ISSN 2348 – 9928
IJAICT Volume 4, Issue 7, July 2017 Doi:01.0401/ijaict.2014.08.04 Published on 05 (7) 2017

ENHANCE THE SENSITIVE HIDING RULES VIA NOVEL APPROACH

Mr. A. Karthikeyan ***Ms. M. S. Vinmathi*** ***Mr. M. Mahendran***

Assistant Professor,
Computer Science and Engineering,
Panimalar Engineering College,
Chennai, Tamilnadu, India

Abstract— Data mining techniques have been extensively used in a variety of applications. However, the mistreatment of these techniques may lead to the discovery of sensitive information. Researchers have newly made efforts at hiding sensitive association rules. The Hiding of Sensitive Rules is done by Modification Schemes. However, undesired side effects, e.g., no sensitive rules wrongly hidden and spurious rules falsely generated, may be formed in the rule hiding process. In this paper, we present a novel approach that intentionally modifies a few transactions in the transaction database to diminish the supports or confidences of sensitive rules without producing the side effects. Since the correlation between rules can make it impracticable to accomplish this goal, in this paper we recommend heuristic methods for increasing the number of secreted sensitive rules and reducing the number of modified entries. The experimental consequences show the value of our approach, i.e., undesired side effects are avoided in the rule hiding process. The outcome also report that in most cases, all the sensitive rules are hidden without spurious rules falsely generated. Moreover, the good scalability of our approach in terms of database size and the authority of the correlation among rules on rule hiding are experimental.

Keywords - Modification Schemes, supports confidences.

I. INTRODUCTION

Web server register a (web) log entry for each single access they get in which they save the URL requested, the IP address from which the demand originated, and a timestamp. With the rapid development of World Wide Web (WWW) technology, a enormous number of web log access log records are being unruffled.

It is not effortless to perform systematic analysis on such huge amount of data, though many people realized the possible usage of data to make successful use of web access history for server presentation, system design improvement, or customer targeting in electronic commerce [2]. With site pulling out, the overall quality and value of the pages at the site can be evaluated. The different modes of usage called user profiles can be exposed using a clustering that extract access patterns from the click streams stored in web log files. Using web log files, studies have been conducted on analyzing system performance, improving system design, accepting the nature of web traffic, and perceptive user reaction and motivation. Web sites that recover themselves by learning from user access patterns. Most of the web log examination tools have boundaries with regard to the size of the web log files. Dissimilar assumptions are made for each web analysis tools consequences in different information with the same log file.

Web server log files include useful information from which a well deliberate can discover of assistance information. Web server log files routinely contain: the domain name (or IP address) of the request; the date and time of the request; the technique of the request (GET or POST); the name of the file requested; the result of the request (success, failure, error etc.); the URL of the referring page[6]. A log entry is mechanically added each time a demand for a supply reaches the web server.

In this paper, data mining techniques are projected to analyze web log records. Mass profiling is based on general trends of procedure patterns compiled from all users on a site, and can be achieved by mining abuser profiles from the chronological data stored in server access logs [7]. I have presented an evolutionary approach, called Hierarchical Unsupervised Niche Clustering (H-UNC), for simultaneously mining Web routing patterns and maximally numerous context-sensitive URL item sets from the celebrated user access data stored in Web server logs.

H-UNC necessitates fixing the number of clusters in advance, is insensate to initialization, can handle noisy data, general non differentiable correspondence measures, and automatically provides profiles at multiple promise levels. Unlike content based connection methods, this advance also discovers associations between dissimilar Web pages based only on the user admittance patterns and not on the page contented. Its hierarchical mode, very small residents sizes contributed to making H-UNC very reasonably priced from a computational position, especially when

ISSN 2348 – 9928
Doi:01.0401/ijaict.2014.08.04 Published on 05 (7) 2017

compared to standard evolutionary computation based data mining techniques.

The Web access patterns on a web site are very dynamic in nature, appropriate to the dynamics of Web site content and structure, changes in the user's interests, and thus their navigation patterns [4]. The access patterns can be observed to alteration depending on the time of day, day of week, week of month, month of year. Here, an advance that considers the Web usage data as a suggestion of a dynamic environment which consequently requires dynamic learning of the access patterns. This evolutionary computation based approach can be generalized to fit the needs of mining dynamic data or huge data sets that do not vigorous in main memory.

The remaining of this paper is organized as follows. In Section 2, it describes the design of a data mining system for web log records. Implementation efforts are presented in section 3. Finally, Section 4 and 5 summarizes how the work is concluded.

II. DESIGN OF DATA MINING SYSTEM FOR WEB LOG RECORDS

Profile discovery based on web treatment mining which starts with the amalgamation and preprocessing of Web server logs and server comfortable databases, includes data cleaning and sessionization, and then continues with the data mining/pattern discovery via clustering. This is followed by a post processing of the clustering consequences to obtain Web user profiles and finally ends with tracking profile progression.

Web server log files enclose useful data from which a well deliberate data mining system can discover beneficial information. In this mining developing user profiles project, the data collected in the web logs goes through three stages. In the first stage, the data is filtered to confiscate irrelevant information and a database is created containing the meaningful enduring data. This database facilitates information extraction based on individual attributes resembling client-IP, resource, day etc. In the second stage, it continues with pattern discovery via clustering and summarizes gathering clusters into user profiles. As a final point, in the third stage the current profiles are tracked with existing profiles. The log entries with figures (gif, jpg, etc.) are disinterested.

A common technique for a server site to divide the log records into sessions. A session is a set of page references from one source site throughout one logical period. A session would be well-known by a user logging into a computer, the stage work, and then logging off. The login and logoff represent the consistent start and end of the session.

2.1 Database construction from server log files

The data filtering stride may filter out requests in order to contemplate on data pertaining to actual page hits. The data filtering was adopted mainly transforms the data into a supplementary meaningful representation. Cleaning the data and time field of the log entry, it is simply reorganized in a set of fields to specify the day, month, year, hour, minute and seconds. The renovation process replaces the request sequence by the diplomat URL. After the cleaning and transformation of the web log entries, the web log is overloaded into a relational database.

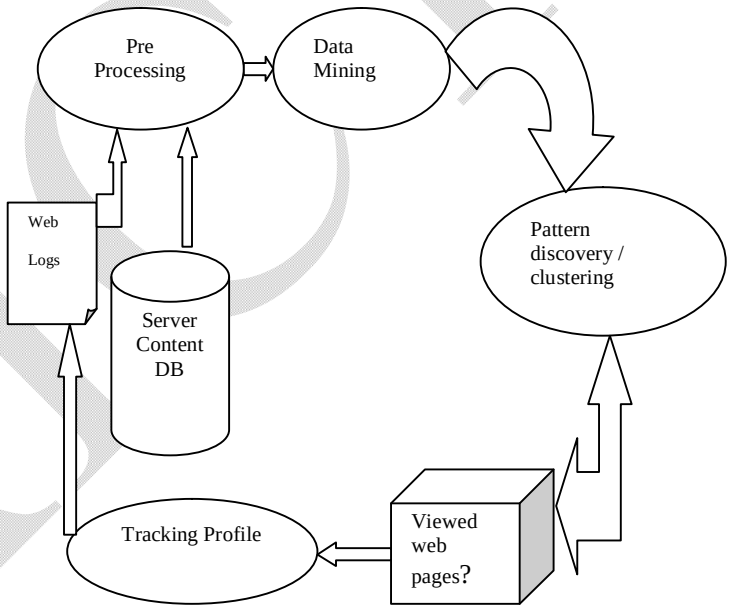

Fig 1 : Web Usage Mining Process

Log Record before preprocessing

Fields: date time s-sitename s-computername s-ip cs-method cs-uri-stem cs-uri-query s-port cs-username c-ip cs-version cs(User-Agent) cs(Cookie) cs(Referrer) cs-host sc-status sc-substatus sc-win32-status sc-bytes cs-bytes time-taken.

Log record after preprocessing

Fields: time c-ip cs-method cs-uri-stem sc-status

2.2 Clustering User sessions into an optimal number of categories

To cluster client sessions, a discordant hierarchical description of a vigorous clustering come up to (UNC) that use Genetic

ISSN 2348 – 9928
Doi:01.0401/ijaict.2014.08.04 Published on 05 (7) 2017

Algorithm to evolve the population of candidates. GA starts by apply the operators' selection, crossover and alteration.

Special Crossover is the interests perform an independent intersect for each sub chromosome in the stumpy level. First, a measure of the aloofness between the sub chromosomes of the parents is compute, and each sub chromosome from solitary parent is matching with the nearly all comparable unpaired sub chromosome from the added parent.

After that, a one top crossover stuck between the paired sub chromosomes is ended (the entire sub chromosomes participate in the crossover). Finally, the commencement facts of the elevated level is cross, by performing arts a solitary point cross linking each pair of parallel high plane activation string (the correspondence is obtained from the matching between the little level corresponding sub chromosomes).Hierarchical unofficial Niche Clustering (H-UNC), for drawing out both user outline clusters and URL associations.

This comes up to prove to be successful in mining cluster from large web session data. H-UNC can grip noise in the data and robotically determine the figure of cluster. It is mentioned richly in segment.

2.3 Tracking Evolving User Profiles

Tracking the different user profile events among different time schedule generate better understanding of the development of user access patterns. Both user profiles and click streams are naturally developing; each profile p_i is revealed with a measure of scale σ_i that represents the amount of inconsistency or dispersal of the user sessions in a given cluster around the cluster agent. This measure is used to agree on the boundary around each cluster and thus allows us to automatically determine whether two profiles are well-suited.

The notion of compatibility between profiles is essential for tracking evolving profiles. After mining the Web log of a given period, perform a mechanized comparison between all the profiles revealed in the current group and the profiles discovered in the previous batch by a series of SQL queries on the profiles that have been stored in a database.

III. IMPLEMENTATION EFFORTS

3.1 Mining of Strong Association Rules in Database

Association rule mining is a process of ruling the set of strong association rules in an operation database. An item set is a set of essentials recorded in a database record set. The prop up and Confidence of all the item set in the transaction database is calculated using Apriori Algorithm.

The Strong Rules are clean out by using the predetermined Minimum Support Threshold and Minimum Confidence Threshold values. The Rules (set of item sets) which comprise the Support and Confidence levels more than the minimum Threshold levels are filtered out as Strong Rules by implementing the Apriori Algorithm. The Sensitive Rules among the Strong connection Rule are elected based on the Business interest and are separated to one side to be targeted for Rule Hiding.

3.2. Evaluation of Modification Schemes for Rule Hiding

The different adjustment schemes are evaluate for Efficient Rule hiding process. The scheme like deletion, Insertion and exchange (both deletion and insertion together) are calculated. The range of the exact Modification plan is done based on the factors 1) The modification scheme must affect minimum number of transaction item sets to hide the rule 2) No Side Effects - Weak Rules trouncing and new Spurious generation have to arise.

It can be see that a modification schemes consists of three elements, including the adaptation scheme (either deletion or insertion), the item, and the transactions to be modified. The effects of a modification can be different if one of its elements is changed. It may enlarge or decrease both Support and assurance.

A adjustment is said to be valid if it will not produce any face effect in the modified database. Because a number of valid modifications can be functional to hide a sensitive rule, the end of this module is to select an appropriate subset of valid modifications to hide the susceptible rules. Specifically, the adjustment that affects the least number of dealings and helps to secrete the most digits of sensitive rules has priority over the others.

3.3 Index Construction and Template Generation

We relate to the an algorithm for association rule taking out with MST 20% and MCT 60% and obtain the frequent item sets and the muscular rules .We encode each item as a unique prime number such that all the communication and rules are converted into products of prime figures. The numerous items are sorted by their counts and map to the prime numbers in reverse order. In this way, the product of prime numbers Stored in the index and tables will not be too large. Not only the communication but also the rules are representing in the form of prime numbers or foodstuffs of prime numbers. The strong rules are spread into two tables. Since the number of sensitive rules is often a large amount less than the come to of non sensitive rules, we do not maintain

ISSN 2348 – 9928
Doi:01.0401/ijaict.2014.08.04 Published on 05 (7) 2017

the susceptible policy in the transaction-rule index. With this guide the contact and no susceptible rules rewarding a sound form can be quickly identified. . The efficiency of cut-out age band depends on both the numbers of sensitive rules and items in the individual sensitive rules. If there are common items among the aware rules, the expenses in this stage can be condensed.

3.4. Implementation of Properties of Sensitive Rule Hiding Based on Constraints

The constraint to be well thought-out while implement of Properties of Sensitive Rule

- The Rule which is not to be hidden denoted by N-T-H.
- The Rule which is not to be generated denoted by N-T-G.

The constraints are carefully checked iteratively step by step and the property of Sensitive Rule thrashing is selected from the listings which satisfy the above said constraint.

In this module, we espouse Scheme 3 i.e., addition and deletion together to alter the directory, and use the constraint N-T-H and N-T-G as the guide to avoid the side effects. To consider both the numbers of hidden sensitive rules and tailored entries, we contain the association among policy in our amendment plot.

IV. RESULTS AND DISCUSSION

This is the crucial form which displays the form details of login course of action, getting threshold worth, generate rule, taking out development and thrashing method.

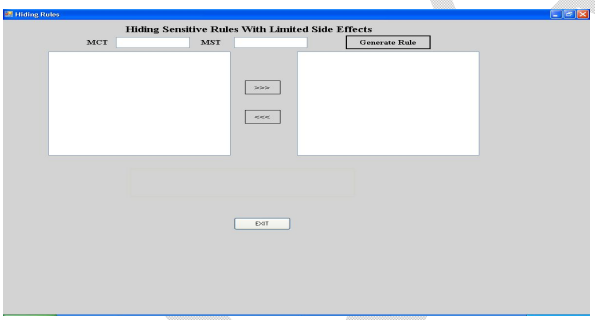

Fig 2 : Login process

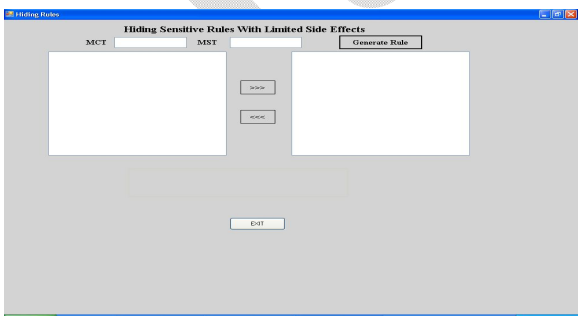

Fig 3 : Threshold value

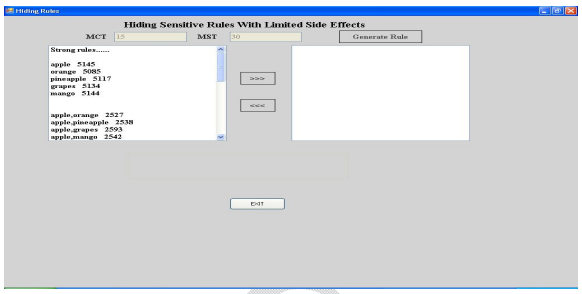

Fig 4 : Generating rule

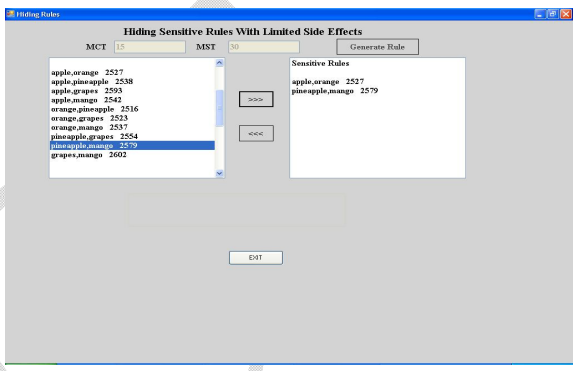

Fig 5 : Mining Process

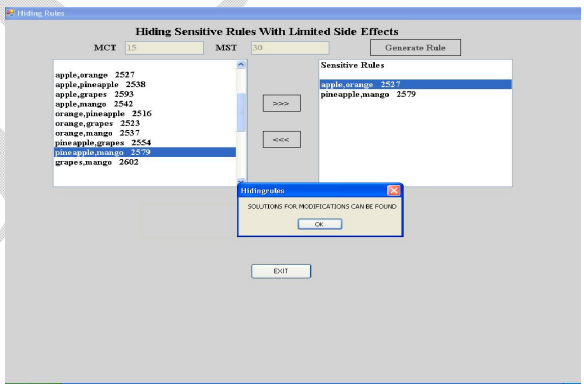

Fig 6 : Hiding Process

V. CONCLUSION AND FUTURE WORK

In this project a heuristic method for growing the digit of out of spectacle susceptible rules and reducing the number of modified entry. The untried results illustrate the effectiveness of this approach, i.e., undesired side effects are avoided in the rule thrashing course of action. The outcome also details that in most cases, all the insightful rules are hidden without specious rules misleadingly generated. Furthermore, the good quality scalability of this approach in stipulations of folder size and the pressure of the relationship in the midst of strategy on top of rule thrashing are achieved. Further a solution of susceptible rule alteration scheme will be planned.

ISSN 2348 – 9928
Doi:01.0401/ijaict.2014.08.04 Published on 05 (7) 2017

References

[1] Yi-Hung Wu, Chia-Ming Chiang, and Arbee L.P. Chen "Hiding Sensitive Association Ruleswith Limited Side Effects" IEEE transactions on knowledge and data engineering, vol. 19, no. 1, January 2007.

[2] R. Agrawal, T. Imielinski and A. Swami "Mining Association Rules between Sets of Items in Large Datasets" Proc. ACM SIGMOD '93, pp. 207-216, 1993.

[3] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," Proc. 20th Int'l Conf. Very Large Data Bases (VLDB '94), pp. 487-499, 1994.

[4] D. Agrawal and C.C. Aggarwal, "On the Design and Quantification of Privacy Preserving Data Mining Algorithms," Proc. ACM Symp. Principles of Database Systems, pp. 247-255, 2001.

[5] C. Clifton and D. Marks, "Security and Privacy Implications of Data Mining," Proc. ACM Workshop Research Issues in Data Mining and Knowledge Discovery, 1996.

[6] E. Dasseni, V.S. Verykios, A.K. Elmagarmid, and E. Bertino,"Hiding Association Rules by Using Confidence and Support,"Proc. Information Hiding Workshop, pp. 369-383, 2001.