in terms of performance over time.

Table 2 Execution Time for Different Confidence Level

Confidence Level	**Execution Time(in Secs)**	
	Apriori	**FP-growth**
0.6	18	2
0.8	15	3
0.7	55	4

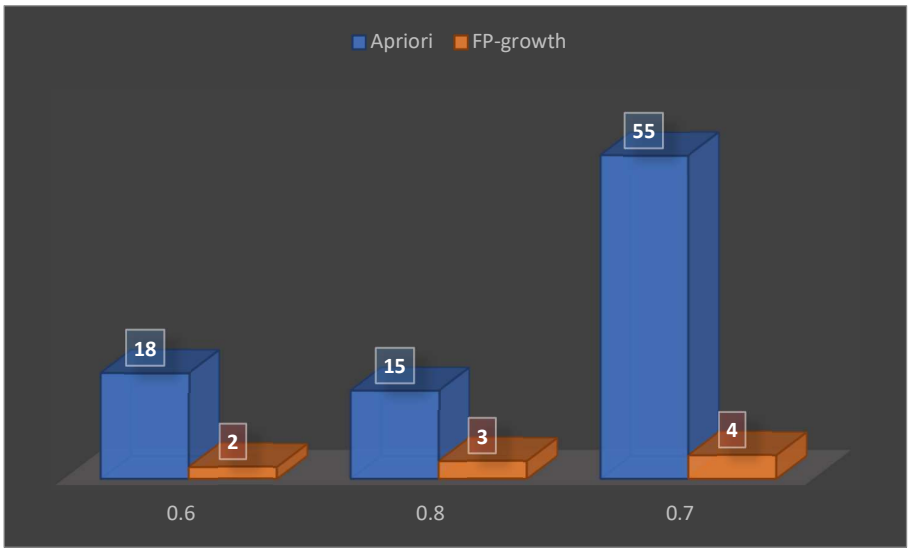

Figure 2: Comparison between Apriori and FP-growth with respect to Execution Time for Different Confidence Level

Table & figure highlights the time it takes to run Apriori and FP-growth at different levels of confidence. When Confidence is high, both methods require a lot of time. The time required to construct the association rule is 2 seconds in FP-growth compared to 18 seconds in Apriori for the same degree of confidence (0.6).

It's stated that, for any given level of confidence, the execution time of FP-growth is smaller than that of Apriori. Therefore, the FP-growth Algorithm provides a scalable and effective means of mining the full complement of common patterns.

VI. CONCLUSION

The significance of association rule hiding algorithms extends beyond the confines of this research. The insights derived from this study contribute to both the academic discourse and practical implementation of privacy-preserving techniques. These algorithms hold the potential to reshape the data sharing and analysis paradigm, enabling organizations and individuals to harness the power of data while ensuring responsible and ethical practices. In a landscape marked by evolving regulations, increasing data breaches, and a growing awareness of privacy concerns, the fusion of data utility and privacy preservation becomes not merely a choice, but a necessity. Association rule hiding algorithms stand as a testament to human ingenuity in addressing this challenge. They embody the convergence of mathematics, computer science, and ethics in a manner that empowers us to unlock valuable insights from data while safeguarding individual rights.

REFERENCES: -

1. Mohan, S. & Angamuthu, Tamilarasi. (2021). Association Rule Hiding in Privacy Preserving Data Mining. 10.4018/978-1-7998-8954-0.ch044.

2. Gayathiri, P. & Balakrishnan, Poorna. (2017). Association Rule Hiding for Privacy Preserving Data Mining : A Survey on Algorithmic Classifications.

3. Ahmed, Gehad & Abd_Ellatif, Laila & Sharaf, Ahmed. (2016). Association Rules Hiding for Privacy Preserving Data Mining: A Survey. International Journal of Computer Applications. 150. 34-43. 10.5120/ijca2016911664.

4. Kumar, Umesh & Singh, Anju. (2016). Approaches for Privacy Preserving Data Mining by Various Associations Rule Hiding Algorithms – A Survey. International Journal of Computer Applications. 134. 21-26. 10.5120/ijca2016908042.

5. Gayathiri, P. & B, Dr. (2015). Association Rule Hiding Techniques for Privacy Preserving Data Mining: A Study. International Journal of Advanced Computer Science and Applications. 6. 10.14569/IJACSA.2015.061232.

6. Al-Janabi, Sufyan & Jumaa, Alaa & Qadir, Nizar. (2014). Hiding Association Rules over Privacy Preserving Distributed Data Mining. Kirkuk University Journal- Scientific Studies. Vol. 9. pp. 59-72..

7. Sreenivasa Rao, Kuncham & Mandhala, Venkata & Bhattacharyya, Debnath & Kim, Tai-hoon. (2014). An Association Rule hiding Algorithm for Privacy Preserving Data Mining. International Journal of Control and Automation. 7. 393-404. 10.14257/ijca.2014.7.10.36.

8. Jain, Yogendra & Yadav, Vinod & Panday, Shirshendhu. (2011). An Efficient Association Rule Hiding Algorithm for Privacy Preserving Data Mining. International Journal on Computer Science and Engineering. 3.

9. Gkoulalas-Divanis, Aris & Verykios, Vassilios. (2010). Association Rule Hiding for Data Mining. 10.1007/978-1-4419-6569-1.