International Journal of
Emerging Research in Management &Technology
ISSN: 2278-9359 (Volume-6, Issue-8)

Research Article **August 2017**

Extracting of Positive and Negative Association Rules

D. Gandhimathi*
Department of Computer Science,
Alagappa University, Tamilnadu, India

N. Anbazhagan
Department of Mathematics,
Alagappa University, Tamilnadu, India

Abstract –

Association rules analysis is a basic technique to expose how items/patterns are associated to each other. There are two common ways to measure association such as Support and Confidence. Several methods have been proposed in the literature to diminish the number of extracted association rules. Association Rule Mining is one of the greatest current data mining techniques designed to group objects together from huge databases aiming to take out the motivating correlation and relation with massive quantity of data. Association rule mining is used to discover the associated patterns from datasets. In this paper, we propose association rules from new methods on web usage mining. Generally, web usage log structure has several records so we have to overcome those unwanted records from large dataset. First of all the pre-processed data from the NASA dataset is clustered by the popular K-Means algorithm. Subsequently, the matrix calculation is progressed on that data. Further, the associations are performed on filtered data and get rid of the final associated page results. Positive and negative association rules are gathered by using new algorithm with Annul Object (AO). Wherever the object "AO" is presented those rules are known as negative association rule. Otherwise, the rules are positive association rules.

Keywords - Positive Association Rules, Negative Association Rules, Frequent Patterns

I. INTRODUCTION

Web usage mining is the procedure of extracting useful information from web based data. It means how usage data from web logs can be analyzed/ mined to build user profiles and how these could be use to improve the user's browsing activities. The area of web usage mining is an important source of methods for the implementation of personalization functionality. Traditionally, Knowledge Discovery in Data has been used to examine data/ pattern accumulated on the web and take out valuable knowledge. This effort was named Web mining and one subdivision of it is interested with the analysis of usage data, i.e., records of how a web service is used. Association rule mining refers to ascertain association relationships among different attributes. The problem of mining strong association rules consists of two main steps, finding frequent item sets from a database and generating rules based on found frequent item sets. The first phase of this mining task is the most time consuming. This is because building association rules based on the found frequent pattern is simple. An association rule is an implication relation in the form (X→Y) between two separate sets of items X and Y. An example of an association rule on "Market Basket Data" is the 80% of customers who purchase pen with eraser together. Each association rule has two quality measurements such as support (S) and confidence (C). The rule has a support "S" in the transaction set D if S% of transactions in D contains X∪Y. The rule X→Y has confidence "C" if C% of transactions in the set of transactions D that contains X also contains Y. The problem of mining association rules is to discover all association rules that have a support and a confidence exceeding the user-specified threshold of minimum support known as Min-Sup and threshold of minimum confidence called Min-Conf respectively. A set of items is referred to as an item set. An item set that contains k items is a k-item set. The set {pen, eraser} is a 2-itemset. The occurrence frequency of an item set is the number of transactions that enclose the item set. This is known as the frequency, support count, or count of the item set.

Frequent pattern mining can be classified in various ways such as "Based on the completeness of patterns to be mined", "Based on the levels of abstraction involved in the rule set", "Based on the number of data dimensions involved in the rule", "Based on the types of values handled in the rule", "Based on the kinds of rules to be mined, "Based on the kinds of patterns to be mined" [6]. We can extort the complete set of frequent item sets, the closed frequent item sets, and the maximal frequent item sets, given a suitable minimum support threshold. We are able to extract constrained frequent item sets (those that satisfy a set of user-defined constraints), approximate frequent item sets (those that derive only approximate support counts for the mined frequent item sets), near-match frequent item sets (i.e., those that tally the support count of the near or almost matching item sets), top-*k* frequent item sets (i.e., the *k* most frequent item sets for a user-specified value, *k*), and so on. Some techniques for association rule mining can find rules at differing levels of abstraction that is known as "Based on the levels of abstraction involved in the rule set".

Based on the number of data dimensions involved in the rule represents if the items or attributes in an association rule reference only one dimension, then it is a Single-Dimensional Association Rule. Suppose, if a rule references two or more dimensions, then it is a Multi-Dimensional Association Rule. Based on the types of values handled in the rule means if a rule involves associations between the presence or absence of items, it is a Boolean Association Rule. If a rule describes associations between quantitative items or attributes, then it is a Quantitative Association Rule. Quantitative values for items or attributes are partitioned into intervals. Based on the kinds of rules to be mined describes Frequent pattern analysis can generate various kinds of rules and other interesting relationships. Association rules are the most popular kind of rules generated from frequent patterns. Based on the kinds of patterns to be mined that mean many kinds of frequent patterns can be mined from different kinds of data sets. The support is calculated by the equation 1:

$$\text{Support } (X \Rightarrow Y) = P\ (X \cup Y) \qquad (1)$$

Once the frequent itemsets/ patterns from transactions in a database "D" have been found, it is straightforward to generate strong association rules from them where strong association rules must satisfy both minimum support (MS) and minimum confidence (MC). The confidence is basically calculated by the equation 2:

$$\text{Confidence}(X \Rightarrow Y) = P\ (Y|X) = \text{support } (X \cup Y)/\text{support}(X) = \text{support count}(X/Y) \qquad (2)$$

II. RELATED WORKS: AN OVERVIEW

The authors propose the algorithms to create the usefulness based non-redundant association rules and methods for reconstructing all association rules. Furthermore, the researcher describes the algorithms which generate high utility itemsets (HUI) and high utility closed itemsets with their generators. These proposed algorithms are implemented using both synthetic and real datasets which are performed by [13]. The researchers propose a set of algorithms for discovering both positive and negative association rules (NAR) in databases is presented. A variant of the Apriori, traditional association rules algorithm, is attained by using support and confidence to discover two types of NAR; the confined negative association rules (CNR), and the generalized negative association rules (GNAR) all works are expressed in [18]. This paper has contributed to this emerging topic by cataloguing, and categorizing problems with, existing negative AR definitions and mining approaches, and by proposing a new Apriori-based algorithm (PNAR) that exploits the upward closure property of negative association rules those are discussed in [2]. This study proposes an efficient method for extracting CARs with the itemset constraint based on a lattice structure and the difference between two sets of object identifiers (diffset). Firstly, a lattice structure is built to store up all frequent itemsets in the large dataset. To decrease memory usage, instead of the whole set of object identifiers, the diffset is used. Secondly, the lattice is traversed to produce only rules which satisfy the itemset constraint all tasks are discussed in [10]. The authors proposes a new measure VARCC which merges correlation coefficient and minimum confidence is proposed and a corresponding algorithm PNAR_MLMS is proposed to make PNARs correctly from the frequent and infrequent itemsets discovered by the MLMS model which are concentrated in [4]. In this paper the authors develop a novel model-based frequency constraint as different to a single, user-specified minimum support. The constraint utilizes knowledge of the process generating transaction data by applying a simple stochastic mixture model (the NB model) which permits for transaction data highly skewed item frequency distribution all works are performed by [5]. This paper is extended version of existing work, the authors have compared the results of Apriori and K-Mean algorithms against their implementation in Weka and XLMiner all are explained in [8]. In this paper the researchers propose an algorithm that broadens the support-confidence framework with a sliding correlation coefficient threshold. In addition to finding confident positive rules that have a strong correlation, the algorithm realizes negative association rules with strong negative correlation between the antecedents and consequents all are explained in [1].

This article on how to develop the algorithm efficiency, diminish the irrelevant rules are studied, Apriori improved algorithm is proposed based on SQL. A little of this algorithm is to save storage space and reduce I/O load; avoiding repeated scrutinizing database has nothing to do a lot of difficulties those methods are expressed in [9]. The proposed model is incorporation among two algorithms, the Positive Negative Association Rule (PNAR) algorithm and the Interesting Multiple Level Minimum Supports (IMLMS) algorithm, to propose a new approach (PNAR_IMLMS) for mining both negative and positive association rules from the attractive frequent and infrequent itemsets extracted by the IMLMS model all are discussed in [17]. Authors propose a method to trim rules that are statistically insignificant with respect to more general rules. Such rules may exist in the presence of high-confidence rules, which is often the case in web usage data. The method effectiveness is validated on two real-life web usage data sets all details are explained in [3].

The researchers propose and evaluate two different algorithms based respectively on crisp rules and fuzzy rules, concluding that fuzzy meta-association rules are appropriate to integrate to the meta-mining practice the obtained quality assessment provided by the rules in the first step of the process, although it consumes more time than the crisp approach those works are explained in [12]. In this paper, authors blue deeply focused on the need for minimal multilevel and cross-level association rules in blue real-life datasets. That algorithm for mining the minimal cross-level association rules from the frequent closed itemset blue lattice by efficiently traversing the hierarchically built lattices those works are discussed in [7]. Researchers propose semantically enriched Web Usage Mining method (SWUM), it merges the fields of Web Usage Mining and Semantic Web. In the proposed method, the undirected graph obtained from usage data is enriched with rich semantic information extracted from the web pages and the web site structure and the results are given in [14]. Authors use the Bitmap table to build a data structure called Peano Tree stored as a binary file on which they apply a new algorithm called BF-ARM (extension of the well known Apriori algorithm). The database is loaded into a binary file, their proposed algorithm will traverse this file, and the processes of association rules extractions will be based on the file stored on disk which is explained in [16]. The authors focus privacy problem by considering the algorithmic requirements for providing confidentially and improve the performance at the time when database stores and retrieves huge amount of data all are realized by [11]. In this article, researchers provide a brief review and analysis of the current status of frequent pattern mining and discuss some promising research directions those are expressed in [15].

III. PROPOSED METHODOLOGY

This section initiates association rules terminology and produce both positive and negative association rules mining. The new Alpha Numeric ($\mathcal{AN}$) matrix algorithm is used to mine association rules. To extract the positive association rules in the form of X→Y. We have used the new matrix based algorithm. This means that whatever rules don't contain Annul Object ($\mathcal{AO}$), such rules are positive rules.

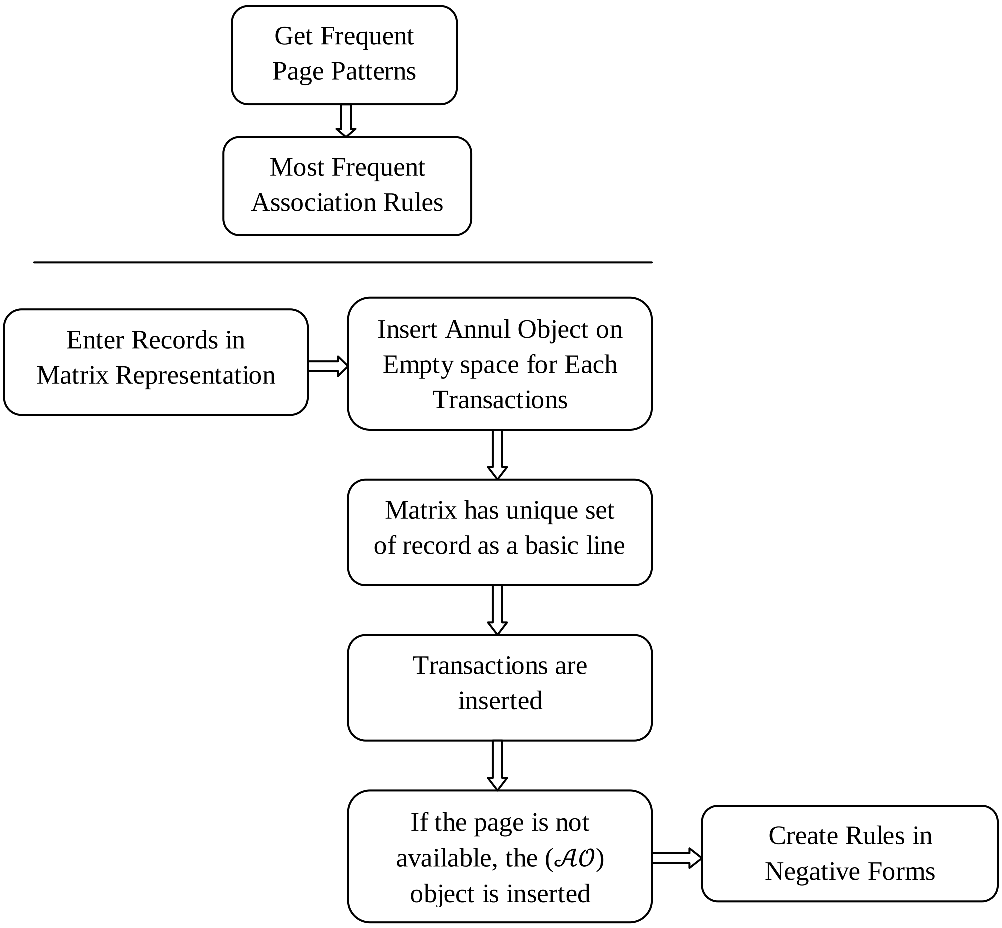

Fig 1. General Framework of the Proposed Work

Then, negative association rules such as X→ ┐Y, ┐X → Y, ┐X→ ┐Y are derived from the matrix. Newly, in this paper we have used an Annul Object ($\mathcal{AO}$) to mine the negative association rules from the matrix. Major advantage of this positive and negative association rules is we needn't check the entire records again for negative association rules by this matrix algorithm. Fig 1 represents the general framework of the proposed work is mainly used for making the positive and negative association rules. Just we can make the positive association rules from the frequent data but we cannot extract the negative associations from frequent data. At that time we have used this algorithm for making the positive and negative rules through the single scan of the records. The purpose of negative rules is we can get whatever pages are not accessed with other pages. From this we can gather the negation of the patterns also, so that the business analyst can improve the quality of those pages. The business analyst will have given the less concentration on whatever pages are not accessed together. So, we have to identify the negative association rules based on three various forms.

A. Forms of the positive and negative association rules based on sustainability of page count

In this paper, we have analyzed positive and negative rules based on our $\mathcal{AN}$ matrix methodology. We express only the association among various pages that are accessed by the online visitor.

X→Y = this means that, when both column values are increased simultaneously. For example: we consider two different page requests such as P_{20}→P_{27} then, $\mathcal{S}_1$=(1,1), $\mathcal{S}_2$ = (2,2). Here the page requests P_{20} and P_{27} are increased in both sessions ($\mathcal{S}_1$, $\mathcal{S}_2$). This strategy can be applied in various session results so all are considered as the most positive association rules.

X→ ┐Y = this means that, when first count value is incremented but the second count value is in "*sustain mode*". For example: we consider two different page requests to express the sustainability such as P_{20}→P_{27} then, $\mathcal{S}_2$=(2,2), $\mathcal{S}_3$ =(3,2). Here the page request P_{20} is increased in session ($\mathcal{S}_3$). But, in the session 3 ($\mathcal{S}_3$) the page P_{27} is not increased. This strategy is known as *"sustainability"*.

┐X → Y = this means that, when first count value is in the "*sustain mode*" but the second count value is increased by one. This means that one page is visited by the online user another one is not accessed by the user. For example: we consider two different page requests to express the sustainability of the first page such as P_{20}→P_{27} then, $\mathcal{S}_3$=(3,2), $\mathcal{S}_8$ =(3,3). Here the page request P_{27} is increased in session ($\mathcal{S}_8$). But, in the session 8 ($\mathcal{S}_8$) the page P_{20} is not increased. Here, we should check the previous row count value for that particular column. Suppose, the same value is on the previous row and the current row the count value is sustained. This means that the page is not visited by the user at that session ($\mathcal{S}$).

$\neg X \rightarrow \neg Y$ = this means that, both column values are in the "*sustain mode"*. Here, $P_{20} \rightarrow P_{27}$ then, $\mathcal{S}_8$=(3,3), $\mathcal{S}_9$= (3,3) so whenever these sustainable values are available at that place we have inserted an *"Annul Object"* that is ($\mathcal{AO}$). Suppose, the new value is on the particular column at first we can count it. But, the same value is available on the same column in the next row the value is sustained.

Table I. Explain The Sustainability And Increased Mode

Sessions	Count Value of page P_{13}	Form of the mode
$\mathcal{S}_8$	$P_{13} = 1$	New count value
$\mathcal{S}_9$	$P_{13} = 2$	Increased value
$\mathcal{S}_{10}$	$P_{13} = 2$	Sustain mode
$\mathcal{S}_{11}$	$P_{13} = 2$	Sustain mode
$\mathcal{S}_{12}$	$P_{13} = 3$	Increased value

Table I explains the Sustainability and Increased Mode. Here, five various sessions of the page 13 (P_{13}). Form of the mode is changed based on the user accessing about pages. New count value is started in the session ($\mathcal{S}_8$) then, in the session ($\mathcal{S}_9$) the value is increased by one so mode is changed here. In the third row the same previous value is sustained on the sessions ($\mathcal{S}_{10}$, $\mathcal{S}_{11}$). Then, the session $\mathcal{S}_{12}$ the page (P_{13}) is incremented by one.

B. Mining of Positive Association Rules

Association rules are defined as follows: Let page requests are $P_i = \{P_1, P_2, \ldots, P_m\}$ be a set of items. Let $\mathcal{S}$ be a set of transactions/ Sessions, where each transaction $\mathcal{S}$ is a set of items. A transaction $\mathcal{S}$ is said to contain P_1, a set of items in P_i, if $P_1 \subseteq \mathcal{S}$. An association rule is an implication of the form "$P_1 \rightarrow P_2$", where $P_1 \subseteq P_i$; $P_2 \subseteq P_i$, and $P_1 \cap P_2 = \Phi$. The rule $P_1 \rightarrow P_2$ has support "s" in the transaction set D if s% of the transactions in $\mathcal{S}$ contains $P_1 \cup P_2$. The model of the proposed work for extracting positive association rule mining is based on Min-Sup and Min-conf that is minimum support and minimum confidence respectively. The positive association rules are derived from matrix.

Table II. Example Of A Transactional Dataset In The Matrix

$\mathbf{P_i}$	$\mathbf{P_1}$	$\mathbf{P_2}$	**………**	$\mathbf{P_m}$
T_1/S_1	$Page_{11}$	$Page_{21}$	………	$Page_{m1}$
T_2/S_2	$Page_{12}$	$Page_{22}$	………	$Page_{m2}$
………	……..	………	……….	………
……..	……..	………	……….	………
………	……...	………	………..	……….
………	………	………	………..	……….
………	………	………	………..	…….....
T_n/S_n	$Page_{1n}$	$Page_{2n}$	……….	$Page_{mn}$

Table II represents the Example of a Transactional Dataset in the Matrix. Transactions in the dataset that is matrix oriented representation and each transaction has more number of pages. The last row of values is n^{th} values up to $page_{mn}$. The dataset records are located within the ($\mathcal{AN}$) matrix. Matrix has unique set of records as a base line $\mathfrak{B}$ those records are available only once in the matrix. We have gathered most positive associations from this new matrix methodology. The base line structure is $P_i = \{P_1, P_2, \ldots, P_m\}$ be a set of individual page patterns. Here, we have to concentrate one important task such as we can make most positive association rules from matrix data. We are able to get only most positive associations. Top most page results only we get from the ($\mathcal{AN}$) matrix. From each level of records we are able to get some number of frequent association results. Set of records are processed here as most positive associations. In the existing works, the pattern's association rule results satisfy the minimum support and minimum confidence. If the rule doesn't satisfy the minimum support (MS) and minimum confidence (MC), it will be considered as a negative rule.

In this paper we have applied the Alpha Numeric Matrix based algorithm, and then we have done permutations for all columns in the matrix. Scanning of the matrix is done at once at that time we have checked whether the page is available or the $\mathcal{AO}$ is available. Finally we realize more frequent page patterns. In this paper we displayed only some important association rules. But, all of the permutations are performed by this new methodology. Before doing the permutation performance, we additionally include the "$\mathcal{AO}$" object on "sustainable values". We must add one more important task as acting object to get both positive and negative association rules. Getting of best positive association rules only in the form of X→Y by using the matrix based algorithm. But we have to know the other positive and negative forms of the rules also. So, that we may use "Annul Object" is used for getting both positive and negative rules also. Perhaps, if the rules don't contain the Annul Object, then those are called positive association rules. Finally check the minimum support and minimum confidence values whether it is greater than or not.

Table III. Some Of The Page Requests From Transactional Dataset With Sample Count

Pages	P_{20}	P_{27}	P_{30}	P_{35}	P_{42}	P_{48}
Sample Count	6	5	7	5	5	7

Table III represents some of the page requests from transactional dataset with sample count. All those page requests are incremented simultaneously in various sessions ($\mathcal{S}$) in the dataset. Min-Sup is 2 percentages and Min-Conf is 0.50 percentages. For example: Let us focus on the access of page request (P_{42}) and page request (P_{48}). Suppose $S(P_{42}) = 5$, $S(P_{48}) = 7$, $S(P_{42} \cup P_{48}) = 1$, MS=2 and MC=0.50. Support $(X \Rightarrow Y) = P (X \cup Y)$ Because, $S(P_{42} \cup P_{48}) = 1 < MS$, $P_{42} \cup P_{48}$ is an infrequent itemset, $P_{42} \Rightarrow P_{48}$ cannot be extracted as a rule. But $S(P_{42} \cup \neg P_{48}) = S(P_{42}) - S(P_{42} \cup P_{48}) = 5 - 1 = 4 > MS$, and Confidence $(X \Rightarrow Y) = P (Y|X)$ that is, $C(P_{42} \Rightarrow \neg P_{48}) = S(P_{42} \cup \neg P_{48})/S(P_{42}) = 2/3 = 0.67 > MC$. Therefore, $P_{42} \Rightarrow \neg P_{48}$ should be extracted as a negative rule. Here, we derive the positive rules traditionally.

Table IV. Some Results Of Positive Association Rules

Associated Pages	X→ Y(support, confidence)
P_{12},P_{14}	s=4,c=1
P_{16}, P_{12}	s=3,c=1
P_{16}, P_{14}	s=3,c=1

Table IV represents different web pages. It represents some results of Positive Association Rules. The pages were visited by the user such as P_{12}, P_{13}, P_{14}, P_{15}, P_{16}. Those positive rules are derived by using the matrix based algorithm. We have decided the Min-Sup and Min-Conf are (2,1) respectively.

Calculate the support by the equation 3:

$$S(X \Rightarrow Y) = S(X \cup Y) = S(P_{12},P_{13}) = 1 < \text{Min-sup} \qquad (3)$$

Calculate the confidence by the equation 4:

$$C(X \Rightarrow Y) = \frac{S(X \cup Y)}{S(X)} = \frac{S(P_{12} \cup P_{13})}{S(P_{12})} = \frac{1}{4} = 0.25\,, \quad \frac{S(Y \cup X)}{S(Y)} = \frac{S(P_{13} \cup P_{12})}{S(P_{13})} = \frac{1}{3} = 0.34 < \text{Min-conf} \qquad (4)$$

The main thing is the $\mathcal{AO}$ does not available in the resulting rules. Such kind of rules is known as positive association rules. The confidence value "C" is 1 and the results are basically two. Because, the (X→ Y) and (Y→ X) both are performed. Finally, the confidence value is changed by using the vice versa operation. The rules are called *"Strong Positive Association Rules"*. Here, "X" represents the first value and "Y" represents the second value. Wherever the $\mathcal{AO}$ is not included in the rule and also the resulting value must satisfy the min-support and min-confidence, those are known as Positive Association Rules. For example, we have mentioned two different pages and calculate the support and confidence. According to the equation 4, two different results are there. Both are not satisfied the Min-conf and by the equation 3 it don't satisfy the Min-sup also. So, this is not a strong association rule. Finally, X⇒Y cannot be extracted as a rule so this is infrequent pattern from the alpha numeric "$\mathcal{AN}$" matrix. This same procedure is applied throughout the entire matrix. It only represents the positive association rules because; no one rule doesn't contain the "Annul Object". If the rule must satisfy the minimum support and minimum confidence those are the strong rules. We have explained only three various rules are extracted by the matrix algorithm in sample set of pages. The rules are generated based on the vice versa also. All the other rules don't satisfy the Min-Sup and Min-conf.

C. Mining of Negative Association Rules

In order to obtain more valued negative association rules, more infrequent patterns should be determined. Infrequent patterns are the complement of frequent patterns theoretically and they are too huge to be discovered entirely. The purpose of negative association rules is we have to analyze the negation of pages because whatever pages are not accessed by the same user. So, we realize that whatever associations are generally made by the user. All the pages are not associated with all other pages by the online user. We needn't scan again for extract the negative association rules. The task of our algorithm is when we do process for positive association rule at the same time the negative association rule process is also doing. If the $\mathcal{AO}$ is present in the checking column the rule is negative or otherwise the rule is positive. So, the time utility is basically reduced by this matrix algorithm. To discover the association rules, whenever the same value is sustained on the matrix the "Annul Object" is located on that. The symbolic representation is $\mathcal{AO}$. Fig 2 represents our new algorithm. In negative association, we checks whether the "$\mathcal{AO}$" is available or not. Suppose, the availability of "$\mathcal{AO}$" is true it must be considered as the negative association rules. The algorithm is processed and the permutations are performed, whenever the $\mathcal{AO}$ is available on the matrix that is null. The particular page request was not accessed on that particular session ($\mathcal{S}$). Negative association rules are in three forms but here we explained only one form of the rule for example. We gave considered two distinguish pages such as P_{13},P_{16}. We have explained a negation form like X→ ┐Y.

Given: A set of session $\mathcal{S}_t = \{\mathcal{S}_1, \mathcal{S}_2, .., \mathcal{S}_q\}$, $\forall$ t=1…..n, a set of pages $P_i = \{P_1, P_2, \ldots, P_m\}$, count=0;
Begin

1. For each page i = 1 to m
2. If (page=T in one session/transaction)
3.   Count=count+1
4. Else if (page=F in one session)
5.   Count=$\mathcal{AO}$
6. End if
7. If (page=T & previous one of that column=count value)
8.   Count=count+1
9. Else if (page=T & previous one of that column= $\mathcal{AO}$)
10. Else if (check before one of the previous one)
11.   Count=count+1
12. End if
13. End for

End

Fig 2: Algorithm-1

Calculate the support by the equation 5:

$$S(X \Rightarrow \neg Y) = S(X) - S(X \cup Y) \Rightarrow \quad S(X \cup \neg Y) = 3 > \text{Min-Sup} \tag{5}$$

Calculate the confidence by the equation 6:

$$C(X \Rightarrow \neg Y) = \frac{S(X \cup \neg Y)}{S(X)} = \frac{P_{13} \cup \neg P_{16}}{P_{13}} = \frac{3}{3} = 1 \text{ , } \frac{S(\neg Y \cup X)}{S(\neg Y)} = \frac{P_{16} \cup \neg P_{13}}{P_{16}} = \frac{3}{3} = 1 > \text{Min-conf} \tag{6}$$

Table V. Some unique sessions on the Numerical Matrix with $\mathcal{AO}$

Sessions	P_{12}	P_{13}	P_{14}	P_{15}	P_{16}
$\mathcal{S}_1$	1	1	1	1	$\mathcal{AO}$
$\mathcal{S}_2$	2	$\mathcal{AO}$	2	$\mathcal{AO}$	1
$\mathcal{S}_3$	3	$\mathcal{AO}$	3	$\mathcal{AO}$	2
$\mathcal{S}_4$	4	$\mathcal{AO}$	4	2	3
$\mathcal{S}_5$	$\mathcal{AO}$	2	$\mathcal{AO}$	3	$\mathcal{AO}$
$\mathcal{S}_6$	$\mathcal{AO}$	3	5	$\mathcal{AO}$	$\mathcal{AO}$

Accordingly to the $\mathcal{AO}$, whenever a particular page is unavailable in the current session, this $\mathcal{AO}$ is presented on that position within the numeric matrix $\mathcal{AN}$. Whether the page is presented on single session/transaction the count increased otherwise the $\mathcal{AO}$ is positioned on it. Perhaps, the numeric value is available on the previous one the count value is incremented otherwise check the before of the previous one. Table V represents the basic line of this paper is presented (Some unique sessions on the Numerical Matrix with $\mathcal{AO}$). For example, we considered two various sessions such as $\mathcal{S}_1$, $\mathcal{S}_2$ and concentrate on the columns (P_{13}, P_{14}) two and three both are increased in $\mathcal{S}_1$. But, obviously in $\mathcal{S}_2$ the page P_{13} is unavailable so it is filled up by the $\mathcal{AO}$. Then, page P_{14} is available in $\mathcal{S}_2$ so this is increased in the session $\mathcal{S}_2$. The same methodology is used throughout the whole matrix. Finally, we take the same columns but various sessions such as $\mathcal{S}_5$, $\mathcal{S}_6$. The page P_{13} in session $\mathcal{S}_6$ is incremented and previous one also. But, P_{14} in session $\mathcal{S}_6$ is increased but the previous one is positioned by $\mathcal{AO}$.

Table VI. Some Results Of Negative Association Rules

Pages	$X \rightarrow \neg Y$(S,C)	$\neg X \rightarrow Y$(S,C)	$\neg X \rightarrow \neg Y$(S,C)
P_{12},P_{13}	s=3,c=0.75,1	s=2,c=1	-
P_{12}, P_{16}	-	-	s=2,c=1,0.67
P_{13}, P_{14}	-	s=3,c=1,0.6	-
P_{13}, P_{16}	s=3,c=1,1	s=3,c=1,1	-
P_{14}, P_{15}	s=3,c=0.6,1	-	-

Table VI represents the negative forms of the association rules. It gives some results of Negative Association Rules. The negative forms such as $X \rightarrow \neg Y$, $\neg X \rightarrow Y$, $\neg X \rightarrow \neg Y$. All of the matrix records are participated in this methodology.

The sample pages are calculated for explain the negative rules. We have taken sample set of pages then, perform the associations between different pages. The permutations are calculated for all pages in the matrix. Scan for a negative association rule is never required in this algorithm.

So, time and space utility is saved. Whenever the page is not accessed by the user within one session ($\mathcal{S}$), it will be filled up by the "Annul Object". The purpose of $\mathcal{AO}$ is we don't take any care about negative rules. Simply we use this object for matrix then the permutations are automatically performed by the matrix algorithm. Lastly, we get the negative rules also and we have given some of the important results only. We realize the strong negative association rules. But, in the confidence value both results are given for vice versa operations. For example, in the first row and first column of pages such as P_{12},P_{13} the support result is 3 and confidence result is 0.75 and 1 so, this is a negative association rule.

And the vice versa operation $\neg$ Y→ X the result is 3,1 support and confidence results respectively. Then, this is the strong negative association rule. So we have given only wherever the support and confidence results are satisfied and its reverse operation also. In the third row and third column the confidence vice versa rule has the value 0.6. This result shows the rule $P_{13} \rightarrow P_{14}$ has less minimum confidence that is 0.6 < Min-Conf. So, the rule is not a negative association rule. Whenever both Min-Sup and Min-Conf are satisfied by the association rules those rules are positive or negative association rules.

D. Performance Evaluation

The system oriented details are Intel(R) Pentium (R) CPU B960 @ 2.20 GH$_z$ 2.20 GH$_z$ Processor, 32-bit Windows 7 Operating System and 4.00 GB Memory in which the experiments are conducted. Every contribution of the matrix algorithm has been implemented using .net programming language. We have used NASA dataset with 1,891,715 numbers of entries and 523,160 total Page-Requests. The records are filtered and finally the strong positive and negative association rules are produced by the matrix algorithm.

Fig 3 represents positive association rules. The overall NASA records are grouped with five various levels then each level of records are processed and the rules are derived. Basic line "$\mathfrak{B}$" is used within the matrix for uniqueness of the visited pages. This is the main advantage of this matrix algorithm because, every page is available only once in the basic line. Based on the Min-Sup and Min-Conf levels the overall matrix records are diminished and finally it displayed the most positive association rules and best negative association rules.

Fig 4 represents the negative association rules of the dataset. It may contain frequent 2 rules or frequent 3 rules. We can derive the frequent 1 page pattern rules from the frequent 2 and 3 rules. The axis shows the percentage ranges of the negative association rules. The final numbers of records are only given here.

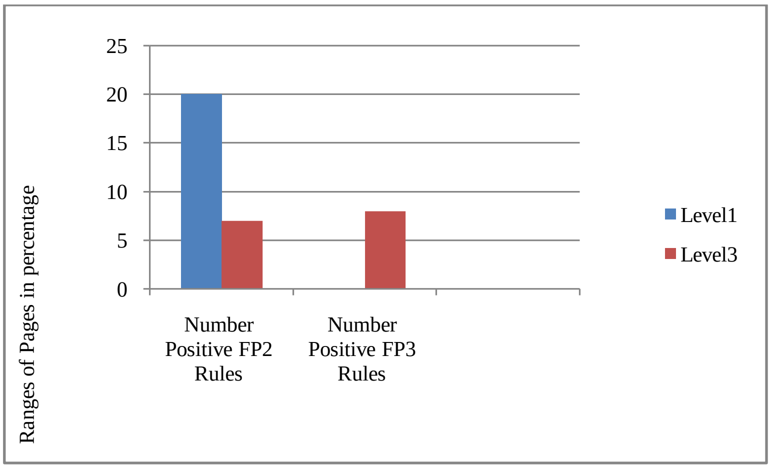

Fig 3. Positive Association Rules

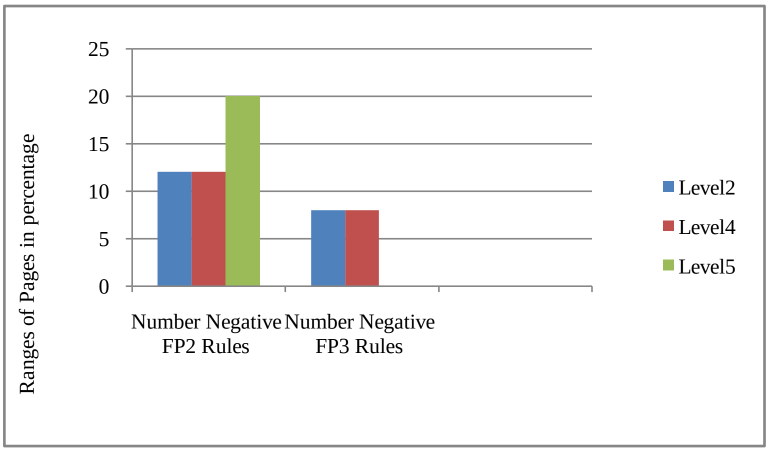

Fig 4. Negative Association Rules

IV. CONCLUSION

The major drawback in some existing works is it gives the positive and negative association rules with more time and space utility. Because the prior works flow throughout the whole data or dataset and make a new modified algorithm for negative association rules. In this paper, we proposed a matrix based algorithm that extracts both positive and negative association rules in single scan process. When the scan is started at that time the matrix algorithm checks the annul object is available or not. We check wherever, the object is presented within one session record in the NASA dataset. The annual object is presented in any comparative association pages in three various negation forms. All those tasks are performed in single scan of the Alpha Numeric Matrix. Our new technique generates positive and negative association rules by using support and confidence framework. We needn't scan the data again for mining negative association rules by using $\mathcal{AO}$. So, the time consumes are diminished with space utility by using this new matrix algorithm. We mined frequent 2 rules and frequent 3 rules then, the frequent 1 page pattern rules are derived.

REFERENCES

[1] Antonie, M-L., & Za¨ıane, O.R. (2004). Mining Positive and Negative Association Rules: An Approach for Confined Rules. Knowledge Discovery in Databases, 3202, 27-38.

[2] Cornelis, C., Yan, P., Zhang, X., Chen G. (2006). Mining Positive and Negative Association Rules from Large Databases. IEEE Conference on Cybernetics and Intelligent Systems DOI: 10.1109/ICCIS.2006.252251.

[3] Dimitrijevic, M., Bosnjak, Z. (2014). Pruning statistically insignificant association rules in the presence of high-confidence rules in web usage data. In proceedings of the 18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems - KES2014, Procedia Computer Science, (35), 271 – 280, DOI: 10.1016/j.procs.2014.08.107.

[4] Dong, X., Niu, Z., Shi, X., Zhang, X., and Zhu, D. (2007). Mining Both Positive and Negative Association Rules from Frequent and Infrequent Itemsets. Advanced Data Mining and Applications, 4632, 122-133.

[5] Hahsler, M. (2006). A Model-Based Frequency Constraint for Mining Associations from Transaction Data. Data Mining and Knowledge Discovery, 13(2), 137–166, DOI: 10.1007/s10618-005-0026-2.

[6] Han, J., & Kamber, M. Data Mining: Concepts and Techniques. 500 Sansome Street, Suite 400, San Francisco, CA 94111, Morgan Kaufmann Publishers is an imprint of Elsevier.

[7] Hashem, T., Ahmed, C.F., Samiullah, Md., Akther, S., Jeong, B-S., Jeon, S. (2014). An Efficient Approach for Mining Cross-level Closed Itemsets and Minimal Association Rules Using Closed Itemset Lattices. Expert Systems with Applications, 41(6), 2914-2938.

[8] Khattak, M., Khan, A.M., Lee, S., and Lee, Y-K. (2010). Analyzing Association Rule Mining and Clustering on Sales Day Data with XLMiner and Weka. International Journal of Database Theory and Application, 3(1), 13-22.

[9] Li, D. (2015). Improvements in Data Mining Association Rules Algorithm. International Journal of Database Theory and Application. 8(2),1-10, DOI: http://dx.doi.org/10.14257/ijdta.2015.8.2.01.

[10] Nguyen, D., Nguyen, L.T.T., Vo, B., and Pedrycz, W. (2016). Efficient Mining of Class Association Rules with the Itemset Constraint. Knowledge-Based Systems, 103(1), 73–88, doi: http://dx.doi.org/10.1016/j.knosys.2016.03.025.

[11] Rao, K.S., Mandhala, V.N., Bhatacharyya, D., Kim, T-H. (2014). An Association Rule hiding Algorithm for Privacy Preserving Data Mining. International Journal of Control and Automation, 7(10), 393-404.

[12] Ruiz, M.D., Romero, J.G., Solana, M.M., Campa˜na, J.R., Bautista, M.J.M. (2016). Meta-association rules for mining interesting associations in multiple datasets. Applied Soft Computing, 49, 212–223, doi: http://dx.doi.org/10.1016/j.asoc.2016.08.014 .

[13] Sahooa, J., Dasb, A.K., Goswami, A. (2015). An efficient approach for mining association rules from high utility itemsets. Expert Systems with Applications, 42(13), 5754–5778, doi: http://dx.doi.org/10.1016/j.eswa.2015.02.051.

[14] Shirgave, S., & Kulkarni, P. (2013). Semantically Enriched Web Usage Mining for Predicting User Future Movements. International Journal of Web & Semantic Technology (IJWesT), 4(4), 59-72, DOI :10.5121/ijwest.2013.4405.

[15] Slimani, T., Lazzez, A. (2014). Efficient Analysis of Pattern and Association Rule Mining Approaches. International Journal of Information Technology and Computer Science (IJITCS), 6(3), 70-81, DOI: 10.5815/ijitcs.2014.03.09.

[16] Slimani, T. (2013). New Approach to Optimize the Time of Association Rules Extraction.IJCSI International Journal of Computer Science Issues, 10(5), No 1, 234-243.

[17] Swesi, I.M.A.O., Bakar, A.A., Kadir, A.S.A. (2012). Mining Positive and Negative Association Rules from Interesting Frequent and Infrequent Itemsets. 9th International IEEE Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 650 - 655, DOI: 10.1109/FSKD.2012.6234303.

[18] Tsiligaridis, J. (2013). Mining Positive and Negative Association Rules: An Approach for Binary Trees. International Journal of New Computer Architectures and their Applications (IJNCAA), 3(3), 81-87.