International Journal of Computer Sciences and Engineering Open Access
Survey Paper Volume-3, Issue-9 E-ISSN: 2347-2693

A Survey on Effective Mining of Negative Association Rules from Huge Databases

E. Bala Krishna[1*], B. Rama[2] and A. Nagaraju[3]

[1*]*Dept. of CSE, JNTU Hyd, India*
[2]*Dept. of Computer Science, Kakatiya University, India*
[3]*Dept. of CSE, Rajasthan Central University, India*

www.ijcseonline.org

Received: Sep/02/2015 Revised: Sep/10/2015 Accepted: Sep/24/2015 Published: sep/30/ 2014

***Abstract*—** This paper describes about research work carried out by various authors, in which different methods are adapted to mine negative association rules. Many research surveys indicate that negative association rules are very important as positive association rules. The discovery of negative association rules are based on the interest patterns and non-interest patterns from data base (that is mining negative association rules from frequent item sets and infrequent item sets). A negative association is referred to as a negative relation between two item sets. This negative relation implies a negative rule between the two item sets.

Keywords: Negative Association Rules, Infrequent Item Sets, Non- Interest Patterns

I. INTRODUCTION

Negative association rule mining is one of the important data mining techniques to discover needed patterns in a database. In order to discover the negative association rules, first we have to identify either interest pattern or non-interest pattern (frequent or infrequent item sets) from the large data set. Next we can mine the valid negative association rules from the frequent or infrequent item sets. Mining of negative association rules are used to generate useful information for data bases ([2], [6], [7], [14], and [15]). For illustration, in transactional data base we may find out a negative association rules such as {if item A appears, then Item B & Item C will not appear}. This kind of negative association rule is useful for increasing the sales in the super market. Negative association rules help users to decide which ones are important instead of checking too many rules like positive association rules.In this paper our main focus is on reviewing work which is carried out by various researchers towards negative association rule mining. The remainder of the paper is structured as follows. Section II provides basic concepts and negative association rule mining algorithms. Section III concludes the paper besides providing directions for future work.

II. BASIC CONCEPTS & NEGATIVE ASSOCIATION RULES ALGORITHMS

Negative association rules are association rules in which either the ancestor or the consequent or both are negated. The negation of an itemset X is denoted by ~X, which means the absence of the itemset X. For example, for the rule X → Y, the negative rules are: X→ ~Y (X implies not Y), ~X → Y, ~X→~Y.

Brin S. et al. [9] developed the idea of negative relationship between the item sets for the first time, which is used to validate the independence between item sets. Chi-square test is performed on the item sets to verify the positive or negative relationship. If items were really independent the value of chi-squared statistic would be 0. If the value is higher than a cutoff value then items are not independent. Since negative association rules represents independent items.

Savasere A. et al. [1] implemented a new concept to find negative association rules. In their model, frequent item sets are merged with domain knowledge in the form of taxonomy to discover negative association rules. But, their algorithm is inflexible to generalize, since it is domain dependent and requires a pre-defined taxonomy. The following steps are required to generate negative association rules:
(a). Identify all the item sets at all levels in the taxonomy whose Support is greater than or equal to the user specified minimum_Support (b). Find the candidate negative item sets based on the item sets and the taxonomy, assign them as Expected Support. (c). Finally, count the Actual Support for the candidate item sets and keep only the negative item sets. The interesting measure called "Rule of Interest" (RI) of negative association rule X→ ~Y, as
RI= {E [supp (X U Y)] - supp (X U Y)}/ supp (X) Where E [supp (X)] is the Expected Support of an item set X. By using RI the negative association rules can be generated.

Bill P. Buckles et al [8] introduced an algorithm for mining negative association rules as follows: An negative association rule encapsulates relationship between the occurrences of one set of items with the absence of the other set of items. i.e., X → ~Y, has Support S% in the data sets, if S% of transactions in T contain item set X, while do not contain item set Y. The rule X → ~Y holds in the given data set (database) with confidence C%.

The Support of negative association rule is given by the following formulas:

$$Supp(\sim x) = 1\text{-}supp(x)$$

$$Supp(x \rightarrow \sim y) = supp(x)\text{-}supp(xUy)$$

$$Supp(\sim x \rightarrow y) = supp(y)\text{-}supp(xUy)$$

$$Supp(\sim x \rightarrow \sim y) = 1\text{-}supp(x)\text{-}supp(x)+supp(xUy)$$

The confidence of negative association rule is given by the following formulas:

$$Conf(x \rightarrow \sim y) = \{supp(x)\text{-}supp(xUy)\}/\ supp(x)$$

$$Conf(\sim x \rightarrow y)=\{supp(y)\text{-}supp(xUy)\}/(1\text{-}supp(x))$$

$$Conf(\sim x \rightarrow \sim y)=\{1supp(x)\text{-}supp(y)+supp(xUy)\}/(1\text{-}supp(x))$$

D.R. Thiruvady et al. [4] proposed an algorithm for discovering negative association rule using generalized rule discovery (GRD). GRD is an algorithm for interesting rule (k) discovery. In contrast to association rule discovery, GRD does not need the use of minimum Support condition in rule discovery process, but the user must specify a no. of rules to be generated i.e., k.

According to Nikky et al. [12] negative association rules are generated using multiple constraints and correlation factor. Multiple constraints are like knowledge base, data, measure and rule constraints. Correlation factor is used to calculate the relationship between items. The negative association rules discovery seeks rules of the three forms with their Support and confidence greater than, or equal to user-specified min_sup and min_conf thresholds respectively. These rules are referred to as an interesting negative association rule. When mining positive and negative association rule at the same time, we will find that the mining rules are contradictory frequently. For example, the rules of the forms $X \rightarrow \sim Y$ and $\sim X \rightarrow Y$ may be mined together, but the two rules are contradictory. In order to resolve these contradictions, we can judge the types of mining association rules by the correlation coefficient.

Let X and Y for the two item sets the correlation coefficient (denoted as $Q_{X, Y}$) can show the relevance of the two item sets. As follows:

$$Q_{X, Y} = \frac{Supp(xy)supp(\sim x\sim y) - supp(x\sim y)supp(\sim xy)}{Supp(xy)supp(\sim x\sim y) + supp(x\sim y)\ supp(\sim xy)}$$

The value of Yule's correlation coefficient $Q_{X, Y}$ exist the following situations:

(a) If $Q_{X, Y} > 0$ then X and Y are positive correlation. The more X occurs in a transaction the more Y will likely also occur in the same transaction and vice versa.

(b) If $Q_{X, Y} < 0$ then X and Y are negative correlation. The more X occurs in a transaction the less Y will likely also occur in the same transaction and vice versa.

By the definition of correlation coefficient, we can conclude the below lemmas:

Lemma 1: If the item set X and Y are positive correlation, then the forms of $X \rightarrow Y$ or $\sim X \rightarrow \sim Y$ will be mined.

Lemma 2: If the item set X and Y are negative correlation, then the forms of $X \rightarrow \sim Y$ or $\sim X \rightarrow Y$ will be mined.

Another interesting measure called "conviction" [17] was introduced for generating negative association rules. The conviction of a rule is defined as

$$Conviction(x \rightarrow y) = (1\text{-}supp(y))/(1\text{-}conf(x \rightarrow y))$$

Conviction $(X \rightarrow Y)$ can be interpreted as the ratio of the expected frequency that X occurs without Y. Conviction measure depends on expected frequency. But this approach generates huge amount of negative association rules which are not of interest.

Yun Sing Koh and Russel Pears [19] developed an algorithm for finding negative association rules without Support threshold by considering the complementary problem of negative association rule mining, which generates rules describing the absence of item sets from transactions. They described a new approach called MINR (Mining Interesting Negative Rules) to efficiently find all interesting negative association rules. Here they only consider the presence or absence of item sets that are strongly associated. Their approach does not require a user defined Support threshold, and is based on pruning items that occur together by coincidence. For every individual item set they will calculate two custom thresholds based on item sets Support: the positive and negative chance thresholds. They compared their implementation against Pearson Φ correlation.

Maria-Luiza Antonie and Osmar R. Zaiane [15] proposed an algorithm that extends the Support and confidence framework with a sliding correlation coefficient threshold. In addition to finding confident positive rules that have a strong coorrelation, the algorithm discovers negative association rules with strong negative correlation between the antecedents and consequents

Kundu. G et al. [20] implemented accurate classifier for mining of negative association rules, this classifier is better than decision tree classifier. Accurate classifiers have added a new direction to the ongoing research. In some cases, but in some cases the accurate classifier contains the incorrect rules. The idea is to put accurate negative rules in place of inaccurate positive association rules to discover adequate number of negative rules competently so that classification precision is improved. Associative classifier with negative rules is proficient method for classification point of view.

M. L. Antonie et al. [15] proposed a novel approach for discovering negative association rules. In his concept, if the correlations between item combinations X and Y of an item set XY, where X and Y are item sets, is negative, negative association rules are generated when their confidence is high enough. The produced rules have either the antecedent or the consequent negated: (~X → Y and X → ~Y), even if the Support is not higher than the Support threshold. However, if the correlation is positive, a positive association rule with the classical Support_confidence idea is discovered. If the Support is not sufficient, a negative association rule that negates both the antecedent and the consequent is discovered when its confidence and Support are above the minimum Support_confidence. The algorithm produces all positive and negative association rules that have a strong correlation. If no rule is found, either positive or negative, the correlation threshold is automatically lowered to ease the constraint on the strength of the correlation and the process is redone.

Rakesh et al. [10] proposed an association rule mining rule that does not require a pre-set Support threshold value. A user needs to determine a Support threshold in order to obtain only the frequent item sets. Having users to determine a Support threshold attracts a number of issues. Often, the number of association rules, even though large in number, misses some interesting rules. As a result, decision making using these rules could lead to risky actions. He also proposed a framework to discover domain knowledge report as coherent rules. Coherent rules are discovered based on the properties of propositional logic, and therefore, requires no background knowledge to generate them. From the coherent rules discovered, association rules can be derived objectively and directly without knowing the level of min_supp_threshold value required. We provide analysis of the rules compare to those discovered via the Apriori. The framework is developed based on implication of propositional logic via Negative and positive association algorithm. The experiments show that this approach is able to identify meaningful association rules within an acceptable execution time. This framework develops an algorithm based on coherent rules so that users can mine the items without domain knowledge and it can mine the items efficiently when compared to association rules.

In conventional approach finding negative association rules come across the large finding space and computational time. Li-Min Tsai et al. [13] proposed an improved approach called "generalized negative association rule" (GNAR) algorithm describe negative rules are as important as positive rules. By using GNAR algorithm to decide which rule as important rather than checking too many rules. This method mainly avoids the huge computing cost of mining negative association rules and eliminates most infrequent negative association rules.

Sandeep Singh Rawat et al. [11] developed a methodology for finding of negative association rules. In his approach an effort has been made to mine infrequent association rule with multiple min_ Supports. Discover the probability and proposed multiple min_Support based on apriori approach called "Probability Apriori Multiple Minimum Support" to capably discover rare association rules. The major proposal of this algorithm is to count the probability of each item and used to make next generation of candidate and frequent items without scan the dataset each time.

Idheba Mohammad Ali [3] proposed to find efficient positive and negative association rules but this algorithm prvduce some difficulty, such as rules invention, unrelated data set, and how to manage single minimum Support value.

Another approach [21] has proposed for mining negative association rules using binary tree concept to generate negative rules, for this purpose the conditional probability can be used. The conditional probability P (A/B) is compared with the marginal (P (A)). The marginal can be taken from Apriori algorithm.

If P (A/B) < P (A) we have a negative rule and
If P (A/B) > P (A) we have positive rule and

The measure of correlation is used to find the rules.
For the correlation: P (A ∩ B) / (P (A) P (B)).
If correlation < 1, possible to find negative rules and
If correlation > 1, possible to find positive rules.

A new concept has derived in [6] to generate both positive and negative association rules. The following steps are required, (a). Generate the set of frequent item sets of interest (P) and the set of infrequent item sets of interest (N) (b). Generate positive rules of the form X→Y in P, and negative rules of the forms X → ~Y, ~X →Y and ~X→~Y in N. By using fipi(X, Y), iipis(X, Y) and cpir(Y/X) functions negative association rules can be generated. Where fipi () means frequent item set of potential interest, iipiis () means infrequent item set of potential interest and cpif () means conditional probability increment ratio.

A method has developed in [22]. In this approach mining of positive and negative association rules as follows: (a). Find all positive frequent item sets P (FI) (b). For all item sets I in P (FI), find negative frequent item sets of the form ~(I1 I2) (c). Find all negative frequent item sets ~I1~I2 (d). Find all negative frequent item sets I1~I2 and (e). Find all valid positive and negative association rules. Authors discovered negative rules without adding additional interesting measure(s) to Support confidence frame work.

III. CONCLUSION AND FUTURE WORK

Mining of Negative association rule is an important research topic in the data mining area. Negative association rulemining is still in a stage of searching and progress. Still there are some critical issues that are needed to revise in mining negative association rules. A

few problems for mining of negative association rules are listed here:

- Frequent pattern mining a fundamental task in data mining, So much research effort is looked-for to further extend of pattern-based mining methods.
- Methods for clear understanding and study of patterns, e.g., explanation for infrequent patterns, and background revision of frequent patterns are desired.
- The majority concepts are based on some severe assumption. So we should acquire a broad view on the algorithms.
- Further new well-organized techniques for negative association mining rules should be improved.
- Single scan mining techniques could be implemented to find knowledge or patterns from data sets.
- Methods for mining negative association rules in large-databases should be explored.
- Efficient algorithms should be developed for frequent patterns and mining of negative association rules.
- The applications of mining negative association rules should be developed

REFERENCES

[1] Savasere, E. Omiecinski and S. Navathe," Mining for Strong Negative Associations in a Large Database of Customer Transactions", Proc. Intl. Conf. on Data Engineering, **1998, pp 494–502.**

[2] R. Agarwal, E. R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," Proc. Of the 20th International conference on very Large Databases,**pp.487-499**, Santiago, Chile, **1994**.

[3] Idheba Mohammad Ali O, Swesi, Azuraliza Abu Bakar and Anis Suhailis Abdul Kadir "Mining Positive and Negative Association Rules from Interesting Frequent and Infrequent Itemsets" **978-1-4673-0024/10/@26 @2012**.

[4] D.R. Thiruvady and G.I. Webb," Mining Negative Association Rules Using GRD", Proc. Pacific-Asia Conf. on Advances in Knowledge Discovery and Data Mining, **2004,** pp **161–165**.

[5] Wu X., Zhang C., Zhang S.: Mining both positive and negative association rules. In: Proc. Of ICML **(2002) 658-665**

[6] X. Wu, C. Zhang and S. Zhang," Efficient Mining of both Positive and Negative Association Rules", ACM Trans on Information Systems, vol. **22(3), 2004,** pp **381–405**.

[7] P. Yan, G. Chen, C. Cornelis, M. De Cock and E.E. Kerre," Mining Positive and Negative Fuzzy Association Rules", LNCS **3213, 2004, pp 270–276**.

[8] X. Yuan, B.P. Buckles, Z. Yuan and J. Zhang, "Mining Negative Association Rules", Proc. Seventh Intl. Symposium on Computers and Communication, Italy, **2002**, pp **623–629.**

[9] S. Brin, R. Motwani and C. Silverstein," Beyond Market Baskets: Generalizing Association Rules to Correlations", Proc. ACM SIGMOD on Management of Data, **1997, pp 265-276.**

[10] Rakesh duggirala, P.Narayan: IJCTT on "Mining Positive and Negative Association Rules Using Coherent Approach" ,**2013**

[11] Sandeep Singh Rawat and Lakshmi Rajamani "Probability Apriori based Approach to Mine Rare Association Rules" Data Mining and Optimization (DMO) **28-29 June 2011,** Selangor, Malaysia **978-1-61284-212-7/11/$26.00 ©IEEE 2011.**

[12] Nikky Suryawanshi, Susheel Jain, Anurag Jain, "A Review of Negative and Positive association rule mining with Multiple constraint and correlation factor", **2012**, **pp 778-781**

[13] Li-Min Tsai, Shu-Jing Lin, and Don-Lin Yang "Efficient Mining of Generalized Negative Association Rules" Granular Computing **978-0-7695-4161-7/10 $26.00 © IEEE 2010**.

[14] O. Daly and D. Taniar, "Exception rules mining based on negative association rules", lecture notes in computer science, vol. **3046, 2004** and **pp 543-552**.

[15] M. L. Antonie and O. R. Zaiane, "Mining Positive and Negative Association Rules: an Approach for confined rules", Proc. Intl. Conf. on Principles and Practice of Knowledge Discovery in Databases, **2004**, **pp 27-38**.

[16] Arun K Pujari "Data Mining Techniques" @ Universities press (India) private limited **2001, 2009:81-99.**

[17] Kavitha Rani et al. "Mining Negative Association Rules" International Journal of Engineering and Technology **Vol.3 (2), 2011, 100-105.**

[18] Sumalatha et al. "Mining Positive and Negative Association Rules" International Journal of Engineering and Technology **Vol.02, No.09, 2010, 2916-2920**

[19] Yun Sing Koh, Russel Pears, "Efficiently Finding Negative Association Rules With Out Support Threshold": AI **2007**, LNAI **4830, pp. 710–714, 2007.** C_Springer-Verlag Berlin Heidelberg **2007**

[20] Kudu, G. Islam, M.M; Munir, S.; Bari, M.F. "An associative classifier with negative rules" **978-1-4244-2172-5@31@2008**

[21] John Tsiligaridis, Toppenish, WA" for mining negative association rules: An Approach for Binary Trees": International Journal of New Computer Architectures and their Applications (IJNCAA) 3(3): **81-87**. The Society of Digital Information and Wireless Communications, **2013 (ISSN: 2220-9085)**

[22] Chris Cornelis, peng Yan, Xing Zhang: IEEE conference on " Mining Positive and Negative Association Rules from Large Databases", **2006**.