An Analysis on Characteristics of Negative Association Rules

Mohammad Javad Kargar*, Fatemeh Hajilou
Department of Computer Engineering,
University of Science and Culture,
Tehran, Iran
kargar@usc.ac.ir, fatemeh.hajilou@yahoo.com

Received:2019/12/25 Revised:2020/01/28 Accepted:2020/02/24

Abstract- Association rules are one of the data and web mining techniques which aim to discover the frequent patterns among itemsets in a transactional database. Frequent patterns and correlation between itemsets in datasets and databases are extracted by these interesting rules. The association rules are positive or negative, and each has its own specific characteristics and definitions. The mentioned algorithms of the discovery of association rules are always facing challenges, including the extraction of only positive rules, while negative rules in databases are also important for a manager's decision making. Also, the threshold level for support and confidence criteria is always manual with trial and error by the user and the proper place or the characteristics of datasets is not clear for these rules.

This research analyses the behavior of the negative association rules based on trial and error. After analyzing the available algorithms, the most efficient algorithm is implemented and then the negative rules are extracted. This test repeats on several standard datasets to evaluate the behavior of the negative rules. The analyses of the achieved outputs reveal that some of the interesting patterns are detected by the negative rules, while the positive rules could not detect such helpful rules. This study emphasizes that extracting only positive rules for covering association rules is not enough.

Keywords— Negative Association Rules, Data Mining, Frequent Itemsets, Association Rule Mining.

1. Introduction

There are lots of content about association rules, while most of them only focused on the positive association rules. Even many of literature did not pointed out that there are two types of association rules which are positive and negative. The negative association rules are beneficial for analyzing the web log files, for finding the strong association between the web data's and discover the useful patterns in web world [1].

Agrawal [2] worked on association rules and for explaining the issue of the discovery of these rules in database used the sales transaction. Despite the importance of these types of rules in business, and presenting the main and unusual exceptions and rules, but still do not pay particular attention to negative association rules. Association rules represent logical and common information, while negative rules and exceptions are unusual and surprising facts in the data[3].

According to the conducted studies, it seems applicable and efficient to suggest a method by using combination of examined algorithms and finding hidden information in these types of rules, such as: the proper dataset to apply these rules or certain characteristics that the data sets have for negative rules. Since the discovery of these negative rules cost more than positive rules in terms of time, one of the most important goals that is considered is the data mining of negative association rules that can be used to state more precisely that under what conditions and with what characteristics negative association rules have better function and whether there is a particular pattern in them.

The association rules are divided into two groups, positive and negative, and each has its own specific characteristics and definitions[4]. In general, negative rules can be achieved by discovering positive rules, definitions, and formulas. In positive association rules the items are either positive or negative on both sides of these rules, and are defined using the following two criteria (Equation 1):

$$Support(A \rightarrow B) = P(A \cup B)$$
$$Confidence(A \rightarrow B) = P(B|A) \quad (1)$$

But in negative association rules, at least one item is negative and they are defined as follow (Equation 2):

$$Support(A \rightarrow \sim B) = Support(A) - Support(A \cup B)$$
$$Confidence(A \rightarrow \sim B) = \frac{Support(A \rightarrow \sim B)}{Support(A)}$$

$$Support(\sim A \to B) = Support(B) - Support(A \cup B)$$
$$Confidence(\sim A \to B) = \frac{Support(\sim A \to B)}{1 - Support(A)}$$
$$Support(\sim A \to \sim B) = 1 - Support(A) - Support(B) + Support(A \cup B)$$
$$Confidence(\sim A \to \sim B) = \frac{Support(\sim A \to \sim B)}{1 - Support(A)} \quad (2)$$

Absence of item sets cannot be programmed and even if it is, that leads to generation of millions of negative rules that may not be of use to the manager. Hence, the objective is to find only "interesting" negative association rules that can be acted upon by the managers[5]. There are other criteria that make these rules valuable and important and merely measure their correlation, interest, and validity, such as the Correlation, Coefficient, Lift.

2 . LITERATURE REVIEW

In the field of negative association rules various works have been conducted so far, and to extract negative association rules various algorithms have been presented by different individuals. Regarding to the analysis of the proposed algorithms related to negative association rules, they can be divided based on their function as follow:

1) Negative association rules algorithms based on the reduction of search space
2) Negative association rules algorithms based on evaluation and interest criteria
3) Combinational negative association rules algorithms based on space reduction and extraction of interest rules.

In the following, these types of negative association rules algorithms will be examined and introduced.

- **Negative association rules algorithms based on the reduction of search space:**

Paul [6] improved the PNAR algorithm to discover negative association rules by using correlation coefficient and dual confidence criteria. He finds the relationship between the items and positive and negative variables relative to each other by using the correlation coefficient, and reduces the search space for negative rules by using the dual confidence and therefore has a higher speed for discovering the rules. The results of applying this method show many rules will be generated that are not necessarily appropriate if confidence value is low. On the other hand, if confidence value is high, the number of rules that are negative on both sides will be unnecessary generated and positive rules may be mistakenly removed, therefore, using dual confidence mode is very appropriate for this method.

Agrawel [2] proposed an approach based on PSO set (particle swarm optimization) in which the Saric algorithm is used to generate negative association rules and has two stages: the first phase of pre-processing and the second phase of data mining.

The first phase involves binary conversion to measure the effective value of the items in a way that the data is converted into binary form, and on the other hand, discovers negative rules using the two criteria of the item range (IR) to determine the search scope and the correlation calculations. The second phase decodes the rules using front and back pointers, then generates the best population, and finally selects the best rule using correlation coefficient.

Cokpinar [7] was looking for a way to discover negative rules in XML flow data. After a while, this method was discussed more precisely and more seriously by providing an algorithm called PNRMXS by Kavi [8]. Its function is in such a way that it maps the relevant codes and uses these mappings and performs data mining and discovers the rules almost based on the advanced FP-Tree method. The most important thing in this method is that the search space of the rules be reduced correctly.

The frequent pattern growth (FP-Growth) approach is one of numerous algorithms and methods to discover and find out negative and positive association rules. FP-Growth is well-organized and capable method for finding the item sets which are frequent, without the generation of candidate item sets. The drawback of FP-Growth is it discovers a huge amount of conditional FP-Tree. Krishna et al.[9] proposed improved FP-Tree for extracting negative association rules without generating conditional FP-Tree. Chen[10] presented another classic algorithm FP-Tree algorithm and solved the Apriori repeatedly scanning database and produce the problem such as a large number of candidate sets.

Rai et al. [11]was looking for a method called MIPNAR_GA to extract negative association rules based on the genetic algorithm and its characteristics. The use of genetic algorithms and the generation of appropriate populations to reduce the search space, which certain steps are respectively conducted based on mutations and chromosomes in the genetic algorithm, is what that is conducted in this algorithm.

- **Negative association rules algorithms based on evaluation and interest criteria:**

In a tourism system [12], negative association rules have been used to facilitate information access. These evaluation criteria are: interest (Lever) and Correlation level (CPIR), in the above system, first the items and their relevant characteristics are generated, and then their positivity or negativity is discovered based on the leverage and use of the support system. At the end of the work and generating all the rules using the CPIR calculation, it is found that the extracted rules are positive or negative.

An algorithm based on knowledge-based space-time systems [13] and the combination of negative rules at first provides a specific definition of this category of rules, and considers negative rules as follow (Equation 3):

$$A \to \sim B \qquad (3)$$

a) A&B = Ø no common element

b) Supp(A) > ms; Supp(B) > ms; Supp(AUB) < ms

c) Supp(A →~ B) = Supp(AU~B)

d) Conf(A→~B)= Supp(AU~B)/Supp(A)

In this method, a criterion is used for their interest after extracting the rules, which in general is as follows (Equation 4):

$$Interest(X,Y) = |Supp(X \cup Y) - Supp(x)\ Supp(Y)| \qquad (4)$$

Multi-objective evolutionary algorithm [14]for the set of positive and negative quantitative and interest rules (MOPNAR), first uses a criterion for evaluating rules which is based on support to reduce the item set of positive and negative rules and select the set properly, and shows the relationship between values that its value is defined in [-1,1] distance, and is defined as (Equation 5):

$$netconf(X \to Y) = \frac{SUP(XY) - SUP(X)SUP(Y)}{SUP(X)(1 - SUP(X))} \quad (5)$$

In this method, in addition to using criteria such as conviction (for calculating the correlation between rules) and Lift (for comparing and calculating confidence and measuring interest among rules), a more important criterion called CF is used that this shows the probable variations of the item Y, provided the presence of item X, and actually represents the correlation relationship (positivity and negativity) of the rules.

Since the MOPNAR algorithm has high complexity and focus on positive rules rather than negative rules, Naredi [15] proposed BMPNAR algorithm i.e. Best M Positive Negative Association Rules Algorithm. This algorithm is an extension to MOPNAR algorithm. They expect the algorithm to produce M best rules with better space and time efifciency. It is a combination of MOPNAR and TOP□ k Algorithm to produce a new algorithm.

- **Negative association rules algorithms based on the presence or absence of items**

Ravi [16] and Kotiyal et.al [17] obtained an appropriate algorithm for items with finite number called k-map, and first converted transactions into binary forms using the rational methods, and put them in an accuracy chart and then drew the relevant karnaugh map according to the number of items and then extracts the rules using the table categorization. Actually, in this method, first, the items and their presence or absence in transactions are converted to binary, and their accuracy chart is formed.

The algorithm for discovering negative rules based on the presence or absence of frequent items was one of the best methods introduced by Kadir [18] . The algorithm presented by this method is based on the algorithm and Apriori’s candidates and discovers both positive and negative rules. This algorithm offers items that are not actually present (absence itemsets) and their support levels are higher than minsupport.

One of the uses of negative association rules was introducing negative association rules to the field of query expansion, and puts forward new models of query expansion by Liu [19]. This algorithm can produce frequent and infrequent feature terms according to the vector inner product, and get positive and negative association rules between terms.

According to the records and examination of various algorithms for extracting negative rules, most of them have two problems of quality and interest of the obtained negative rules and a large search space for these rules. Therefore, this study seeks to discover the characteristics and the uses of these negative rules to overcome these problems and in addition, achieve the best rules by introducing and combining different assessment criteria.

In the end, Table 1 compares the related works in the scope of negative association rules.

3 . METHODOLOGY

Contrary to positive association rules that there are many pseudo code and executable algorithms of them, we encounter the problem related to the execution of most of the pseudo-code or introduced algorithms in examining and testing negative association rules algorithms. Hence, in order to determine which one is practically applicable and appropriate for testing, first it was necessary that the proposed algorithms be implemented in a form of trial and error. R-language and SPSS were used for implementing algorithms and statistical analysis, respectively.

One of the best methods of finding negative rules was an algorithm designed for the first time in late 2011 by Kadir et al. [18] which is still the basis of many other methods. Although in this method execution

TABLE 1. COMPARING PREVIOUS WORKS IN THE FIELD OF NEGATIVE ASSOCIATION RULES

Criteria	Shortcomings	Advantageous	Algorithm Name	Reference Number	Year
Correlation coefficient confidence dual	Some complexity in the pruning method	Better precision and speed	PNAR	[6]	2016
Correlation coefficient and partitioning	Complexity of partitions and their decoding	Use the precision criteria to select the item range	Set- Pso	[2]	2015
Binary conversion and accuracy chart	Item Limit, Complexity of karnaugh map	Proper precision for less items	K-map	[16, 17]	2014
Interest criterion	dependency on data conditions	Simplicity in defining negative rules	Spatio-temporal	[13]	2013
Criteria Lever & CPIR	Sup scans and calculations for each item	categorizing the final rules	Tourism system	[12]	2012
Covariance and standard deviation	Long execution time	Getting the proper space to search rules	XML-data	[7, 8]	2014
Conviction criteria and lift	High complexity in calculations	Select the correct item set	MOPNAR	[14]	2014
Correlation coefficient and frequency and infrequency criteria of items	Long calculations for population generation	Generating rules based on genetic	MIPNAR_GA	[11]	2014
Interest and Collective Strength and Piatetsky Shapiro	Generating rules and a large search space	Discovery of rules based on the absence of items	Frequent item set	[18]	2011

and extraction time of positive and negative rules is longer, but results are more interesting and more efficient. Thus, this research used this algorithm as basic algorithm and then and attempt to modify the algorithm for extracting optimal negative rules.

- **Algorithm Implementation**

As mentioned above, the introduced algorithms must first be modified and then tested. To implement the algorithm, the transactions or any other data type must first be defined after installing the packages in R. Then, the data will be converted to a transactional list. At this stage, the minimum support value must be first determined, and then it is need to find the items in the transactions. Fig.1 shows the pseudo code of algorithm.

After completing the above steps, frequent 1 item and 2 items and so on should be extracted. Fig.2 shows the pseudo code.

Finding frequent 1-items L1: In this method, the number of transactions that have the intended item are counted with countP and the number of transactions that do not have the intended item are counted with

```
l←Ø, f←Ø
scan the database and find the set of items (f)
for all items f in F do
    for all transaction t ∈ D do
    if t contains f then
        countP++;
    else
        countN++;
if countP≥ms or countN≥ms
    l1←l1∪{f}
return l1
```

Fig.1. Pseudo code of the basic algorithm extracted and modified from [18].

countN. According to the L1 algorithm, the items enter that the value of countP for them are larger than ms. It is also the same for negative items. Each item as "1 C" is a positive item and as "0 D" is a negative item.

In the following, it is required that the Lk sets be produced for values k = 2,3, ... All selections of a k number set of 1-items are found by combination instruction, and the items in transaction 1 are counted

```
for (k = 2; l_{k-1}≠Ø; k++) do
    C_k= l_{k-1} ⋈ l_1
    foreach i ∈ C_k
      for each item in i
        if item is negative item then
          if item not contains in t then
            count++;
        else
          count++;
      if count≥ms
        l_k←l_k∪{i}
return l_k
```

Fig.2. Pseudo Code for Extracting Frequent K-items

with the grep instruction, and stored in the cP matrix. In the matrix, 1 shows the presence of the item and 0 is the absence of that item.

According to the results of the cP matrix, the number of transaction that contains each set of cP is counted. Then the sets that are count> = ms are selected as Lk.

In the end, the negative and positive rules are selected. Each of the Lk sets are divided into two sub-sets Aj and Bj, and the support values of the sets Aj, Bj and AjBj are calculated, and then the confidence is calculated. If ConfAB> = mf, it will be considered as a rule. The right-hand side of the rule is saved on Rt and the left-hand side rules in Lt. The achieved rules are as follow:

> Lt items of the right hand side of the rules

[[1]]

[1] "0 H"

[[1]]	**~H → C**
[[35]]	**{E,G} → ~H**

> Rt items of the left hand side of the rules

[[1]]

[1] "1 C"

This algorithm can extract all possible negative rules according to the type and size of the dataset. It is enough to change support and confidence values to see a significant change in the number of rules.

5 . Findings

Negative rules extraction algorithms were tested on several datasets in order to monitor behavior of this algorithm on different datasets. R programming language and the Arules package have been used to implement this algorithm. For first round of the test, the transaction data in Table 2 were used.

Positive and negative item sets with 1, 2, and 3 items is given in Table 3. The min support value to extract this item sets is 0.35.

According to the results, the number of 1-item sets is the same for positive and negative items. For 2-item sets, only one positive set {C, I} is obtained and the other sets have negative item. Also for 3- item sets, all sets have negative item. Thus, the considered dataset to examine positive and negative rules seems appropriate. 1-, 2-, and 3-item set with the highest frequency are shown in Table 3.

By implementing negative rules algorithm on the above dataset, negative and positive rules are reported in Table 4. The results show that most of the obtained rules are negative.

The rules with a maximum item length of 3 have been achieved and the rules containing 2 items are more likely to occur and have higher values of confidence and support.

- **Testing Different Datasets by the Negative Association Rule algorithm**

First, negative and positive association rules algorithms have been investigated for five different datasets and the results are shown in Table 5. According

TABLE 2. TRANSACTION DATA

TID	Transaction
1	G,A,C,F,I,D
2	B,H,J,C
3	G,I,E,C,D
4	H,J,B,F,A
5	I,A,C,D
6	E,J,I,G
7	G,H,F,B,D
8	J,F,C,B
9	I,E,C,B
10	J,F,D,H,A
11	J,E,B,G,A
12	I,F,D,H,B
13	F,G,E,B,C,I
14	D,J,A,H,B
15	G,A,C,E,F,I
16	H,F,I,C,B,D
17	A,F,J,C,H,E
18	E,G,C,D,A,B
19	I,J,C,H,A,B
20	E,C,I,A,G,D,J

TABLE 3. ITEM SETS WITH THE HIGHEST FREQUENCY

	Frequent 1 item sets		Frequent 2 item sets		Frequent 3 item sets	
1	C	0.65	B, ~ E	0.4	B, ~ E, ~ G	0.35
2	B	0.6	B, ~ G	0.4	B, ~ E, H	0.35
3	I	0.55	C, ~ H	0.45	C, ~ H, I	0.35
4	D	0.5	C, I	0.45	C, ~ H, ~ J	0.35
5	~ D	0.5	C, ~ J	0.4	C, I, ~ J	0.35
6	F	0.5	~ E, ~ G	0.45	~ E, ~ G, H	0.35
7	~ F	0.5	~ E, H	0.4	E, G, ~ H	0.35
8	J	0.5	E, ~ H	0.4		
9	~ J	0.5	G, ~ H	0.4		
10	~ G	0.55	~ G, H	0.4		
11	~ E	0.55	~ H, I	0.4		
12	~ H	0.55	I, ~ J	0.4		

to the low volume of data, rules with a maximum item length of 3 are extracted in these data. Table 5 shows the number of extracted rules along with their support and confidence values. Table 6 shows statistics of the extracted negative and positive rules separately.

According to the results from Table 5 and Table 6, sports datasets and data of purchasing computer have a higher support value for negative rules. Whereas Test dataset in positive association rules section have a very high support value. By comparing these results with the results from the Apriori algorithm, that are presented in the following, it is found that Test dataset has more positive rules with the same support and confidence value than other dataset, while this dataset is not compatible with negative association rules. The achieved results in this experiment indicate somewhat that negative association rules are not compatible with binary datasets.

Table 8 shows number of extracted negative and positives rues separately.

Support and confidence value of the rules have been calculated in order to make better analysis of the achieved results to obtain the characteristics of negative association rules. Table 9 shows these values.

Fig.3 shows summary of the results along with confidence and support of extracted positive and negative rules on 9 Datasets.

6. Discussion and Findings

According to Figure 3 that examined and tested different datasets in many ways such as the type of recognized rules, data type, data size, and so on, the following interpretations can be presented:

1) Considering that in some datasets only negative rules are extracted, the necessity of using negative association algorithm is more clear and important than before because in some databases the extracting only positive association rules does not result and leads to miss some o interesting patterns in dataset. Sometimes missing the interesting and significant patterns may have financial or even human losses.

2) The results suggest that although there is not a lot of difference between support and confidence in a positive and negative state in some datasets, but in some others such as sport data, computer purchasing, or adults, this amount is more important and considerable. Moreover, negative rules generated from these datasets have higher confidence than the same positive rules. This fact indicate that sometimes negative rules are more beneficial for our business.

3) Finally, the results indicates the characteristics of datasets influence effectiveness of negative association rules.

According to the achieved results and the tests conducted on different datasets in this study, it can be said that these negative association rules are not at fully compatible at least with binary datasets. The results of the tests on both small and large datasets shows negative association rules could not extract patterns on binary datasets as well in compare with positive association rules. However with the scale of this research cannot prove negative association rules have weak performance on binary datasets, but this can be an alarm to investigate this issue on a bigger list of datasets. Moreover, the results of this research shows characteristics of datasets influence effectiveness of negative association rules. It seems that the discovery of completely hidden patterns and the difference between these types of rules and the positive extracted rules is one of the factors that affects the success of the negative association rules because these rules are completely opposite and contrary to everyday beliefs and behaviors.

TABLE 4: THE OBTAINED RULES FROM NEGATIVE ASSOCIATION RULES ALGORITHM

Rule			Support	Confidence
~E	→	B	0.4	0.727273
~G	→	B	0.4	0.727273
~H	→	C	0.45	0.818182
I	→	C	0.45	0.818182
~J	→	C	0.4	0.8
~E	→	~G	0.45	0.818182
~G	→	~E	0.45	0.818182
~E	→	H	0.4	0.727273
~H	→	E	0.4	0.727273
~H	→	G	0.4	0.727273
~G	→	H	0.4	0.727273
~H	→	I	0.4	0.727273
I	→	~H	0.4	0.727273
I	→	~J	0.4	0.727273
~J	→	I	0.4	0.8
~H~J	→	C	0.35	1
C~J	→	~H	0.35	0.875
C~H	→	~J	0.35	0.777778
~J	→	C~H	0.35	0.7
C~J	→	I	0.35	0.875
CI	→	~J	0.35	0.777778
~J	→	CI	0.35	0.7
I~J	→	C	0.35	0.875
C~J	→	I	0.35	0.875
CI	→	~J	0.35	0.777778
~J	→	CI	0.35	0.7
I~J	→	C	0.35	0.875
~EH	→	~G	0.35	0.875
~GH	→	~E	0.35	0.875
~E~G	→	H	0.35	0.777778
~GH	→	~E	0.35	0.875
E~H	→	G	0.35	0.875
EG	→	~H	0.35	1

7. Conclusion

The aim of this research was to analysis characteristics of negative association rules. For this, a base and efficient negative algorithm was selected, modified and implemented. Then, this negative rules extraction algorithm was examined on different datasets with different characteristic such as: size and type of variables and data. The results of tests revealed that type of datasets and their characteristics influence effectiveness of the negative association rule algorithm. The results showed the negative association rules are not at fully compatible at least with binary datasets. On the other hand, the results indicate the negative association rules are fully appropriate and compatible with the transactional data such as store shopping basket. Considering that in some datasets only negative rules are extracted, the necessity of using negative association algorithm is inevitable.

TABLE 5. RESULTS OF EXPERIMENT ON FIVE DIFFERENT DATASETS

Dataset	Rules	Number	Support	Confidence
Transaction	2 item	28	0.383929	0.732792
	3 item	46	0.35	0.850725
	Total	75	0.362838	0.806102
Sport	2 item	11	0.583333	0.864646
	3 item	11	0.571429	0.89899
	Total	23	0.577381	0.881818
Computer	2 item	35	0.497959	0.819965
	3 item	45	0.431746	0.84515
	Total	80	0.577381	0.881818
Cheating	2 item	3	0.8	0.925
	3 item			
	Total	106	0.8	0.899371
Test	2 item	12	0.99	0.99
	3 item	24	0.99	0.99
	Total	36	0.99	0.99

TABLE 6. CATEGORIZING DIFFERENT POSITIVE AND NEGATIVE RULES ON THE 5 DATASETS

Dataset	A → B	A → ~B	~A→ B	~A →~B
Transaction	1	16	20	4
Sport	0	5	5	6
Computer	0	5	5	7
Cheating	0	0	0	3
Test	36	0	0	0

TABLE 7. RESULTS OF EXPERIMENT ON 9 FINAL DATASETS

Dataset	Rules	Number	Support	Confidence
Transaction	2 item	28	0.383929	0.732792
	3 item	46	0.35	0.850725
	Total	75	0.362838	0.806102
Sport	2 item	11	0.583333	0.864646
	3 item	11	0.571429	0.89899
	Total	23	0.577381	0.881818
Computer	2 item	35	0.497959	0.819965
	3 item	45	0.431746	0.84515
	Total	80	0.577381	0.881818
Cheating	2 item	3	0.8	0.925
	3 item			
	Total	106	0.8	0.899371
Test	2 item	12	0.99	0.99
	3 item	24	0.99	0.99
	Total	36	0.99	0.99
Restaurant	2 item	33	0.681875	0.829299
	3 item	64	0.620645	0.820419
	Total	96	0.641489	0.823442
Adults	2 item	156	0.782308	0.887466
	3 item	109	0.759197	0.878065
	Total	264	0.772841	0.88422
Mushroom	2 item	10	0.723141	0.871656
	3 item			
	Total	10	0.723141	0.871656
Binary	2 item	6	0.7984	0.898001
	3 item	12	0.722769	0.860478
	Total	18	0.743778	0.870901

TABLE 8. CATEGORIZING DIFFERENT POSITIVE AND NEGATIVE RULES ON 9 FINAL DATASETS

Dataset	A → B	A → ~B	~A→ B	~A →~B
Transaction	1	16	20	4
Sport	0	5	5	6
Computer	2	5	5	6
Cheating	0	0	0	3
Test	36	0	0	0
Restaurant	0	0	0	96
Adults	21	49	52	33
Mushroom		5	5	
Binary	0	0	0	18

8. Recommendation for Future Works

Based on the results achieved on the different datasets to find other characteristics of negative association rules, including the proper place for using them, the best type of datasets and other characteristics, it is recommended that these negative association rules again examine different datasets, including those examined in this study, and their points of failure and success be analyzed again because the examination of the points of failure in the mentioned dataset and algorithms can help to get the other background characteristics of the association rules.

On the other hand, considering the difference between the tested datasets in many ways such as size and number of records, in order to better understanding of characteristics of negative association rules and the reasons of their success and failure in some datasets, further analysis of this algorithm on the dataset, including binary, is recommended.

References

[1] R. Kumar, P. K. Pattnaik and Y. Sharma, "Web Data Analysis Using Negative Association Rule Mining," in Information Systems Design and Intelligent Applications, ed: Springer, 2016, pp. 513-518.

[2] J. Agrawal, S. Agrawal, A. Singhai and S. Sharma, "SET-PSO-based approach for mining positive and negative association rules," Knowledge and Information Systems, vol. 45, no. 2, pp. 453-471, 2015.

[3] N. Shukla and A. Sen, "An Improved Association Rule Mining with Frequent Itemset Relationship," International Journal of Scientific Research in Science Engineering and Technology, vol. 2, no. 5, pp. 46-52, 2016.

[4] P. K. Singh, N. Rahim, S. Tiwari and N. Sahu, "Study of Various Association Rule Mining Techniques with Positive and Negative Integration," International Journal of Research, vol. 3, no. 10, pp. 83-90, 2016.

[5] R. Sethi and B. Shekar, "Issues in Negative Association Rule Mining with Business Analytics Perspectives," DHARANA-Bhavan's International Journal of Business, vol. 11, no. 2, pp. 13-20, 2018.

[6] A. Paul, "Positive and Negative Association Rule Mining Using Correlation Threshold and Dual Confidence Approach," in Computational Intelligence in Data Mining—Volume 1, ed: Springer, 2016, pp. 249-260.

[7] S. Çokpınar and T. İ. Gündem, "Positive and negative association rule mining on XML data

TABLE 9: SUPPORT AND CONFIDENCE VALUE FOR 9 FINAL DATASETS

		AB	~AB	A~B	~A~B	Other
Adults	Support	0.782	0.786684	0.786531	0.765263	0.7575
	Confidence	0.882766	0.904919	0.87288	0.87655	0.954079
Restaurant	Support				0.641489	
	Confidence				0.823442	
Mushroom	Support		0.723141	0.723141		
	Confidence		0.977227	0.746668		
Binary	Support				0.743778	
	Confidence				0.870901	
Sport	Support		0.585714		0.571429	0.571429
	Confidence		0.945455		0.793266	0.876263
Transaction	Support	0.45	0.383333	0.358333	0.4	0.383333
	Confidence	0.818182	0.728788	0.804545	0.759091	0.728788
Test	Support	0.99				
	Confidence	0.99				
Cheating	Support				0.8	
	Confidence				0.944444	
Computer	Support	0.571429	0.614286	0.614286	0.571429	
	Confidence	0.844444	0.897778	0.897778	0.814815	

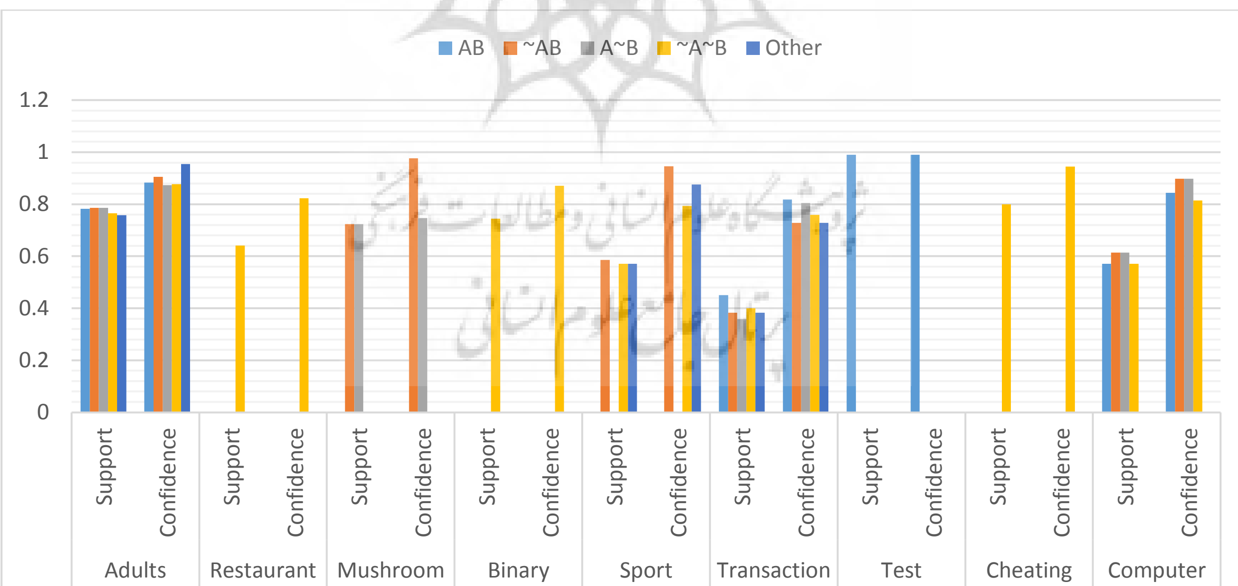

Fig.3. Summary of the Results of Extracted Rules on 9 Datasets

streams in database as a service concept," Expert Systems with Applications, vol. 39, no. 8, pp. 7503-7511, 2012.

[8] V. Kavi and D. Joshi, "A Survey on Enhancing Data Processing of Positive and Negative Association Rule Mining," International Journal of Computer Sciences and Engineering, vol. 2, no. 3, pp. 139-143, 2014.

[9] E. B. Krishna, B. Rama and A. Nagaraju, "Mining of negative association rules using improved frequent pattern tree," in Computer and Communications Technologies (ICCCT), 2014 International Conference on, 2014, pp. 1-6.

[10] C. Chen and D. Wang, "Research on Association Rules Mining Base on Positive and Negative Items of FP-tree," In 2016 6th International Conference on Machinery, Materials, Environment, Biotechnology and Computer. Atlantis Press. 2016.

[11] N. S. Rai, S. Jain and A. Jain, "Mining Interesting Positive and Negative Association Rule Based on Improved Genetic Algorithm (MIPNAR_GA)," International Journal of Advance Computer Science and Applications, vol. 5, no. 1, pp. 160-165, 2014.

[12] J. Rong, H. Q. Vu, R. Law and G. Li, "A behavioral analysis of web sharers and browsers in Hong Kong using targeted association rule mining," Tourism Management, vol. 33, no. 4, pp. 731-740, 2012.

[13] M. Shaheen, M. Shahbaz and A. Guergachi, "Context based positive and negative spatio-temporal association rule mining," Knowledge-Based Systems, vol. 37, pp. 261-273, 2013.

[14] D. Martin, A. Rosete, J. Alcalá-Fdez and F. Herrera, "A new multiobjective evolutionary algorithm for mining a reduced set of interesting positive and negative quantitative association rules," IEEE Transactions on Evolutionary Computation, vol. 18, no. 1, pp. 54-69, 2014.

[15] S. Naredi and R. A. Deshmukh, "Improved extraction of quantitative rules using Best M Positive Negative Association Rules Algorithm," 2015 IEEE International Conference on in Electronics, Computing and Communication Technologies (CONECCT), 2015, pp. 1-5.

[16] C. Ravi and N. Khare, "EO-ARM: an efficient and optimized k-map based positive-negative association rule mining technique," in Circuit, Power and Computing Technologies (ICCPCT), 2014 International Conference on, 2014, pp. 1723-1727.

[17] B. Kotiyal, A. Kumar, B. Pant and R. Goudar, "A Novel Concept for Mining Negative and Positive Rule Through Association Based K-MAP," Journal of Engineering Science and Technology, Special issue 1 , pp. 68-76, 2014.

[18] A. S. A. Kadir, A. A. Bakar and A. R. Hamdan, "Frequent absence and presence itemset for negative association rule mining,"11th International Conference on Intelligent Systems Design and Applications (ISDA), IEEE, 2011, pp. 965-970.

[19] C. Liu, R. Qi and Q. Liu, "Query expansion terms based on positive and negative association rules," In 2013 IEEE Third International Conference on Information Science and Technology (ICIST), IEEE, 2013, pp. 802-808

Mohammad Javad Kargar is an Associate Professor at the Department of Computer Engineering in University of Science and Culture, Tehran, Iran. His research interest is data science and distributed systems.

Fatemeh Hajilou has received master of software engineering at the Department of Computer Engineering in University of Science and Culture, Tehran, Iran.. Her research interest is data mining.