IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 14, Issue 5 (Sep. - Oct. 2013), PP 60-64
www.iosrjournals.org

Analysis and Implementation of Efficient Association Rules using K-mean and Neuralgas Algorithm

Mohnish Patel[1], Prashant Richariya[2], Anurag Shrivastava[3]
[1](M-Tech Scholar, Computer Science Engineering, NIRT/ RGPV, Bhopal, India)
[2](Guide, Computer Science Engineering, NIRT/ RGPV, Bhopal, India)
[3](HOD, Computer Science Engineering, NIRT/ RGPV, Bhopal, India)

Abstract : *Efficient Privacy Preserving association rule mining has emerged as a latest research issue. In this thesis work, existing algorithms, Increase Support of Left Hand Side and Decrease Support of Right Hand Side are implemented successfully on the real data for Privacy Preserving Association Rule Mining. To provide privacy to sensitive data we also propose and implement a new algorithm .The performance of new algorithm is also compared with existing algorithms on the basis of number of rule pruned. The result show that proposed algorithm is more efficient as it performs privacy preserving mining by pruning more rules. Securing these against unauthorized access to the long-term goal of the database security research base community and the government statistical agencies. Whether data is personal or corporate data, data mining offers the potential to reveal what other regard as sensitive (private). In some cases, it may be of mutual benefit for two parties' even competitors to be share their data for analysis task. They would like to it will be ensure their own data remains private. In other good words, there is a need to protect sensitive knowledge during a data mining process. For Experimental work, we have used a realistic database of Doctor Patient Evaluation is taken from Medical College.*

Keywords: *Association rule, Apriori Algorithm, Clustering, k-Mean, Neural Network.*

I. INTRODUCTION

Privacy preserving association rule mining [1] [2] [15], there has some change or modification in the original database in the symbol "?" is used if the transaction does not contains the items. For example, the value in the ith position of a transaction is 1 if the transaction contains (or supports) the ith item and, the value is 0 otherwise. A "?" mark in the ith position of a transaction means that we do not have any information regarding whether the transaction contains the ith item or not. Instead of a single value for the support of an itemset A, the support interval, [minsup value (A); maxsup value (A)] where the actual support of itemset [14] A can be any value between minsup value (A) and also maxsup value (A). The minsup value (A) is the percentage of the transactions that contain 1's for all the items in A and maxsup value (A) is the percentage of the transactions that contain in either 1 mark for all the items in A. Similar value of the confidence formula, instead of a single value for the confidence of a rule A => B, there is a confidence interval [minconf(A=> B); maxconf(A => B)], where the actual confidence of a rule A => B can be any value between minconf(A=> B) and maxconf(A=> B). Given the minimum and maximum support values of itemsets AB and A, the minimum confidence value for a rule A => B is, minconf(A => B) = minsup(AB) X 100/maxsup(A), and the maximum confidence value is maxconf(A => B) = maxsup(AB) X 100/minsup(A).

In this situation when minconf(A=>B) = maxconf(A=> B), and minsup(AB) = maxsup(AB) then there are no unknown values in the database. During the sanitization process is marks, the minimum and maximum values will start to set apart, and in this case, the degree of uncertainty for the rule, will increase.

II. RELATED WORK

Kasthuri S and Meyyappan T [6] proposed a heuristic approach for hiding sensitive association rules [16]. It makes the representative rules to hide the sensitive rules. Balancing the confidentiality of the disclosed with the legitimate needs of the data user is the major challenge in association rule. They proposed an approach on the basis of modification of database transactions.

M. Mahendran etal. [7], proposed an more Efficient Heuristic approach method to hide association rule. The objective of this algorithm is to extract relevant knowledge from large amount of data, while protecting at the time sensitive information. The proposed method focused on hiding set of frequent items containing highly sensitive knowledge that only remove information from transactional database with no hiding failure.

Janakiramaiah Bonam et al [8] discuss different data restriction methods from sanitization process. They introduce the taxonomy of sanitization algorithms and validate all data restriction algorithms against real and synthetic data sets. They also considered a set of metrics to evaluate the effectiveness of the algorithms by

perform the experimental studies on different data restriction algorithms. This work concluded that SWA algorithm has an advantage over the IGA algorithm. The advantage is that SWA allows a database owner to set a specific disclosure threshold for each sensitive rule.

Cornelia Győrödi et al [9], presents a comparison between classical frequent pattern mining algorithms that use candidate set generation and test and the algorithms without candidate set generation. In order to have some experimental data to sustain this comparison a representative algorithm from both categories mentioned above was chosen (the Apriori, FP-growth and DynFP-growth algorithms). The compared algorithms are presented together with some experimental data that lead to the final conclusions. Also, the performance of the FP-growth algorithm is not influenced by the support factor, while the performance of the Apriori algorithm decreases with the support factor.

Yogendra Kumar Jain [10] proposed a new algorithm solves the problem of them. That can increase and decrease the support of the LHS and RHS item of the rule correspondingly so that more rule hide less number of modification. The efficiency of the proposed algorithm is compared with ISL algorithms and DSR algorithms using real databases, on the basis of number of rules hide, CPU time and the number of modifies entries and got better results.

Jayashree Patil, Y.C.Kulkarni [11], proposed an algorithm used to secure the sensitive items while extracting the appropriate knowledge from database. In this paper, the methodology used to preserve the privacy in data mining using association rule is used because association rule mining is one of the important aspect in data mining. In this algorithm, secure multiparty computation is used which assures security using cryptography is also discussed in brief. This algorithm ensures better privacy preserving with high efficiency. The proposed evaluation methodologies can be applied in new set of privacy preservation like cryptography-based algorithms.

Gayatri Nayak, Swagatika Devi [12], review recent work on these topics, presenting general frameworks that we use to compare and contrast different approaches. They begin with the problem of focusing on different techniques of privacy preserving. They also present and relate several important notions for this task, followed by distributed privacy preserving approaches and describe some general goals of different approaches also. Then these methods are compared and contrasted and finally we end up with conclusion and future work in succeeding sections.

The privacy-preserving data mining [21] has thus become an important issue in current years. This paper propose an evolutionary privacy-preserving data mining technology which uses data mining technique and network security cryptographic method to secure or preserver the data to find appropriate transactions to be hidden from a database.

Geetika Narang, Anjum Shaikh, Arti Sonawane, Kanchan Shegar, Madhuri Andhale [13], they reviews the major method of privacy preserving on each category and chooses some of them to complete our system. At the end, an improvement of sensitive rule hiding is proposed to make it more accurate and secured. The main approach of privacy preservation when doing association rule mining, construction a system for data mining ,by using Secure computation and TEA encryption technology is carried out. It avoids data leakage which cause by data sharing. The knowledge hiding, using ISL achieve sensitive rules hiding, and present an optimization method to get a better result.

III. MOTIVATION

3.1 ISL Algorithm:-

Assuming that the min supp 33% of min conf 70% the result of hiding item C and then item B using ISL algorithm [17] [18] is to be follows. To hide item C, the rule C = > B (50%, 75%) will be hidden if transaction T5 is to be modified according from 100 to 101 using ISL Increase Support of LHS. The rule C=>B will have support = 50% and confidence = 60%. However, rules C=>A, B=> A; B=> C cannot be hidden by ISL algorithm.

3.2 DSR Algorithm

Association rule X=>Y will be hided [19] [20] if the support of the itemset XχY is deceased or the support of Y (the right hand side of rule) is decreased. DSR algorithm Ayat Jafari, Wang, 2005 [5] decreases the support of the right hand side of the rule by modifying one item at a time in a selected transaction by changing its value from 1 to 0.

3.3 Sensitive Association Rule Hiding

Given a set of rules R extracted from the database with a certain minimum confidence and support threshold. The purpose of the Elmagarmid 2001, [3][4] rule hiding algorithms is to make the sensitive rules invisible to the association rule mining algorithms while giving as little harm as possible to the remaining non-sensitive rules to keep the data quality as high as possible. To hide a rule A => B, either decrease the support of the item set AB below the minimum support threshold, or decrease the confidence below the minimum

confidence threshold. This can be accomplished by placing "?" marks in place of the actual values to increase the uncertainty of the support and confidence of the rules (i.e., length of the support and confidence intervals).

According to the support interval and the minimum support threshold (MST), the following cases for an item set A:

- A is hidden when the minsup(A) is greater than or equal to MST,
- A is still visible when maxsup(A) is smaller than MST,
- A is visible with a degree of uncertainty when minsup(A)<=MST<= maxsup(A)

The same reasoning applies to the confidence interval and the minimum confidence threshold (MCT). It is possible for the support of a rule to be above the MST, and for the confidence to have a degree of uncertainty and vice versa. Also, both the confidence and the support may be above the threshold. From a rule hiding point of view, to hide a rule A=> B by decreasing its support, the only way is to replace 1's by "?" marks for the items in AB. In this way, change the minimum support value while the maximum support value will be the same. As replace 1's by "?" marks for the items in AB, the minimum support value of A=>B will decrease and after some point it will go below the minimum support threshold.

To hide a rule, A=> B, by decreasing its confidence and replace both 1's and 0's by the any mark. The confidence interval of A=> B is.

TID	A	B	C	D
T1	1	1	0	1
T2	0	1	0	0
T3	1	0	1	1
T4	1	1	0	0
T5	1	1	0	1

Table 1

TID	A	B	C	D
T1	1	1	0	1
T2	0	1	0	0
T3	1	0	1	1
T4	1	1	0	0
T5	1	1	0	1

Table 2

[minconf(A=> B); maxconf(A => B)] and the aim is to decrease the minconf(A =>B) below the MCT.

As minconf(A=>) B) = minsup(AB)X 100/maxsup(A), we should decrease minsup(AB) and/or increase maxsup(A). The minsup(AB) can be decreased by either placing a "?" mark in place of either A or B. If place a "?" mark in place of A then minsup(A) will also decrease, which will cause an increase in the maximum confidence value, since maxconf(A=>B)= maxsup(AB)X100/minsup(A). For rule hiding, it would be desirable to keep the maximum confidence as low as possible, and for this reason, it is better to place a any symbol mark for B. In order to increase maxsup(A), replace the 0 values for the items in a are replaced with a any mark. This process may cause an increase in the maximum support values of other rules as a side effect.

IV. MODIFY WORK

We also compare all four algorithms on the basis of number of database scans and no. of cluster. We proposed Neural Gas algorithm, which can efficiently tackle clustering of nonlinearly structured datasets. Compared with several clustering algorithms k-mean algorithm can be less sensitive to initializations due to employing the sequential learning and the neighborhood cooperation scheme. Distortion Sensitive Neural Gas algorithm is also devised to tackle imbalanced clustering issues. Experimental results demonstrate the superior performance of our K-Mean, Neural gas Cluster Algorithm and ISL, DSR Algorithm with Number of Records and cluster . We also discovered that clustering performances of the methods were dependent on the choice of the parameter. Now we are investigating a new way to adaptively determine suitable parameter values for given clustering tasks.

Result:

4.1 Comparison between K-Mean, Neuralgas Cluster Algorithm and ISL, DSR Algorithm with Number of Records and Execution Time. The Table 4 representing the Number of Records and execution time for K-Mean, Neural gas Cluster Algorithm and ISL, DSR Algorithm with and Execution Time.

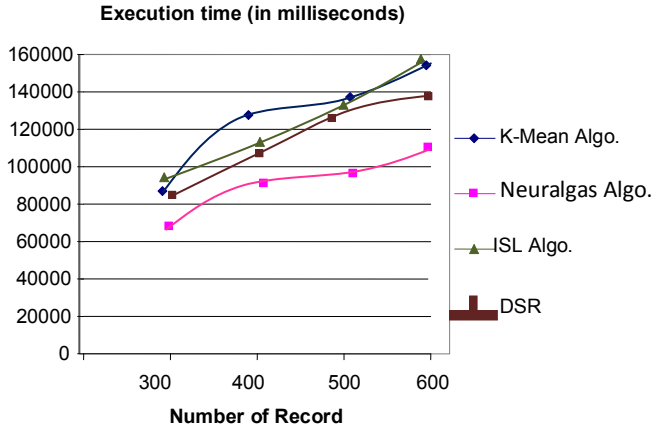
Input to Preprocessing1.cs program is Doctor Patient Evaluations Database and Output for this program is another database Filtered database of Doctor Patient Evaluations which are filter from raw data and preprocessed.

Cold	Phnemoniea	Lungcancer	Size
1	1	1	3
1	1	1	3
1	1	1	3
1	1	0	2
1	0	0	1
1	0	1	2
1	1	1	3
1	1	1	3
1	1	1	3
1	1	0	2
1	0	0	1

Table 3

Number of Record	Time taken to execute (In millisecond) K-Mean Algorithms	Time taken to execute (In millisecond) Neural gas Cluster Algorithm	Time taken to execute (In millisecond) ISL Algorithms	Time taken to execute (In millisecond) DSR Algorithms
300	86342	62636	94241	82233
400	127374	85324	112372	110302
500	136726	94247	138524	129203
600	147203	11531	154363	139562

Table 4



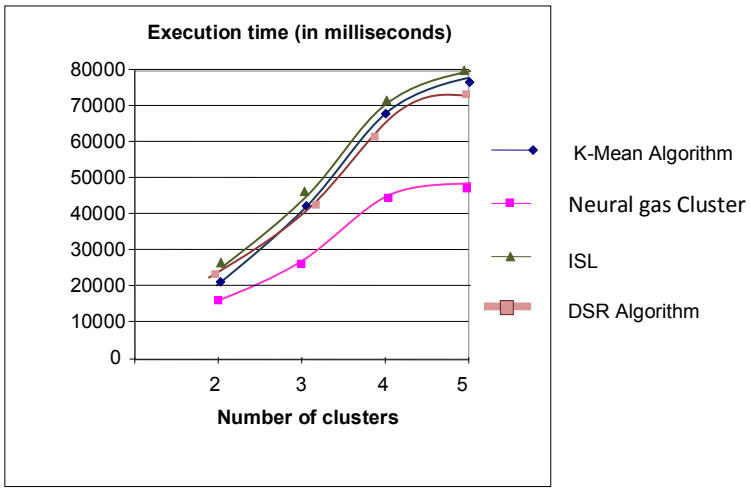
Graph 1

Above figure is shows comparison between K-mean, Neural gas and ISL, DSR algorithms. As Graph 1 show that when number of records is less, neural gas takes less time to execute than the K-mean and ISL, DSR takes More time to execute than. If the number of records is more than it is again true that neural gas Cluster Algorithm takes less time to execute than the K-mean and ISL, DSR. Further this is clear that number of records as is going to increase than more time taken to execute by K-Mean, Neural gas Cluster Algorithm and ISL, DSR Algorithm.

4.2 Comparison between K-Mean, Neuralgas Cluster Algorithm and ISL, DSR Algorithm with Number of Cluster and Execution Time. The Table 5 representing the Number of clusters and execution time for K-Mean, Neural gas Cluster Algorithm and ISL, DSR Algorithm with Number of Cluster and Execution Time, Graph 2 Represent Number of Clusters and Execution Time for K-Mean, Neural gas Cluster Algorithm and ISL, DSR Algorithm with Number of Cluster and Execution Time

Number of clusters	Time taken to execute (In millisecond) K-Mean Algorithms	Time taken to execute (In millisecond) Neural gas Cluster Algorithm	Time taken to execute (In millisecond) ISL Algorithms	Time taken to execute (In millisecond) DSR Algorithms
2	22365	15354	22470	20326
3	42301	28214	43401	40376
4	67812	45522	70010	69219
5	75343	46324	78343	73342

Table 5



Graph 2

V. CONCLUSION

The database privacy problems in data mining have been discussed and an algorithm for hiding sensitive data in association rules mining is proposed. The proposed algorithm is hybrid of two existing two algorithms ISL and DSR and K-mean, Neural Gas Clustering. An example demonstrating the proposed algorithm is given. It has been shown that the proposed algorithm is better as its hides more number of rules in same number of database scans.

REFERENCES

[1] N.V. Muthu & K. Sandhaya Rani, "Privacy Preserving association rule mining in vertically partitioned data", International journal of computer applications, Feb 2012.
[2] Fosca Giannotti, Laks V. S. Lakshmanan, Anna Monreale, Dino Pedreschi, and Hui (Wendy) Wang, "Privacy-Preserving Mining of Association Rules From Outsourced Transaction Databases", IEEE 2012.
[3] Yogendra kumar Jain, Vinod kumar yadav & Geetiks S. Pandey, "An Efficient Association Rule Hiding Algorithm for Privacy Preserving Data Mining", International Journal on Computer Science and Engineering (IJCSE), 7 July 2011.
[4] Krishnamoorthy Siva Kumar, " Spectral Filtering Technique Method", Proceedings of the Third IEEE International Conference on Data Mining, pages 40-48, 2003.
[5] Shyue-Liang Wang,Ayat Jafari, Department of Computer Science, New York Institute of Technology, New York, USA, "Hiding Sensitive predictive Association Rule", 2005.
[6] Kasthuri S and Meyyappan T, "Hiding Sensitive Association Rule Using Heuristic Approach", International Journal of Data Mining & Knowledge Management Process (IJDKP), Vol.3, No.1, January 2013.
[7] M.Mahendran, Dr.R.Sugumar, K.Anbazhagan, R.Natarajan, "An Efficient Algorithm for Privacy Preserving Data Mining Using Heuristic Approach", International Journal of Advanced Research in Computer and Communication Engineering Vol. 1, Issue 9, November 2012.
[8] Janakiramaiah Bonam, Dr.RamaMohan Reddy A, Kalyani G, " An Approach for Privacy Preserving in Association Rule Mining Using Data Restriction", International Journal of Engineering Science Invention, January 2013.
[9] Cornelia Győrödi, Robert Győrödi, Dr. Stefan Holban, "A Comparative Study of Association Rules Mining Algorithms", IJSCE, January 2013.
[10] Yogendra Kumar Jain, "An Efficient Association Rule Hiding Algorithm for Privacy Preserving Data Mining", IJSCE, July 2011.
[11] Jayashree Patil, Y.C.Kulkarni, "Association Rule for Privacy Preserving in Data Mining", IJCTA, Nov-Dec 2012.
[12] Gayatri Nayak, Swagatika Devi, "A Survey On Privacy Preserving Data Mining: Approaches And Techniques", International Journal of Engineering Science and Technology, Vol. 3 No. 3 March 2011.
[13] Prof. Geetika. Narang, Anjum Shaikh, Arti Sonawane, Kanchan Shegar, Madhuri Andhale, " Preservation Of Privacy In Mining Using Association Rule Technique", International Journal Of Scientific & Technology Research Volume 2, Issue 3, March 2013.
[14] Jnanamurthy HK, Vishesh HV, Vishruth Jain, Preetham Kumar, Radhika M. Pai, "Discovery of Maximal Frequent Item Sets using Subset Creation", IJDKP, Vol.3, No.1, January 2013.
[15] Shikha Sharma, "An Extended Method for Privacy Preserving Association Rule Mining", 2012, IJARCSSE.
[16] Mohnish Patel, Aasif Hasan & Sushil Kumar, "A Survey: Paper For Preventing Discovering Association Rules For Large Data Base", International Journal of Scientific Research in Computer Science and Engineering, Volume-1, Issue-2, May-June-2013.
[17] Ramesh Chandra Belwal, Jitendra Varshney, "Hiding Sensitive Association Rules Efficiently By Introducing New Variable Hiding counter", 2008 IEEE.
[18] Shyue-Liang Wang, Ayat Jafari, "Hiding Sensitive Predictive Association Rules", October 11, 2008, IEEE.
[19] K. SathiyaPriya, G. Sudha Sadasivam, N. Celin, "A New Method for Preserving Privacy in Quantitative Association Rules using DSR Approach with Automated Generation of Membership Function", 2011 IEEE.
[20] Suraj P. Patil, Assoc Prof T. M Patewar, "A Novel Approach for Efficient Mining and Hiding of Sensitive Association Rule", 2013 IEEE.
[21] Mohnish Patel, Prashant Richariya, Anurag Shrivastava, "Privacy Preserving Using Randomization and Encryption Methods", Scholars Journal of Engineering and Technology (SJET) Volume 1 Issue 3, Sep 2013.