IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 21, Issue 1, Ser. III (Jan - Feb 2019), PP 62-66
www.iosrjournals.org

A Novel Algorithm for mining Positive and Negative Association rules

Sachin Sharma (Research Scholar)[1], Dr Shaveta Bhatia (Supervisor)[1]
[1]*Faculty of Computer Applications, Manav Rachna International Institute of Research and Studies, Faridabad*
Corresponding Author: Sachin Sharma

Abstract: *So far, severalmethods have been reflected about positive association rules based on frequent itemsets in databases. They have not measured the problem of drawing out negative rules/ itemsets. These negative itemsets also deliberate the identical itemsetsas well as negated itemsets. These negative itemsets are suitable in market-basket analysis to recognize the products that counter and go in sink. This paper presents a mathematical model to extract positive and negative association rules/ itesmsets. The experimental results show that the proposed model NAR gives better results than the previous model.*

Keywords: *Positive, Negative, Support, Confidence, Mininterest*

Date of Submission: 13-02-2019 Date of acceptance: 28-02-2019

I. Introduction

Association rules/ itemsets, a part/domain of data mining, have been deliberatedfruitfully and comprehensively in numeroussignificanceareas such as market-basket review, medical industry, education and others. But, the finding of generating the rules/ itemsets in an effectivemeans has been a mainemphasis of the data mining group/ association.

Mostly all conventional association techniques were progressed and projected to catch positive rules among data items. By referring to positive associations, we associate an association among items that occurs in transactions.Furthermore, in addition to the positive association, there is additionalsignificantkind of association rule mining which is negative association rule mining process, provide essential data to the company and help in planning future implementation. Thus, detection of possible negative rules is vital to make a consistent decision support structure. In the paper, the method of finding of positive as well as negative rules of the forms $A \Rightarrow \neg B$, $\neg A \Rightarrow B$, $\neg A \Rightarrow \neg B$, and so onwards are specified here.

Aprioriis the initialtechniqueprojected in the field of association mining and severalnewvariations were developed from it. This algorithm minespositive rules or itemsets sustaining minimum thresholds. It is acknowledged that thousands of rules have been extracted from various databases. We may get unexpected results as the support threshold is low and many more rules have been generated if support threshold is set to be high. Therefore, to mine the meaningful information, it is necessary to set support threshold low. Some different approaches were specified in the literature to remove this limitation.

II. Related Work

Following are some different approaches to produce positive and negative association rules. The objective of association mining is to generate the relationship between different itemsets.The study of association rules have been used since their introduction by [6] and Apriori is the utmosteminent association mining method.The advantage of this algorithm over other variations is its easy operation. It is the simple process. The limitation of this approach is it uses multiple scans over the database resulting the process to be time consuming. Han et al [2] developed another popular system, called FP-Growth which squeezes the data into an FP-Tree.

Brin et al [5] stated the concept of negative rules in the literature. They used the numericalor statistical test like chi-squared test. The limitation of this approach is that the values are insignificant while using this test.Dong et al. [8] presenteda new approach by using an interest value for producing negative itemsets. The approach is like Apriori for excavating negative association rules.

Sethi et.al. (2017) addressed various issues in negative association rule mining with business analytics perspectives. The authors discussed the various methods with their limitations for generating negative association rules.The limitation is correlation coefficient is complex in calculations because of the denominator.

Li-Min Tsaiet al [1] offered the GNAR (Generalized Negative Association Rule) which is an upgradedmethod, shows negative rules are as imperious as positive rules. It assistancescustomer to markfast decision to investigate which is the pre-eminent association rule. The benefit of this algorithm is cost and time saving but lack in precision and effectiveness.

Xiangjun et al. [5] suggested one more algorithm IMLMS model which made the frequent and infrequent itemsets by means of the minimum correlation coefficient. This systemstates that the interestingness of the rule is high and the minimum support value is easy to set. But lack in precision in the generation of the infrequent itemsets.

Most of the algorithms projected are anticipatedsimply for positive associations and the introduction to the fuzzy methodology has made the procedure of simplifying associations on the foundation of abs*ent item sets for generating rule dependency measures [6].

III. Proposed Approach

For mining both positive and negative itemsets/ association rules, the proposed approach first consider data. Database *D*, minimum confidence *minconf* and minimum interest *mininterest* are given in the algorithm. The resulting data of the algorithm are PR, a set of positive rules of interest, and NR, a set of negative ones.

During the running of the algorithm, Freq, a set of frequent itemsets must be generated at the *k*th pass of the algorithm, and based on them, the sets of positive and negative itemsets, namely P*k*, must be also producedindividually, where P*k* is the similar as L, frequent itemsets in the outmoded algorithms of excavatingthe association rules.

We have produced frequent itemsets. After producing all frequent itemsets, list of negative itemsets is considered by using the formula $TEMP_k$ - P_k, For trimming of uninteresting itemsets from Pkon the base of mininterest, the interest measure of each itemset is designed by using |sprt(A U B)- sprt(A)sprt(B)|.

Similarly, if |sprt(A U B)- sprt(A)sprt(B)| <mininterest for any X and Y, then it is the uninteresting frequent itemset and rest of the itemsets will be interesting.

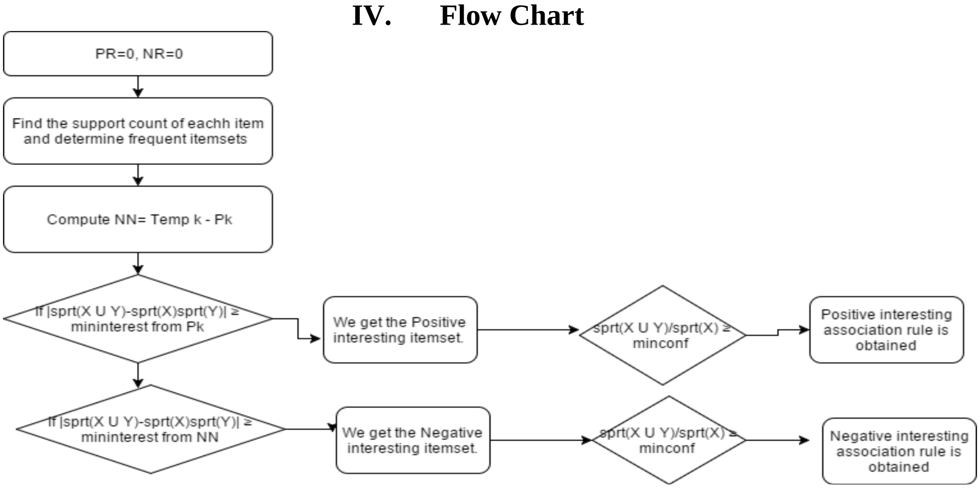
Steps: Proposed Algorithm (NAR)

a) Assume PR =null and NR=null
b) Count the support of each item.
c) Find the various frequent itemsets according to the method described in the paper [ ].
d) Calculate list of negative itemsets by using the formula

$NN= Temp_k-P_k$ where temp is list of candidate itemsets

e) On the basis of mininterest, positive uninteresting and interesting itemsets will be pruned from P_k and negative interesting and uninteresting itemsets will be pruned from NN.
f) If |sprt(X U Y)- sprt(X)sprt(Y)| ≥ mininterest, the particular itemset will be positive interesting rule otherwise it is positive uninteresting rule. In the similar way, we will check the condition from NN and get the negative interesting and uninteresting rule.
g) To reduce the space searched, a formula sprt(X U Y)/sprt(X) ≥ minconf is applied on PR and NR resulting interesting Positive and Negative itemsets.
h) Generate the possible subsets on the basis of positivs and negative itemsets and find all rules.
i) Find the confidence value of each rule and compare with minimum confidence threshold. If it is greater, it is an interesting positive / negative association rule.

IV. Flow Chart



V. Performance Evaluation

Here, the proposed method is equated with the existing versions, to discover positive and negative association rules without using multiplesupportthresholds.Toverifytheefficacyandproficiencyoftheproposed method, several experiments are conducted using various datasets with different features. In these tests, we measurethe performance with respect to time and memory space.

A. EXPERIMENTAL ENVIRONMENT AND DATASETS

We conduct the experiments by means of different kind ofdatasets to find the performance and efficiencyof the proposed method. The data sets are executed and tested on machine Intel Core 2, 2.00 Ghz with 64 bit Operating system and are implemented in MATLAB.

We used three realworld datasets (Wine and Iris). The real world datasets are taken fromFIMI repository, The important characteristicsreal world datasets are given inTable 5.

TABLE V. CHARACTERISTICS OF DATASETS

Data Set	Instances	Attributes
Wine	178	13
Iris	150	5

TABLE VI. RESULTS OBTAINED FOR WINE

Data Sets	Time Execution (in seconds)	No. of positive association rules	No. of interesting positive association rules	No. of negative association rules	No. of interesting negative association rules
178	32.67	22	2	35	1

TABLE VII. RESULTS OBTAINED FOR IRIS

Data Sets	Time Execution (in seconds)	No. of positive association rules	No. of interesting positive association rules	No. of negative association rules	No. of interesting negative association rules
150	0.2187	9	4	2	0

The execution time evaluation of proposed and various existing versions are given in Tables.The performance of NAR isexalted on the stated datasets. The experimental results divulge that proposed method is substantively faster than earlier versions.

TABLE IX. Comparison between Proposed & Existing System (Wine dataset)

	No. of interesting positive association rules	No. of interesting negative association rules
Existing [Rai N]	130	245
Existing [Swesi I]	37	52
Proposed	22	35

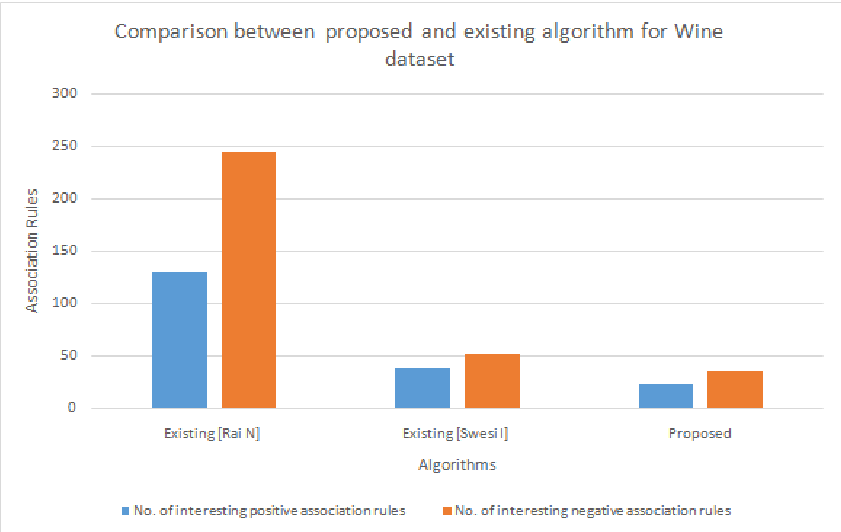


Fig 1: Comparison between Proposed and Existing (Wine)

TABLE X. Comparison between Proposed & Existing System (Iris dataset)

	No. of interesting positive association rules	No. of interesting negative association rules
Existing [Rai N]	5	18
Existing [Swesi I]	8	8
Proposed	9	2

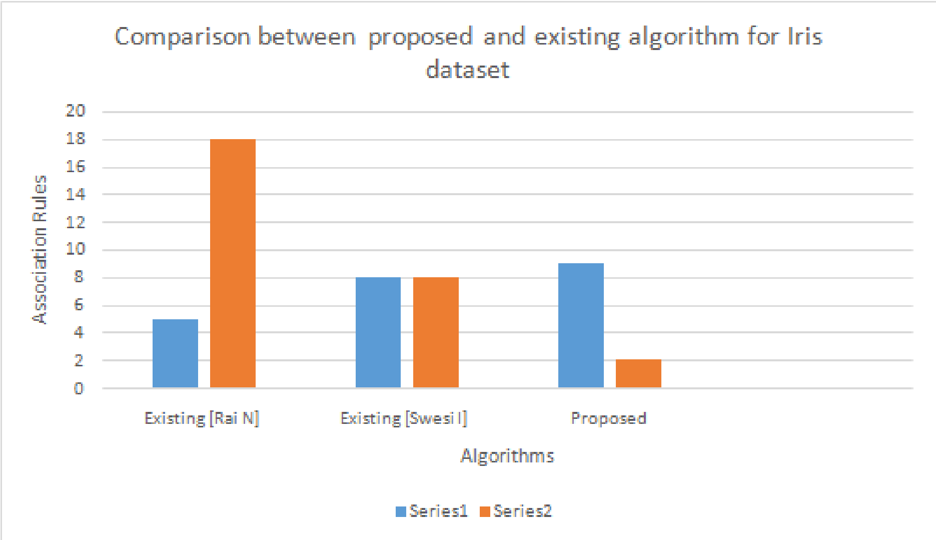


Figure 2: Comparison between Proposed and Existing (Iris)

Figure 1 and Figure 2 shows the comparison between proposed and existing algorithms. The purpose of the paper is to generate interesting positive and negative association rules form positive and negative itemsets. Here, the proposed algorithm identifies only interesting positive and negative association rules which are filtered from various association rules. From the results, 22 positive and 35 negative interesting association rules are obtained in case of wine dataset, and in the case of Iris dataset, 9 positive and 2 negative interesting association rules are obtained. These values are very less as compared to earlier versions which shows that the proposed algorithm is better than previous versions.

VI. Conclusion

In this paper, the authors have proposed a novel algorithm(NAR) that captures interesting positive and negative association rules from frequent and infrequent itemsets both. There may be various versions of existing approach showing the negative itemsets but it was not sure from the results that which itemsets are interesting or not. In the NAR, the authors have concentrated on identification of interesting positive and negative association rules only. The experimental results shows that proposed system is much better than Existing versions. With this approach, the user is able to determine quality based decision.

References

[1]. Li-Min Tsai, Shu-Jing Lin and Don-Lin Yang, "Effective Mining of Generalized Negative Association Rules", *Proceedings of IEEE Conference on Granular Computing*, pp. 113-117, 2010.
[2]. J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," ACM SIGMOD Record,vol.29,no.2,pp. 1–12, 2000.
[3]. Li Min Tsai, Shu Jing Lin and Don Lin Yang, "Effective Mining of Generalized Negative Association Rules", *Proceedings of International Conference on Granular Computing*, pp. 163-167, 2010.
[4]. M. Delgado, M. D. Ruiz, D. S´anchez, and J. M. Serrano, "A formal model for mining fuzzy rules using the *RL* representation theory," *Information Sciences*, vol. 181, no. 23, pp. 5194–5213, 2011.
[5]. Brin,S., Motwani,R. and Silverstein,C., " *Beyond Market* Baskets: Generalizing Association Rules to *Correlations*," Proc. ACM SIGMOD Conf., pp.265-276, May 1997.
[6]. R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," ACM SIGMOD Record,vol.22,no.1,pp.207–216,1993.
[7]. M.L. Antonie and O. Zaiane, "Mining positive and negative association rules: an approach for confined rules," in Knowledge Discovery in Databases: PKDD 2004,vol.3202, pp.27–38, Springer, 2004.
[8]. W. U. X. Dong, Z. C. Qi, and S. C. Zhang, "Mining both positive and negative association rules," in Proceedings of the 19th International Conference on Machine Learning (ICML '2002), pp. 658–665, 2002.

[9]. Mahmood S., Shahbaz M. and Guergachi A,"Negative and positive association rules Mining from text using frequent and infrequent itemsets", Hindawi Publishing Corporation, Volume 2014
[10]. Parmar K, Shukla R,"Enhanced data using positive negative association mining", IJARIIE, Volume 2, Issue 3, 2016
[11]. Rai N., Jain S. & Jain A. (2014),'Mining positive and negative association rule based on improved genetic algorithm', International Journal of advanced computer science and applications, Volume 5, Issue 1, pp 160-165

IOSR Journal of Computer Engineering (IOSR-JCE) is UGC approved Journal with Sl. No. 5019, Journal no. 49102.

Sachin Sharma. " A Novel Algorithm for mining Positive and Negative Association rules " IOSR Journal of Computer Engineering (IOSR-JCE) 21.1 (2019): 62-66.