Journal of Theoretical and Applied Information Technology
31st December 2017. Vol.95. No 24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MINING OPTIMIZED POSITIVE AND NEGATIVE ASSOCIATION RULE USING ADVANCE ABC ALGORITHM

[1]I.BERIN JEBA JINGLE, [2]J.JEYA A.CELIN

[1]Assistant Professor, Noorul Islam University, Department of CSE, Tamilnadu, India

[2]Assistant Professor, Hindustan Arts and Science College, Department of computer Application, Tamilnadu, India

E-mail: [1]berinjeba@gmail.com, [2] jjeyacelin@gmail.com

ABSTRACT

One of the imperative mining used in data mining is the association rule mining which mines many eventual information and association from outsized databases. The Association rule mining has many research challenges. The generation of accurate frequent and infrequent itemset by reducing the candidate itemset and reducing the space in the memory is one of the challenges. The generation of positive and negative association rule with high confidence and high quality is another challenge. The next research challenge is to mine an optimized positive and negative association rule. Numerous existing algorithms have been used to overcome these challenges but many such algorithms cause data loss, lack of efficiency and accuracy which also results in redundant rules and loss of space in memory. The major issue in using this analytic optimizing method is specifying the activist initialization limit were the quality of the association rule relays on. In the Proposed work, it is proved that an efficient optimized Positive and negative association rule can be mined by the proposed Advance ABC algorithm. The advance ABC (Artificial Bee Colony) algorithm is swarm intelligence based algorithm highly efficient optimization algorithm. The Advance ABC Meta heuristic technique is stimulated through the natural food foraging behaviour of the honey bee creature. The result shows that the proposed algorithm can mine incredibly high confidence non redundant positive and negative association rule with less time.

Keywords: *Data Mining, Association Rule Mining, Apriori Algorithm, Accurate Multi Level And Multi Support,Advanceabcalgorithm,GPNAR*

1. INTRODUCTION

Data mining is a very imperative field in the case of science. The process of data mining is mining of unseen prognostic information from huge databases. The data mining is recent technology which helps business management and market analysis to spotlight the indispensable information in their data warehouse. The data's are extracted from different sources such as Text, Image, and Web etc. Analyzing such data are very important process nowadays the knowledge discovery and the data mining has plays immense responsibility in changing these data into valuable information and patterns.There are many techniques available in data mining for extracting data from various data sources such as association rule, classification, decision tree, clustering, prediction, etc. Generally these techniques are mainly planned for the rationale of developing efficient mining algorithm in order to extract patterns within sensible and tolerable time frame.

The association rule (AR) mining is a technique of data mining which is used to analyze high-dimensional relational data. The association rule mining discovers interesting relationship hidden in a large dataset. The association rule techniques are implemented effectively in application domain such as market basket analysis, intrusion detection, telecommunication and diagnosis decision support. The major focal point of the association rule mining is the research community.

The proposed algorithms extracts high quality, high confidence association rule from transactional database. The optimized advance ABC algorithm mines the optimized association rules by removing the redundant rules and inefficiency and mould the mined PAR and Nar into supreme one.

The advance ABC algorithm uses parameter like Number of Food Sources, Limit and a uniform random number. The random number is uniformly generated and randomly generated during each iteration and does not depend on any component.

The population size is similar to the number of food source and also the Advance ABC algorithm performance depends on the size of the initial population size. Hence as the size of the population size enlarges the performance of Advance ABC algorithm get worse. As the size of the population varies the parameter limit also changes because the parameter limits are the product of size width and population size so a slight change in the limit leads to large changes in the performance of Advance ABC algorithm.

An algorithm MIPNAR_GA [1] has presented by Nikky Rai for mining optimized association rules. This method determines the crisis of negative rule generation and as well as optimized the method of rule generation. This algorithm is categorized into three phases. The generation of frequent and infrequent itemset by Apriori algorithm is the first stage. The mining of positive and negative association rule is the next stage. And finally prune redundant rule has been applied for interest measurements. This algorithm lack in efficiency and time and also large number of candidate itemset is generated

Azadeh Soltan et al [2] presented confabulation-inspired association rule mining (CARM) algorithm for mining frequent and infrequent item sets. The presented algorithm holds two phases of knowledge attainment and rule extraction. Knowledge achievement grasps two modules whereas the axonal association links between these two modules are made to archive all domain knowledge. The second phase of rule extraction is then executed derived from the weight age of these communication links. Li-Min Tsaiet al [3] proposed an improved approach called generalized negative association rule (GNAR) algorithm describe negative rules are as important as positive rules. It helps the user to quickly decide which one is important instead of checking all the rules. The main benefit of this algorithm is reduces the computational cost and reduces uninteresting negative rules.

Christian Hidlern et al [4] introduced an algorithm continuous association rule mining algorithm (CARMA) which was introduced to compute large datasets online. It requires two scans of transaction to produce all large itemsets. It also has two pass, the first pass constructs a lattice of all potentially large itemsets continuously. In the second pass it removes all the itemset which is very small user specified threshold which is done continuously. Xiangjun et al [5] suggested another algorithm IMLMS model which is proposed to find the frequent and infrequent itemset. The Minimum correlation Strength is suggested which a correlation coefficient. It shows that it has better performance than the measure interest and its value is very easy to set. But the performance for generating the frequent and the infrequent itemset is not good

The MLMS algorithm mines the FIS and inFIS itemse, the discovered patterns are not much interesting and are raucous and hence it requires pruning. So the existing method used the modified wu's pruning strategy with IMLMS [7] an algorithm was designed to discover interesting frequent and infrequent patterns. This method has some difficulty to set the threshold value. It uses *interest* measure to calculate the interestingness of the itemset the *interest (A,B)* depends on values of support *s(.)*. This method mainly prunes uninteresting patterns. The discovered patterns lack in efficiency and accuracy.The next existing method rectifies the measure *interest* and uses another measure Minimum correlation Strength (MCS) [4]based on correlation coefficient the performance is better than the measure *interest* here the users finds easy to set the values here $\rho(A,B)$ is calculated instead of *interest*(A,B). However the performance improves but this method still lack in accuracy and efficiency. The AMLMS-GA [18] is the algorithm used to generate accurate frequent and infrequent itemset and from this itemset the positive and negative association rule is generated and optimized using the GA (Genetic Algorithm). The GA (genetic algorithm) is an optimization algorithm it classifies the generated dataset based on their relevancy. This algorithm has better performance, accuracy and the efficiency yet the performance and accuracy can be improved much better with our proposed methods and technique. The proposed *Apriori_AMLMS* [20] algorithm generates frequent and infrequent item set which is also pruned by this algorithm and the modified genetic algorithm is later applied. In order to use genetic algorithm these measurements are needed (i.e.) fitness value, crossover and mutation hence it produces and optimized solution.

2. THE PROPOSED SYSTEM

The proposed algorithm is exceptionally useful in the field of Association rule mining area. The proposed work is categorized into three phases 1) Apriori_AMLMS algoritm 2) GPNAR algorithm 3) Advance ABC algorithm. The first phase generates the frequent itemset (FIS) and the inFrequent itemset (inFIS). The generation of the FIS and the inFIS is based on the two important measures of the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

association rule 1) Support 2) confidence in correspond with the user defined threshold value the minimum support(minisuppo) and the minimum confidence(miniconfi). First the Support (Suppo) for each item in the dataset is calculated. If the support value of each and every item is greater than the user-defined threshold value(minisuppo) then the itemset is considered as the FIS otherwise the inFIS. the next phase is the calculation of confidence of the rule.

The generated FIS and inFIS and the user-defined threshold value (miniconfi) becomes the input of the next phase. The correlation of the rule is also is measured in this phase using the *lift* measure. The *lift* value is measured for each and every rule if the *lift* value is greater than one then the rule has a valid positive correlation otherwise negative correlation. The confidence of each generated rule is calculated along with the *lift*, if the confidence of the rule is greater than the miniconfi and *lift* is greater than one then the rule is PAR otherwise NAR. Similarly generate the PAR and NAR from inFIS also. The generated PAR and NAR are in large numbers many are redundant rules. The proposed systems intend is to produce appropriate output so this can be achieved only through optimized algorithm. The next proposed algorithm is advance ABC algorithm a recent optimizer which is a swarm intelligence algorithm. The working of this algorithm is just like the bee hive. The working of this is classified into 1) the initialization process 2) employee Bee 3) onlooker bee 4) Scout Bee. The input for this algorithm is the generated NAR and PAR. The bee hive is initialized first then the foods are stored in each cell. The employee bees collect the food (rule) and store them in their memory. The Each rule have antecedent (left side) and consequent (right side) parts respectively. The Proposed Advance ABC Algorithm assumes consequent part of rule as bees and antecedent part's first attribute as Food. Each rule has occurrence value (fitness value), means how many times that rule occurring in transaction set. This algorithm finds set of rules which have same bees and same food. If a set of rule have same food and same food the best rule is selected which has highest occurrence value in transaction set (i.e.,) that path is mostly followed by bees and it is shortest path. Now the fitness value of each rule is compared with user defined threshold value minimum confidence (miniconfi). If the fitness value of a rule is greater than the miniconfi value then the bee place the food sources in the memory. After applying the onlooker bee phase the generation of new offspring from old offspring is made. For finding new food sources, the scout bee is sent to the search space.

2.1 Proposed System Architecture

The flow diagram describes about the proposed method. Here for the proposed system implementation the balloon dataset is taken for analysis. The figure 1 shows the overall proposed system which has all three proposed algorithm.

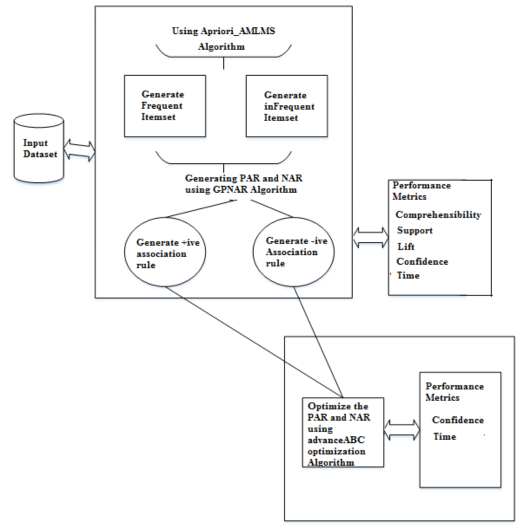

FIGURE 1: Proposed Algorithm flow diagram

The pre-processing step is the data transformation. The data is transformed into a decision table system. Transform transactional database into decision table system, which contains the conditional and decisional attributes. The attributes are compared with the neighbour attributes and arranged in priority manner. After this the rough set theory is applied in the arranged data. This data is taken as input for the first phase. The Apriori_MLMS algorithm now generates the frequent and infrequent itemset from the dataset which is tested for accuracy and time taken to generate. The next proposed algorithm GPNAR is applied to generate the PAR and NAR. The generated rule is measured for comprehensibility, support, confidence and time. Finally the Advance ABC algorithm is applied for optimized PAR and NAR.

2.2 Proposed Apriori_AMLMS Algorithm

A proposed algorithm [20] is used to generate the frequent and infrequent item set in the dataset. The Apriori is incompetent algorithm that uses bottom up search to identify all the frequent itemset from the database. The proposed Apriori_AMLMS algorithm generate both frequent anf infrequent itemset . Let I= $\{i_1, i_2, i_3.....i_m\}$,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

where m is the distinct literals or we call as terms. Let DS be the transaction database (in the form of documents). Each transaction S is set of items S is a subset of I. Each transaction S has a unique identifier S_{ID}. Consider X (antecedent), Y (consequent) are the two set of items where X⇒Y, where $X \subset I$, $Y \subset I$, $X \cap Y = \emptyset$. Here the two eminence measurements are used Support (suppo) and Confident (confi). The Apriori_AMLMS algorithm generates the Frequent itmset FIS) and infrequent itemset(FIS) in varying minisuppo user defined threshold value. As the minisuppo value is increased the generation of Frequent itemset gradually decreases.

2.3 Proposed GPNAR Algorithm

The next is the proposed algorithm is the GPNAR (Generating Positive and Negative Association Rule) for mining positive and negative association rule from the discovered frequent itemset (FIS) and infrequent itemset (inFIS) which is derived from Algorithm 1. The rules are categorized as PAR and NAR based on the confidence and the lift. The lift is the measures the quality of the mined rules.

The GPNAR uses another one measurement for measuring the interestingness of the rule which denoted as *lift* in order to identify the interesting positive and negative association rule. The *lift* value shows the strength or interestingness on the rules generated. Thus if the value of *lift* is greater than one it is consider as strong rule otherwise called as weak rule. Once the Support value is high than the user's suggestion then it is defines as minsuppo, hence only minimum threshold transaction is presented. Once the confidence is greater than user's suggestion then it is defined as minconfi these conditions are possible between X and Y then it is a valid rule. Once the association is a valid one the *lift* measurement identifies the positive and the negative association rule. If the value of *lift* is greater than 1 then it is said as positive rule and if it is less than one then it is negative rule and if the value equals 1 then recognises no association rule generated.

Suppo (X) defines no transaction containing X, suppo(X⇒Y)=Prob(X,Y), denotes the number of transaction where X and Y coalesce. The confidence

$$confi(X \Rightarrow Y) = \frac{prob(X \Rightarrow Y)}{Prob(X)}, \qquad (1)$$

Confidence measure X and Y occurs in the transaction $= \frac{1 + prob(XY) - Prob(X)prob(Y)}{Prob(X)Prob(Y)}$, (2)

each item is denoted $X \subset I$

$$lift(X \Rightarrow Y) = \frac{prob(X,Y)}{prob(X)Prob(Y)}, \qquad (3)$$

the lit measures the strength of the relationship between the X and Y.

$$\text{Comprehensibility} = \frac{\log(1+|Y|)}{\log(1+|X \cup Y|)}. \qquad (4)$$

The comprehensibility describes the clarity of rule. Consider minsuppo(n) is the minimum support of n items where n=1,2,3...and minsuppo (0) is the threshold value.

If suppo (X) ≥minsuppo(number(X)),then the X is frequent itemset

If suppo(X) < minsuppo(number(X), then the X is infrequent itemset

If *lift* (X⇒Y) >1, then it is positive dependency rule

If *lift* (X⇒Y) <1, then it is a negative dependency rule

If *lift* (X⇒Y) =1, then no association rule is generated.

Input
Minisuppo; miniconfi; FIS; inFIS;
Output
NAR (Negative Association Rule)
PAR (Positive Association Rule)

/*algorithm for generating PAR and NAR from FIS
Step1 initialize NAR=∅; PAR=∅;
Step2 For everyitemset I in FIS
Do begin
Each itemset X∪Y=I, X∩Y=∅
Do begin
2.1 If confi (X⇒Y)≥miniconfi && *lift* (X⇒Y) ≥1
Then output (X⇒Y); PAR∪(X⇒Y)
Else
2.2. If confi (X⇒¬Y)≥miniconfi && *lift* (X⇒¬Y) ≥1
Then output (X⇒¬Y); NAR∪(X⇒¬Y)
If confi (¬X⇒Y)≥miniconfi && *lift* (¬X⇒Y) ≥1
Then output (¬X⇒Y); NAR∪(¬X⇒Y)
If confi (¬X⇒¬Y)≥miniconfi && *lift* (¬X⇒¬Y)≥1
Then output (¬X⇒¬Y); NAR∪(¬X⇒¬Y)
End for;
End for;

/*algorithm for generating PAR and NAR from inFIS

For itemset I in inFIS
Do begin
Step 3 for every itemset X∪Y=I, X∩Y=∅
Suppo (X)≥minisuppo and Suppo(Y)≥minisuppo
Do begin
3.1 If confi (X⇒Y)≥miniconfi && *lift* (X⇒Y) ≥1
Then output (X⇒Y); PAR∪(X⇒Y)
Else
3.2. Ifconfi (X⇒¬Y)≥miniconfi && *lift* (X⇒¬Y) ≥1
Then output (X⇒¬Y); NAR∪ (X⇒¬Y)
Ifconfi (¬X⇒Y)≥miniconfi && *lift* (¬X⇒Y) ≥1
Then output (¬X⇒Y); NAR∪ (¬X⇒Y)
Ifconfi (¬X⇒¬Y)≥miniconfi && *lift* (¬X⇒¬Y)≥1
Then output (¬X⇒¬Y); NAR∪ (¬X⇒¬Y)
End for;
End for;
Step 4 Returns PAR and NAR.

Algorithm 1: Proposed Algorithm For Mining PAR And NAR.

The *Lift* is a measure which is used to check the dependency between the association rules. If the *lift* value is greater than 1 then the rule has positive (strong) dependency. If the value of the *lift* is less than 1 then the rule has negative (weak) dependency. So in order to achieve a strong rule the lift value should be 1 or more. The below rule has required minimum confidence and also the lift value is greater than 1 then the rule has positive dependency. Hence we can call the rule as strong NAR. Hence the lift value of the rule is greater than 1 it is considered as positive dependency or strong negative association rule.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Given
[1] $Suppo(X \cup Y) < mintsuppo$,and $Suppo(X \cup Y) \neq \emptyset$
$Suppo(X) \geq mintsuppo\ and\ Suppo(B) \geq mintsuppo$
[2] $If\ Conft(X \Rightarrow Y) \geq mintconft\ and\ lift(X \Rightarrow Y) > 1\ then\ X \Rightarrow Y$ is a valid PAR
[3] Where the value of *lift* is greater than 1 it has positive dependency or strong PAR
[4] $Else\ if\ conft(X \Rightarrow \neg Y) \geq$ minlconfi $and\ lift > 1$ then
[5] $(X \Rightarrow \neg Y)$ Is a strong NAR

Given
[6] $Suppo(X \cup Y) < mintsuppo, Suppo(X) > mintsuppo\ and\ Suppo(B) > mintsuppo$
[7] $If\ Conft(X \Rightarrow Y) \geq mintconft\ and\ lift(X \Rightarrow Y) > 1\ then\ X \Rightarrow Y$ is a valid PAR
[8] Where the value of *lift* is greater than 1 it has positive dependency or strong PAR
[9] $Else\ if\ conft(X \Rightarrow \neg Y) \geq$ minlconfi $and\ lift > 1$ then
[10] $(X \Rightarrow \neg Y)$ Is a strong NAR

Algorithm 2: Proposed Algorithm For Correlation Measure.

2.4 Advance ABC algorithm:

In general the ABC optimization is

Initialization Phase
REPEAT
Employed Bees Phase
Onlooker Bees Phase
Scout Bees Phase
Memorize the best solution achieved so far
UNTIL (Cycle = Maximum Cycle Number)

The advance ABC algorithm is a swarm intelligence algorithm. There are various steps in this algorithm

Step 1: Input is the generated PAR and NAR and the minicofidence.
Step 2: Initialize population using ABC on selected members to discover association rules.
Step 3: Find each association rule fitness function.
Step 4: Check following condition: If (fitness function> miniconfi).
Step 5: Set Q=Q∪(x=>y) /* the rules are added to the temporary variable Q

5.1 In memory, employed bees are placed on food sources;

5.2 Generate new offspring from older offspring after applying onlooker bee phase.

5.3 For finding new food sources, send scout bee to search space.

Step 6: UNTIL (requirements are not met).

2.4.1 Employee Bee:

The preliminary process is the initialization process. First, we initialize the positions of 3 food sources (CS/2) of employed bees, (50, 500) are the range of uniform distribution and they are randomly utilized.

$$y_j^k = y_{\min}^k + rand[0,1] * (y_{\max}^k - y_{\min}^k) \quad (5)$$

Where $y_{\max}^k$ the upper is bound for y_j^k and $y_{\min}^k$ is the lower bound for y_j^k. The $j = 1, 2, \cdots, M\ and\ k = 1, 2, \cdots, D$.

y_j^k is a parameter to be optimized for the jth employed bee on the dimension k of the D-dimensional space. Number of employed bee denoted as M.

Employed bees are look for the local optimization value in the neighbourhood of food source is performed in the employed bee phase. The fitness function is calculated using the particular formula

$$F(j) = \begin{cases} \dfrac{1}{(1 + obj(i))} & if\,(obj(i)) \geq 0 \end{cases} . \quad (6)$$

The *obj* value is rule based on the support, confidence, and comprehensibility. This *obj* value is considered as the food.

After Number of iterations is used for calculate the Fitness function, this iteration will be continuing until the best optimized value is got. Based on the optimized fitness values the bees stores new food in the memory. The new position is calculated by

$$v_j^k = y_j^k + \psi_j^k \left(y_j^k - y_i^k \right) \quad (7)$$

Where, $k = 1, 2, \cdots, D$ and $i = 1, 2, \cdots, M$. In the above equation, y_j^k is the j^{th} employed bee, v_j^k is the new solution for y_j^k, y_i^k is the neighbour bee of

Journal of Theoretical and Applied Information Technology
31st December 2017. Vol.95. No 24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

y_j^k in employed bee population, [-1, 1] is the range of φ and it is randomly selected, D is the dimension of the problem and M denotes the number of employed bee. In the above equation k and i values are selected randomly and it memorize best solution.

2.4.2 Onlooker bee:

Unemployed honey bees comprise of two honey bees gatherings: onlooker honey bees and scouts. Employed honey bees share their food source data to onlooker honey bees holding up in the hive and afterward onlooker honey bees probabilistically pick their food sources relying upon this data. In ABC, an onlooker honey bee picks a food source relay upon probability values figured utilizing fitness values gave by utilized honey bees. For this reason, a fitness based determination strategy can be utilized, for example, roulette wheel selection technique. Onlooker bee can be calculated by using the expression

$$P_m = \frac{F(j)}{\sum_{m=1}^{SN} F(j)} \qquad (8)$$

$F(j)$ Denotes fitness value, SN denotes swarm size.

2.4.3 Scout Bee

The unemployed honey bees who pick their food sources arbitrarily are called scouts. Employed honey bees whose arrangements can't be enhanced through a foreordained number of trials, determined by the client of the ABC calculation and called "limit" or "abandonment criteria" in this, get to be scouts and their answers are surrendered. At that point, the changed over scouts begin to scan for new solutions, randomly. For instance, if solution y_j^k has been abandoned, the new solution discovered by the scout who was the employed bee of y_j^k can be defined. Hence those sources which are initially poor or have been made poor by exploitation are abandoned and negative feedback behaviour arises to balance the positive feedback.

3. EXPERIMENT RESULT

For the proposed method Apriori_AMLMS based Advance ABC algorithm the analysis is done from the balloon database for experimentation the number of information used is five attributes with 3200 instances. The experimental analysis first shows how the frequent and infrequent itemset are generated by the support (minsuppo) value as input for the dataset DS.

Table 1: Accurate FIS and inFIS from various minisuppo

minisuppo	Frequent itemset	Infrequent itemset
0.2	353	434
0.5	365	562
0.7	234	754
1	143	834

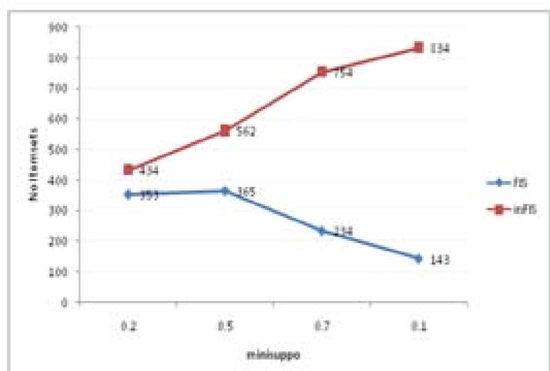

Figure 2: Frequent and inFrequent item

From the experimental result it is clearly seen that there is a gradual decrease in the frequent itemset set as the user defined threshold value minisuppo is increased. This shows that the candidate itemset generation is decreased and database scanning is also decreased. Hence automatically the space is also decreased. The time to generate the FIS and inFIS is also reduced using the proposed Apriori_AMLMS algorithm when compared with the existing algorithms. The rules generated by the proposed algorithm are very huge and many rules are redundant so to achieve optimized high confidence rule, the optimized Advance ABC algorithm is applied. The result analysis shown proves that the achieved rule is well optimized rule.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2: Overall result analysis of the proposed algorithm

minisuppo	miniconfi	FIS		inFIS		comprehensibility	confidence	lift	support
		PAR	NAR	PAR	NAR				
0.2	0.6	243	655	213	553	0.6309	0.75	1.5	0.3
0.5	0.65	162	432	236	544	0.6732	0.65	1.3	0.25
0.7	0.7	223	563	221	672	0.5342	0.72	1.4	0.3
1	0.8	212	435	256	653	0.6309	0.75	2	0.25

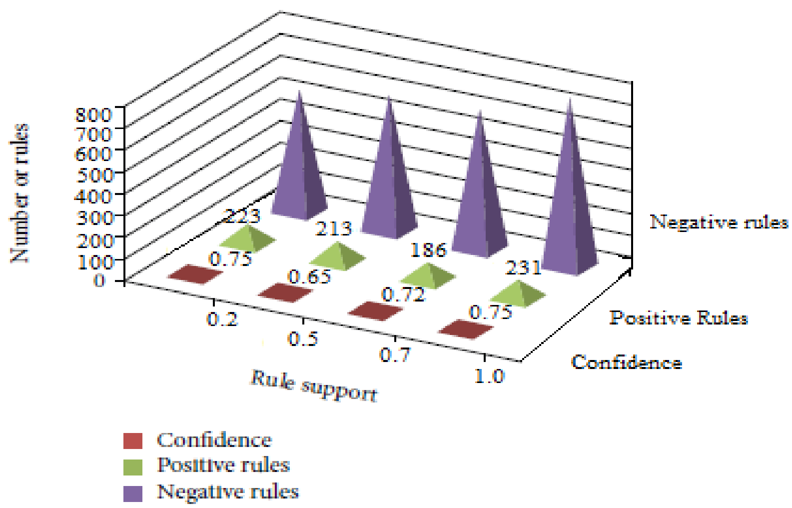

Figure 3: Valid Positive and Negative Association Rule generated using proposed algorithm

Table 3: Optimized PAR and NAR

Miniconfi	Fittness value	Confidence	Apriori_AMLMS and GPNAR		Adv.ABC	
			PAR	NAR	PAR	NAR
0.6	0.6983	0.7	243	655	132	455
0.65	0.6943	0.75	162	432	102	301
0.7	0.7412	0.9	223	563	84	183
0.8	0.8453	0.95	212	435	56	230

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

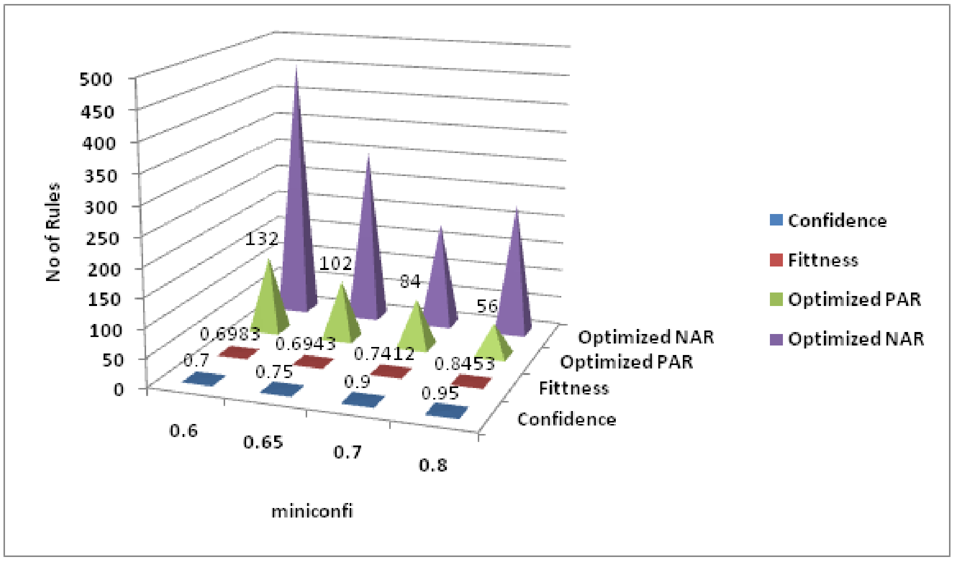

Figure 4: Optimized PAR and NAR with confidence and fitness value

4. CONCLUSION

Mining negative association rule has now got a great interest among the researchers. The proposed method generates both optimized positive and negative associations rule with high confidence, and reduced time is the major assistance. The proposed Advance ABC optimizer produces an efficient optimized rule. It is clearly demonstrated through the experimental analysis that the proposed method is very accurate and provides a very good performance and more efficient. In future this algorithm is applied in various dataset to generate different rule with high quality and confidence with less time.

REFERENCES:

[1] Nikky Surya wanshi Rai, Susheel Jain, Anurag Jain," Mining Interesting Positive and Negative Association Rule Based on Improved Genetic Algorithm" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 5, No. 1, 2014.

[2] AzadehSoltani and M.-R.Akbarzadeh-T," Confabulation-Inspired Association Rule Mining for Rare and Frequent Itemsets", IEEE transactions on neural networks and learning systems, 2014.

[3] Li-Min Tsai_; Shu-Jing Lin_; Don-Lin Yang, "Effective mining of generalized negative association rules", IEEE conference Publication, 2010.

[4] Christian Hidber. Online Association rule mining. SIGMOD '99 Philadelphia PA. ACM 1-58113-084-8/99/05, 1999.

[5] Li-Min Tsai_; Shu-Jing Lin_; Don-Lin Yang, "Effective mining of generalized negative association rules", IEEE conference Publication, 2010.

[6]XiangjunDong,ZhendongNiu,DonghuaZhu,ZhiyunZheng, QiutingJia, "Mining Interesting infrequent and frequent Itemset based on MLMs Model" Springer, 2008, pp 444-451.

[7] Xiangjun Dong, '"Ming Interesting Infrequent and Frequent Itemset Based on Minimum Correlation Strength", Springer, 2011, pp 437-443.

[8]Mrs.K.Mythili, MrsK.Yasodha," A Pattern Taxonomy Model with New Pattern Discovery Model for Text Mining"International Journal of Science and Applied Information Technology Volume 1, No.3, ISSN No. 2278-3083July – August 2012.

[9] CharushilaKadu, Praveen Bhanodia, Pritesh Jain, "Hybrid Approach to Improve Pattern Discovery in Text mining"International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 6, June 2013.

[10] Zhen Hai, Kuiyu Chang, Jung-Jae Kim, and Christopher C. Yang," Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance", IEEE Transactions on

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

knowledge and data engineering, vol. 6, no. 6, June 2012.

[11] Spyros I. Zoumpoulis, Michail Vlachos, Nikolaos M. Freris, Claudio Lucchese, “Right-Protected Data Publishing with Provable Distance-based Mining “ IEEE Transactions on knowledge and data engineering, vol. 21, no. 19, november 2012.

[12] K. Aas and L. Eikvil, “Text Categorisation: A Survey,” Technical Report Raport NR 941, Norwegian Computing Center, 1999.

[13] R. Agrawal and R. Srikant, “Fast Algorithms for Mining Association Rules in Large Databases,” Proc. 20th Int’l Conf. Very Large Data Bases (VLDB ’94), 1994,pp. 478-499.

[14],MissDiptiS.Charjan, Prof. MukeshA.Pund ,” Pattern Discovery For Text Mining Using Pattern Taxonomy”.International Journal of Engineering Trends and Technology (IJETT) – Volume 4 Issue 10- October 2013.

[15] J. Han and K.C.-C. Chang, “Data Mining for Web Intelligence,”Computer, vol. 35, no.11, Nov. 2002,pp. 64-70.

[16] J. Han, J. Pei, and Y. Yin, “Mining Frequent Patterns without Candidate Generation,” Proc. ACM SIGMOD Int’l Conf. Management of Data (SIGMOD ’00), 2000,pp. 1-12.

[17] RupeshDewang, JitendraAgarwal, ”Anew Methos for Generating All Positive and Negative Association rules” International Journal on computer Science and Engineering Volume 4, No.4, ISSN No. 0975-3397 April 2012.

[18] I.Berin Jeba Jingle and Dr.J.Jeya A.Celin, “Markov Model in Dicovering knowledge in Text mining”, Journal of Theoretical and Applied Information Technology”vol.70,no.3,Dec 2014,pp.459-463.

[19]I.Berin Jeba Jingle and Dr .J. Jeya A.Celin” Discovering Useful Patterns in Text Mining using AMLMS-GA algorithm, ”International Journal of Applied Engineering Research” vol.10, no.18, 2015,pp.39763-39767.

[20] Dr .J. Jeya A.Celin and I.Berin Jeba Jingle” Mining Useful Patterns from Text using Apriori_AMLMS-MGA Algorithm” International Journal of Applied Engineering Research” vol.10, no.18, 2015,pp.39763-39767.