© 2018 IJSRSET | Volume 4 | Issue 4 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology

Mining Negative Association Rules in Distributed Environment

Hetal Jadav[1], Kinjal Thakar[2]

[1]Research Scholar, M.E., Department of Information and Technology, Silver Oak College of Engineering and Technology, Ahmadabad, Gujarat, India

[2]Assistant Professor Department of Information and Technology, Silver Oak College of Engineering and Technology, Ahmadabad, Gujarat, India

ABSTRACT

In the data mining field, association rules are discovered having domain knowledge specified as a minimum support threshold. the more accurate the process of setting up minimum threshold, the more accurate we find association between data. The data may be positively or negatively relate to each other based on data values. Even though large in number, some data misses some interesting rules and the rules' quality necessitates further analysis. As a result, we have proposed a hybrid approach based on apriori algorithm for mining frequent item sets. This algorithm will help to discover itemsets which are negatively associated with each other. These association is found on the base of properties of propositional logic, and therefore, requires no background knowledge to generate them. The experiments show that our approach is able to identify meaningful negative association rules within a reasonable execution time. This approach has a new algorithm based on modified apriori, so that users can mine the items without domain knowledge and it can mine the items efficiently when compared to association rules.

Keywords: Data Mining, Distributed Database, Negative Association Rule Mining, K-Anonymity.

I. INTRODUCTION

The main aim of data mining technology is to explore hidden information from large databases. A Real word data coming from many fields. Many data mining techniques are exist such as association rule mining, clustering, classification, regression and so on and have wide applications in the real world for finding the useful data .

Association means looking for a relationship between variables or objects. It aims to extract interesting association, correlations or casual structures among the objects association rule mining generates positive and negative rules. Positive rules specify the presence and positive relationship between the objects whereas negative rules specify the absence and negative relationship between the objects and variables. Association rules (ARs), a branch of data mining, have been studied successfully and extensively in many application domains including market basket analysis, intrusion detection, diagnosis decisions support, and telecommunications. Traditionally, the association rule mining algorithms target the extraction of frequent features (itemsets), that is, features boasting high frequency in a transactional database. However, find the association rule in distributed environment is also a interesting research it is important to find the frequent itemset in distributed environment . Many algorithms for generating association rules have been

proposed. Some well-known algorithms are Apriori, Fp-tree, Eclat and FP-Growth, but they only do half the job, since they are algorithms for mining frequent itemsets. Another step needs to be done after to generate rules from frequent itemsets found in a database.

As pr the basepaper and other research paper we saw the not more work are performed in negative association rule. One thing is arise with this is the privacy problem in distributed environment. We try to solve the privacy issue in distributed environment. In this paper our main aim to reduce time and space complexity in distributed environment and find the negative association rule in large database.

II. METHODS AND MATERIAL

A. Apriori Algorithm

Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as candidate generation), and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found. Apriori uses breadth-first search and a Hash tree structure to count candidate item sets efficiently. It generates candidate item sets of length K from item sets of length K-1. Then it prunes the candidates which have an infrequent sub pattern.

B. Proposed System Flow Chart

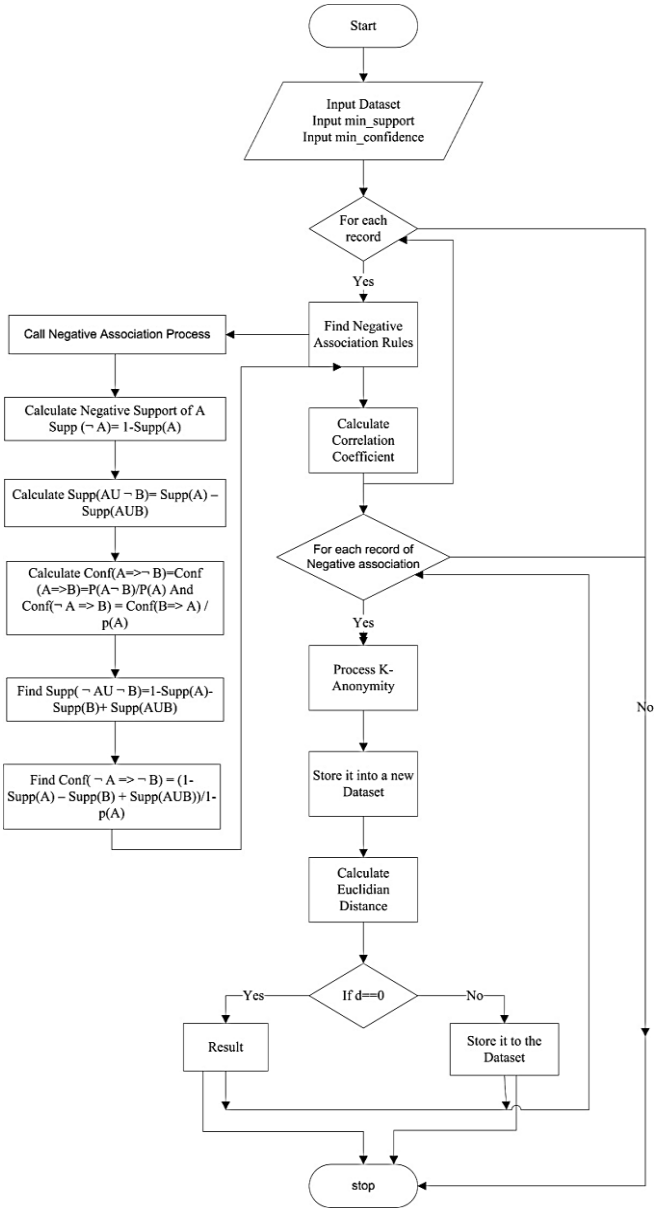

Figure: 1 Proposed System Flow

C. Proposed System Algorithm

Step 1: Input dataset,
Input min_support,
Input min_confidence.
Step 2: For each record in dataset
Step 3: Call negative association process
Step 4: To find correlation coefficient
Step 5: End For

Step 6: For each record in negative association list
Step7: Call K-Anonymity()send it to privacy protection process

Step 8: Store in the new Dataset

Step 9: Calculate euclidian distance for each row.

distance((x, y), (a, b)) = $\sqrt{(x - a)^2 + (y - b)^2}$

Step 10: If d==0

Step 11: Add it to result

Step 12: Else

Store it to the dataset

Step 13: End for

Step 14: Stop

D. Negative Association Rule(NAR) algorithm

Input: dataset

Output: Negative Association Rules

Step 1: Find relationship between of A and B

Calculate Negative Support for A: Supp (¬ A) = 1-Supp (A)

Step 2: Calculate Negative Support for:

Supp (AU ¬ B) = Supp (A) – Supp (AUB)

Step 3: Calculate Confidence of:

Conf (A=>¬ B) =Conf (A=>B)= P (A¬ B)/P (A)

And Conf (¬ A => B) = Conf (B=> A) / P (A)

Step 4: Calculate Negative Support for A and B:

Supp (¬ AU ¬ B) = 1-Supp (A)-Supp (B) + Supp (AUB)

Step5: Calculate Negative Confidence for A and B

Conf (¬ A => ¬ B) = (1-Supp (A) – Supp (B) + Supp (AUB))/1-P (A)

III. RESULTS AND DISCUSSION

The experiment uses bakery dataset. The experiment was executed by Net Beans IDE and weka tool. The language is used Java to perform the algorithm.

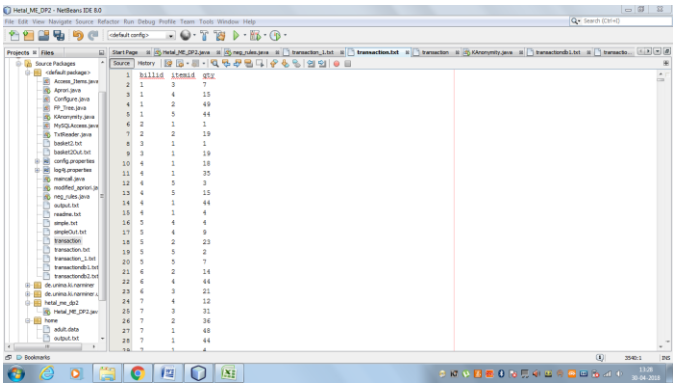

Figure: 2 Bakery Dataset

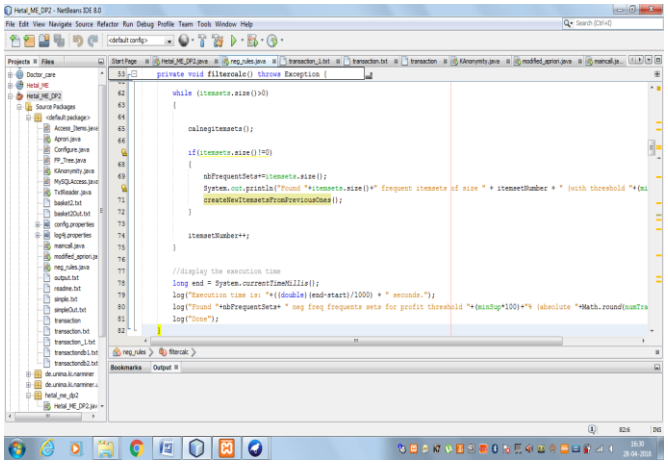

Figure: 3 Negative Association Rule code

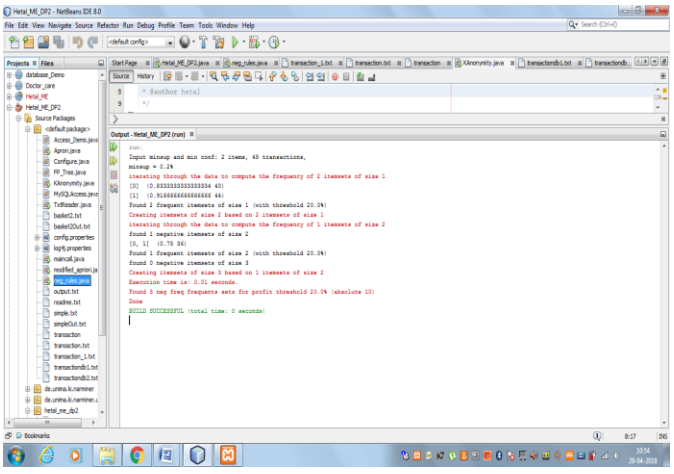

Figure:4 Negative itemsets found in bakery dataset

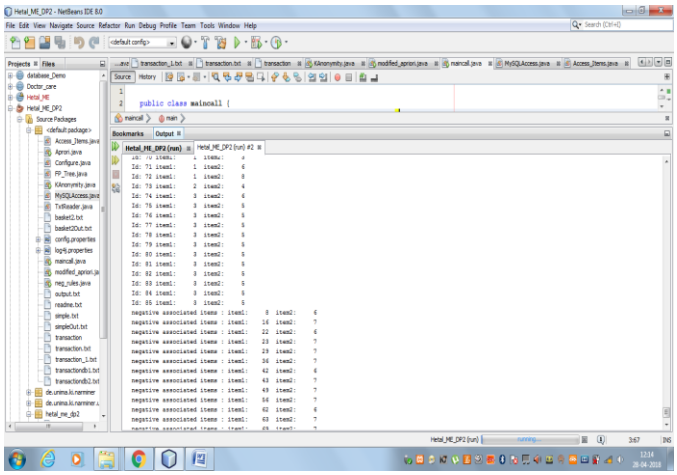

Figure 5 Itemsets combined to generate negative rules

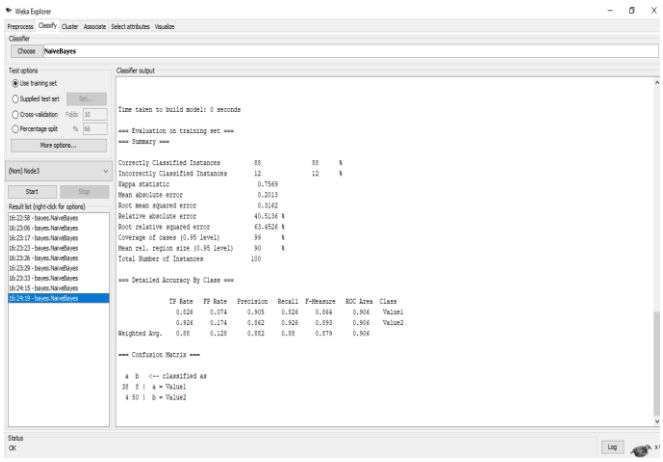

Figure: 6 weka process to classify records of bakery dataset

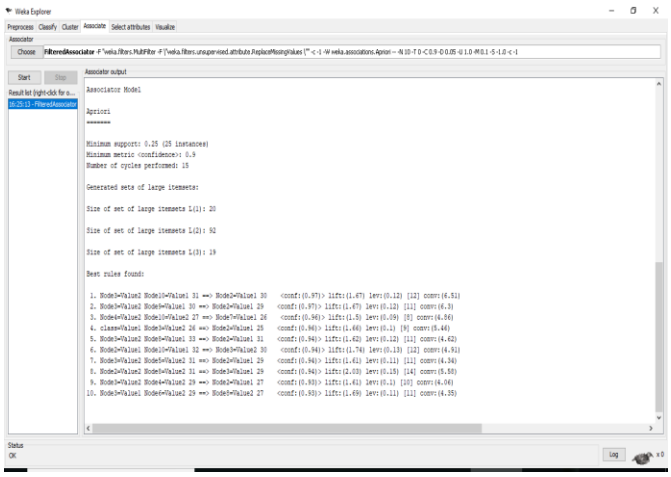

Figure: 7 weka process to display association between rules on bakery dataset

Table: 1 Result of Dataset

Dataset	Negative item rules generated
Bakery	3
Iris	54
Online transaction	127
Mushroom	9

Table: 2 Time consumed for proposed algorithm

Dataset	Time
Bakery	0.01second
Iris	0.012second
Online transaction	0.0198second
Mushroom	0.0193second

IV. CONCLUSION

We have performed algorithm and generating the negative association rules in distributed environment with privacy preservation on the Bakery dataset. And try to reduce the space and time complexity. We also perform preposed algorithm online transaction, iris and mushroom dataset and get the results. We consider the base of Apriori and than generating the result using our proposed algorithm. In future we will implement the algorithm on different dataset and get the better result in distributed environment.

V. REFERENCES

[1]. Aneesh K. Sahu, Raghvendra Kumar, Naazish Rahim : Mining Negative Association Rules in Distributed Environment 978-1-5090-0076-0/15 © 2015 IEEE DOI 10.1109/CICN.2015.183

[2]. Lulu Wanga and Hong Shenb,caSchool of Data and Computer Science, Sun Yat-Sen University, China : Improved Data Streams Classification with Fast Unsupervised Feature Selection 2016 IEEE

[3]. Ms. Sheetal Naredi Mrs. Rushali A. Deshmukh : Improved Extraction of Quantitative Rules using Best M Positive Negative Association Rules Algorithm 2015 IEEE

[4]. Prathibhamol C. P Amala G. S Malavika Kapadia : Anomaly Detection based Multi Label

Classification using Association Rule Mining (ADMLCAR) 2016 Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 21-24, 2016, Jaipur, India

[5]. Fan Fei, Student Member, IEEE, Shu Li, Student Member, IEEE, Haipeng Dai, Member, IEEE, Chunhua Hu, Member, IEEE, Wanchun Dou, Member, IEEE and Qiang Ni, Senior Member, IEEE : A K-Anonymity Based Schema for Location Privacy Preservation DOI 10.1109/TSUSC.2017.2733018, IEEE

[6]. RAKESH DUGGIRALA1, P.NARAYANA2 : Mining Positive and Negative Association Rules Using CoherentApproach ISSN: 2231-2803

[7]. Ming-Syan Chen, Jiawei Han,Yu, P.S., Data mining: an overview from database perspective, IEEE Transactions on Knowledge and Data Engineering, Vol. 8 No. 6, pp 866 –883, 1996

[8]. M Kantarcioglu and C. Clifto. Privacy-preserving distributed mining of association rules on horizontally partitioned data. In IEEE Transactions on Knowledge and Data Engineering Journal, volume 16(9), pp. 1026-1037, 2004

[9]. Xuan Canh Nguyen, Hoai Bac Le, Tung Anh Cao, "An enhanced scheme for privacy-preserving association rules mining on horizontally distributed databases". 978-1-4673-0309-5/12. IEEE 2012.

[10]. Shubhra Rana, Dr. P. anthi Thilagam, "Hierarchical Homomorphic Encryption based Privacy Preserving Distributed Association Rule Mining". IEEE 13th International Conference on Information Technology, 2014.

[11]. Ait-Mlouk Addi, Agouti Tarik, Gharnati Fatima, "Comparative Survey of Association Rule Mining Algorithms based on Multiple-criteria Decision Analysis approach".978-1-4799- 6428-4/14.IEEE,2015

[12]. Vaishali Patil, Ramesh Vasappanavara, Tushar Ghorpade Securing association rule mining with FP growth algorithm in horizontally partitioned database IEEE 978-1-4673-9084-2/16/ 2016

[13]. http://archive.ics.uci.edu/ml/index.php

[14]. Charu C Aggarwal. A survey of stream classification algorithms. Data Classification: Algorithms and Applications, page 245, 2014.

[15]. Fan Fei, Student Member, IEEE, Shu Li, Student Member, IEEE, Haipeng Dai, Member, IEEE, Chunhua Hu, Member, IEEE,Wanchun Dou, Member, IEEE and Qiang Ni, Senior Member, IEEE A K-Anonymity Based Schema for Location Privacy PreservationIEEE TRANSACTION ON SUSTAINABLE COMPUTING 2017

[16]. Rachit V. Adhvaryu, Nikunj H. Domadiya, Privacy Preserving in Association Rule Mining On Horizontally Partitioned Database. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Vol. 3, Issue 5, May 2014.

[17]. Tamir Tassa, Secure Mining of Association Rules in Horizontally Distributed Databases. IEEE Transactions on Knowledge and Data Engineering, Vol. 26, No. 4, April 2014.

[18]. Jiawei Han, Micheline Kamber, and Jian Pei. Data mining: concepts and techniques. Elsevier, 2011.

[19]. X.Wu, C. Zhang, and S.Zhang, Efficient mining of both positive and negative association rules, ACM Transaction Inf. Syst., vol. 22, no. 3, pp. 381-405, 2004.

[20]. Srikant, R. & Agrawal, R., ?Mining quantitative association rules in large relational tables?, SIGMOD Rec., ACM, 1996, 25, 1-12.

[21]. Susan Tony, Ujjwal Harode,A Comparative Study On k-means and Hierarchical Clustering, International Journal of Electronics, Electrical and Computational System (IJEECS), February 2015.

[22]. Jing Gao, Wei Fan, and Jiawei Han. On appropriate assumptions to mine data streams: Analysis and practice. In Data Mining, 2007.

ICDM 2007. Seventh IEEE International Conference on, pages 143–152. IEEE, 2007.

[23]. Shubhra Rana, Dr. P. anthi Thilagam, Hierarchical Homomorphic Encryption based Privacy Preserving Distributed Association Rule Mining. IEEE 13th International Conference on Information Technology, 2014

[24]. Jayanti Dansana, Raghvendra Kumar, Debadutta Dey, Privacy Preservation in Horizontally Partitioned Databases using Randomized Response Technique. Proceedings of IEEE Conference on Information and Cmmunication Technologes (ICT), 2013

[25]. Yongmei Liu, Yong Guan, FP-Growth Algorithm for Application in Research of Market-Basket Analysis. 978-1-4244-2874-8. IEEE 2008

[26]. Charu C Aggarwal and Philip S Yu. Locust: An online analytical processing framework for high dimensional classification of data streams. In Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on, pages 426–435. IEEE, 2008.

[27]. Mahmoud Hussein, Ashraf El-Sisi, and Nabil Ismail, Fast Cryptographic Privacy Preserving Association Rules Mining on Distributed Homogenous Data Base, I. Lovrek, R.J. Howlett, and L.C. Jain (Eds.): KES 2008, Part II,LNAI 5178, pp. 607–616, 2008.