

# Improved Association Rule Mining based on ABC

Sourabh Sahota  
Dept. of CSE  
CT Institute of Engg., Mgt & Tech  
Punjab (India)

Prince Verma  
Dept. of CSE  
CT Institute of Engg., Mgt & Tech  
Punjab (India)

## ABSTRACT

Association rule mining is virtually importance and its use is one of a essential method for data mining. The association rule mining approach significant have been with many minute changes in the apriori although their fundamental opinion proceed same i.e use of support and confidence threshold(s). This paper to find that there is no tasks that have been done in the region of E-Apriori. In this paper have to introduce new algorithm Enhance Apriori i.e(E-Apriori).The E-Apriori algorithm is advance to Enhance the Apriori algorithm by using the median support (supmedian) alternatively of minimum support, to deliver probabilistic item-set alternatively of large item-set. In this paper for optimization the rule with the help of ABC technique i.e (Artificial Bee Colony) and E-Apriori and Apriori algorithm situated on ABC technique (Artificial Bee Colony).

## Keywords

Data Mining ,KDD Process, ABC (Arificial Bee Colony), Association Rule Mining, Apriori , E-Apriori.

## 1. INTRODUCTION

Data mining is called exploratory data analysis, on other things. Data mining is necessary to identification of a problem, along with collection of data that can produce the exceptional understanding, and computer model to give statistical or other means of analysis. This may be based on decision tools, that show data or over vital statistical analysis, such as correlation analysis.

Data mining relate to statistical and or artificial intelligence analysis commonly applied to large-scale datasets. Data mining relate to the resolve of knowledge discovery (study new and useable things) knowledge discovery classify the unsupervised (knowledge Discovery) much of this can be established over automatic means, as see in decision tree analysis. But data mining is not defined to automated analysis. Knowledge Discovery by humans can be enhanced by graphical tools and diagnosis of unpredicted patterns through a accumulation of human and computer communication.

## 1.2 Association Rule Mining

Association rules are generally used in the peddle industry supporting the name “Market Basket Analysis”. In Market Basket Analysis an association rule comprise a set of item that are possible to be acquire together, For example, The rule {Honey}→{Bread , Milk} would state that when a purchaser acquire Honey, he or she is probably to purchase both Bread and Milk as well in the same transaction. Association Rule Mining is virtually vital and factual research method of data mining it was first presented by [Rakesh Agrawal in 1993]. Its target is to obtain impressive statistics frequent patterns, associations or informal structures surrounded by sets of items in the transaction databases or other data depositary. Support and Confidence are important parameters of ARM (Association Rule Mining).

## 1.3 Apriori Algorithm

APRIORI algorithm aims to expose hidden knowledge it is the extensive objective of the data mining. Agrawal et al. was first introduce the APRIORI algorithm in 1993. APRIORI algorithm is extant useful during the candidate production growth. Breath first search pattern is used for the APRIORI algorithm to consider the support of item sets and uses a candidate production function which use the descending closure property. In database the APRIORI algorithm easy to perform and identical use to exploit all frequent item-sets. Enhance techniques used to APRIORI to abide the measure definite item sets, although secure integrity. APRIORI algorithm intentionally manipulate on databases including transactions. APRIORI generate the advantage to decrease the number of items being advised by only prospecting the item-sets whose support compute is larger than the minimum support compute. But already there is drawback that it expend more time, space and memory for item production process and for achieving the sets it expect many scans over the database. so that, it can use few enhancement in it, the enhancement will over come its drawback.

### 1.3.1 Example For Apriori Algorithm

#### Apriori Explanation

As Apriori algorithm produce the candidate item-sets without seeing the transaction in the database. For Example: we are provide with a database D (Table 1) with some set of transaction, supmin =61% and conf min = 70%

Table 1: Database D

Transaction	Items	Transaction	Items
11	C,D,E	21	A,B,C,D
12	A,B,C,D	22	C,D,E
13	E	23	A,B,C,D
14	A,B,C,D	24	E
15	A,B,C,D	25	A,B,C,D
16	A,B,D	26	A,B,C,D,E
17	A,B,C,D,E	27	A,B,D,E
18	C,D	28	C,D,E
19	A,B,C,D	29	A,B,C,D,E
20	B,C,D	30	A,B

Before creating large item-sets (L1, L2 & so on) and candidate item-sets (C2 & so on) candidate, firstly Candidate item-set, C1 is generated from database (i.e. Table 2 from Table 1) and then, Large item-set, L1 is created from C1 using supmin=61% (i.e. Table 3 from Table 2)

Table 2: Candidate Item-set C1

Candidate Item-set, C1	Support
A	65%
B	70%
C	75%
D	80%
E	55%

**Table 3: Large Item-set L1**

Large Item-Set, L1	Support
A	65%
B	70%
C	75%
D	80%

Step 1: Candidate set is produce as follows:

- The candidate item-sets, C<sub>k</sub> can be produce by joining large item-sets L<sub>k-1</sub> items, and
- Deleting those that comprise any subset that is not large.

In Example C<sub>2</sub> is produce from L<sub>1</sub> items by join procedure (i.e. Table 4 from Table 3) and those item-sets are deleted that have some (k-1) subset of c is not in L<sub>k-1</sub> where c ∈ C<sub>k</sub>.

Step 2: Large item-set, L<sub>k</sub> is produce form candidate item-set, C<sub>k</sub> using sup<sub>min</sub>.

The above two steps are repeated until Large Item-set came to be empty.

Here from the candidate item-sets, C<sub>2</sub> elements with sup ≥ sup<sub>min</sub>, Large item-set, L<sub>2</sub> is created (Table 5). As all candidate item-sets, C<sub>3</sub> has sup ≥ sup<sub>min</sub>, so all became Large item-set, L<sub>3</sub> (Table 5 from Table 6)

**Table 4: Candidate Item-set C2**

Candidate Item-set, C1	Support
A,B	65%
A,C	50%
A,D	55%
B,C	60%
B,D	60%
C,D	70%

**Table 5: large Item-set L2**

Large Set L2	Support
A,B	65%
C,D	70%

Large Item-set: (A,B), (C,D) & Association Rules are:

**Table 6: Association Rules**

Large Set	Association Rules	Confidence
A,B	A=>B	65/65=100%
	B=>A	65/70=92.8%
C,D	C=>D	70/75=93.3%
	D=>C	70/80=87.5%

## 1.4 Enhance Apriori Algorithm

In work have been observe that the fundamental problem for APRIORI algorithm. Introduce a advance method to enhance the APRIORI algorithm in better way. Proposed Enhance APRIORI algorithm confine the drawback of original APRIORI algorithm. Enhance APRIORI algorithm is describe in such a way. This algorithm perform completely and efficiently. Proposed Enhance APRIORI algorithm proceed less time and minimum number of scans to generate frequent. The Propose algorithm focuses in enhance the APRIORI algorithm by using average support (sup avg) alternatively of minimum support (sup<sub>min</sub>) to create probalistic item-sets

alternatively of large item-sets. Here average support is not user specify value but it is calculated using the formula

$$supavg = \frac{\sum_{k=1}^m (sup^i(k))}{m}$$

Where m is the number of items elaborate. So this formula would lead to better item-sets and this increase the number of superior association rules that were left trans value by APRIORI algorithm.

### 1.4.1 Example For Enhance Apriori Algorithm

The prospective algorithm target to Enhance the process of Apriori algorithm by using mean support (sup<sub>mean</sub>) instead of minimum support (sup<sub>min</sub>), to produce probabilistic item-set instead of large item-set.

For Example: we are provided with a database D (Table 1) with some set of transaction and conf<sub>min</sub>=70%.

Before crating probability item-sets (P<sub>1</sub>,P<sub>2</sub>& so on) and candidate item-sets (C<sub>2</sub>, C<sub>3</sub> & so on) candidate, firstly candidate item-set, C<sub>1</sub> is generated from database (i.e. Table 8 form Table 7).

Then, Would-be Probability item-set, P<sub>1</sub> is created from C<sub>1</sub> (i.e. Table 9 from Table 8). The item-sets with sup ≥ sup<sub>mean</sub> are inserted into probability item-set. i.e.

$$Using\ supavg = \frac{65 + 70 + 75 + 80 + 55}{5} = 69\%$$

**Table 7: Database D**

Transaction	Items	Transaction	Items
11	C,D,E	21	A,B,C,D
12	A,B,C,D	22	C,D,E
13	E	23	A,B,C,D
14	A,B,C,D	24	E
15	A,B,C,D	25	A,B,C,D
16	A,B,D	26	A,B,C,D,E
17	A,B,C,D,E	27	A,B,D,E
18	C,D	28	C,D,E
19	A,B,C,D	29	A,B,C,D,E
20	B,C,D	30	A,B

**Table 8: Candidate Item-set C1**

Candidate Item-set, C1	Support
A	65%
B	70%
C	75%
D	80%
E	50%

**Table 9: Probability Item-set P1**

Probable Item-set, P1	Support
B	70%
C	75%
D	80%

Now emphasis are performed on item-sets with sup ≤ sup<sub>mean</sub>, item-set and sup ≥ Sp (here Sp = sup<sub>mean</sub>/1.2 = 69/1.2=57.5%) i.e. item A. So, items with sup ≥ Sp is {T<sub>1</sub>} with sup=65%). From these, items with probability > 80% are taken into Probability Item-set. i.e. If items with sup ≥ Sp have maximum sup, say Sp-max then items with support greater than (80% of Sp-max) are inserted into probability item-set, while others are pruned. (here Sp-max=65% and

80% of Sp-max=52%). So, Probability Item-set P1 is given in Table 10.

**Table 10: Probability Item-set P1**

Probable Item-set, P1	Support
A	65%
B	70%
C	75%
D	80%

Candidate item-set C2 is generated from P1 (Table 9 from Table 10). Here from the candidate item-sets, C2 elements with  $sup \geq sup_{min}$ , Probability item-set, P2 is created (Table 12 from Table 11). And for probability item-set generation same process as above is repeated.

**Table 11: Candidate Item-set C2**

Candidate Item-set, C2	Support
A,B	65%
A,C	50%
A,D	55%
B,C	60%
B,D	60%
C,D	70%

**Table 12: Probability Item-set P2**

Probable Item-set, C2	Support
A,B	65%
A,D	55%
B,C	60%
B,D	60%
C,D	70%

**Table 13: Candidate Item-set C3**

Candidate Item-set, C3	Support
A,B,D	55%
B,C,D	50%

From here P3 is created containing (A,B,D)

As its  $sup > 52\%$  and (B,C,D) is pruned (see Table 14).

**Table 14: Probability Item-set P3**

Probable Item-set, C3	Support
A,B,D	55%

Probable Item-set: {(A,B), (B,C), (B,D), (C,D), (A,B,D)}. So, (B,C), (B,D) are more in Probable item-set than large item-set of Apriori algorithm in Section and association rule are prescribed in Table 15.

**Table 15: Association Rules**

Large Set	Association Rules	Confidence
A,B	A=>B	65/65=100%
	B=>A	65/70=92.8%
A,D	A=>D	55/70=78.5%
	D=>A	55/80=68.7%
B,C	B=>C	60/70=85.7%
	C=>B	60/75=80%
B,D	B=>D	60/70=85.7%
	D=>B	60/80=75%
C,D	C=>D	70/75=93.3%
	D=>C	70/80=87.5%
A,B,D	(A,B)=>D	55/65=84.6%
	(A,D)=>B	55/55=100%

$$(B,D) \Rightarrow A \quad 55/60=91.6\%$$

## 1.5 Artificial Bee Colony (ABC)

In recent year, the researcher are attractive to the Swarm intelligence. who work in the respective research domain. Artificial Bee Colony (ABC) introduce by Karaboga in 2005 for actual parameter optimization, just now introduced optimization algorithm and assume the foraging behavior of bee colony. It categorized the braches in evolutionary process. Swarm intelligence determines the measure that present the corporate behavior of social insect colonies or some other animal societies to design algorithm or divide the difficulty – resolving devices. Mostly swarm intelligence for algorithm are practically implemented to solve optimization problems. Artificial Bee Colony achievement are analyzed in 2007. The Artificial Bee Colony algorithm is improved to observe the behavior of the bees on search food source, called the nectar and distribute the knowledge of food sources to the bees in the nest. In the Artificial Bee Colony, the artificial workers are categorized into three types.

1. The Employed Bee.
2. The Onlooker Bee.
3. Scout.

All of them play particular role in the procedure,

The employed bee come to food source and supply the neighborhood of the source in its memory: the onlooker gets the knowledge of food source from the employed bee in the hive and select one of the food source to collect the nectar; and the scout is responsible for discover new food, the new nectar, source.

### 1.5.1 Designing Artificial Bee Colony (ABC)

#### Optimization for Apriori & E-Apriori.

Artificial Bee Colony is to implemented on ARM's Apriori algorithm for optimization of results. So, we implemented the Artificial Bee Colony on Apriori & then on E-Apriori for the better optimization of results. The approach utilized by us to implement ABC over Apriori & E-Apriori is explained as follows. Each rule have antecedent (left side) and consequent (right side) parts respectively. We assume consequent part of rule as bees and antecedent part's first attribute as Food. Each rule has occurrence value means how many times that rule occurring in transaction set. Here we would find those who have same bees and same food. E.g Suppose we see first rule, second rule and third rule that has same food and same bees then among them that rule is best which have highest occurrence value in transaction set means that path is mostly followed by bees and it is shortest path. In this way we obtain various groups which have same food and same bee. The occurrence of each rule is assumed as fitness value for bees. For the each group's best rule is found based on its occurrence. At the end we get only those rules which are optimized by Artificial Bee Colony.

## 2. EXPERIMENTAL RESULTS

To estimate the efficiency of our prospective algorithm we have broadly studied the performance of our algorithm (E-Apriori) in comparison the Apriori algorithm. The parameters for comparison between Apriori, Apriori with ABC(Artificial Bee Colony) and E-apriori, E-apriori with ABC(Artificial Bee Colony) would be:

- i. Minimum Maximum & Median confidence .
- ii. Time Consumed
- iii. Number of Rules Generated.
- a) Total
- b) With confidence $\geq$ 0.9
- c) With confidence $\geq$ 0.8
- d) With confidence $\geq$ 0.7
- e) With confidence $\geq$ 0.7

The test dataset to be used by us for the objective of comparison with the Apriori

Algorithm is accurate and recognized i.e. Smoking Dataset. We use these datasets for the experiment.

### 2.1 Fitting Contact Lenses Dataset.

The Dataset helps to know that patient should be fitted with hard contact lenses, soft contact lenses or none.

- a) Data set Information: The Fitting Contact Lenses Problems is summarized in Cendrowska, J. "PRISM: An algorithm for inducing modular rules", International Journal of Man-machine Studies, 1987,pp-349-370.
- b) Donor: Benoit Julien
- c) Date Donated: 01-08-1990
- d) Attribute Information
  - i. age: young, pre-presbyopic, presbyopic.
  - ii. spectacle-prescription: myope, hypermetrope
  - iii. astigmatism: no, yes
  - iv. tear-prod-rate: reduced, normal
  - v. contact-lenses: soft, hard, none
- e) Results:

Fitting Contact Lenses Dataset are shown in Table (Min-Max Confidence, Total Rules & Time required) & Table (Rules with Conf $\geq$ 90%, Conf $\geq$ 80, Conf $\geq$ 70, Conf $\geq$ 70).

From the results we came to know that Minimum , Maximum & Average confidence is same in case of E-Apriori as some cases of apriori Algorithm (with minsup = 0.07, 0.08, 0.09).The rules generated in E-Apriori are also same as these above defined cases of apriori. But in E-apriori is better in case than Apriori when minsup=0.06. Also we don't need to find the minimum support for finding the best rules in E-Apriori as in the case of Apriori.

This Table show the Fitting Contact Lenses Dataset Results Before ABC.

**Fitting Contact Lenses Dataset Results before ABC**

Fitting Contact Lenses Dataset	Apriori				E-Apriori
	Min sup=0 .06	Min sup=0 .07	Min sup=0 .08	Min sup=0 .09	Median sup=0 .0644
Confmin	0.0666	0.1333	0.1333	0.1333	0.1333
Confmax	1	1	1	1	1
Conf median	0.3545	0.4204	0.4204	0.4204	0.4204

Total Rules	2682	966	966	966	966
Time Required (Milliseconds)	1236	455	356	417	1929
Total Time Required(Milliseconds)=2464					

**Number of Rules (Fitting Contact Lenses Dataset) before ABC**

Fitting Contact Lenses Dataset	Apriori				E-Apriori
	Min sup=0 .06	Min sup=0 .07	Min sup=0 .08	Min sup=0 .09	Median sup=0 .0644
Rules with conf $\geq$ 0.9	233	83	83	83	83
Rules with conf $\geq$ 0.8	6	6	6	6	6
Rules with Conf $\geq$ 0.7	19	19	19	19	19
Rules with conf $\geq$ 0.7	2424	858	858	858	858

**This Table Show the Fitting Contact Lenses Dataset Results After ABC.**

**Fitting Contact Lenses Dataset Results after ABC**

Fitting Contact Lenses Dataset	Apriori				E-Apriori
	Min sup=0 .06	Min sup=0 .07	Min sup=0 .08	Min sup=0 .09	Median sup=0 .0644
Confmin	0.1333	0.1333	0.1333	0.1333	0.1333
Confmax	1	1	1	1	1
Conf median	0.4078	0.4078	0.4078	0.4078	0.2663
Total Rules	610	610	610	610	310
Time Required (Milliseconds)	7426	1786	1852	1680	940
Total Time Required(Milliseconds)=12744					

**Number of Rules (Fitting Contact Lenses Dataset) after ABC**

Fitting Contact Lenses Dataset	Apriori				E-Apriori
	Min sup=0 .06	Min sup=0 .07	Min sup=0 .08	Min sup=0 .09	Median sup=0 .0644
Rules with conf $\geq$ 0.9	43	43	43	43	10
Rules with	11	11	11	11	11

conf $\geq$ 0.8					
Rules with Conf $\geq$ 0.7	19	19	19	19	3
Rules with conf $\geq$ 0.7	537	537	537	537	319

### 3. CONCLUSION

In this paper, The performance of the Enhance Apriori algorithm with ABC (Artificial Bee Colony) has been compared with original Apriori algorithm with ABC (Artificial Bee Colony). The Enhance Apriori with proposed approach is unique in several process that reducing the time consumed in transaction scanning for candidate itemsets by reducing the number of transaction to be scanned. In addition, the simulation results show that the proposed technique converges better quality solutions in case of confidence (max, min or average) and in number of rules (total rules, rules with confidence >90%). In the future our focus is to implement the ABC (Artificial Bee Colony) for ECLAT algorithm to produce better quality solutions.

### 4. REFERENCES

- [1] Ke Wang, Mircea Stan, Kevin Skadron, "Association Rule Mining with the Micron Automata Processor", IEEE International Parallel & distributed Processing Symposium (IPDPS 2015)
- [2] Brijendra Dhar Dubey, Mayank Sharma, Ritesh Shah, "Comparative Study of Frequent Item Set In Data Mining", International Journal of Programming Languages and Applications (IJPLA) Vol.5, No.1, January 2015
- [3] Narges Jamshidian Ghalehsefidi, Mohammad Naderi Dehkordi, "A Survey on privacy preserving association rule mining", ACSJ Advances in Computer Science: an International Journal, Vol 4, Issue 2, No.14, March 2015
- [4] Heydar Jafarzadeh, Rouhollah Rahmati Torkashvand, "Provide a New Approach for Mining Fuzzy Association Rules Using Apriori Algorithm", Indian Journal of Science and Technology, vol 8(8), 707-714, April 2015
- [5] Ms. Dhara Patel, Prof. Ketan Sarvkar, "Algorithms for Frequent Pattern Mining An Analysis", International Journal of Enhanced Research in Science Technology & Engineering ISSN:2319-7463 Vol. 4 Issue 4, April-2015
- [6] Wael Zakaria, Yasser Kotb and Fayed Ghaleb, "MCR-Miner: Maximal Confident Association Rules Miner Algorithm for Up/Down-Expressed Genes", Appl Math Inf Sci 8, No.2, 799-809 (2014)
- [7] Jaishree Singh, Hari Ram, Dr. J.S. Sodhi "Improving Efficiency of Apriori Algorithm Using Transaction Reduction" International Journal of Scientific and Research Publications, vol. 3, issue 1, January 2013
- [8] R. Patel Numisha Sheetal Mehta, "A Survey on Mining Algorithms" International Journal of Soft computing and Engineering, Vol. 2, issue 6, 2013, pp 460-463
- [9] Suhani Nagpal "Improved Apriori Algorithm using logarithmic decoding and pruning" International Journal of Engineering Research and Applications, vol. 2, issue 3, pp. 2569-2572, May-Jun 2012.
- [10] Y. Ramamohan, K. Vasantharao, C. Kalyana Chakravarti, A.S.K. Ratnam "A Study of Data Mining tools In Knowledge Discovery Process", International Journal of Soft Computing and Engineering (IJSCE) ISSN:2231-2307, Volume-2, Issue-3, July 2012
- [11] Sheila A. Abaya, "Association Rule Mining Based On Apriori Algorithm in Minimizing Candidate Generation", International Journal of Scientific & Engineering Research Volume 3, Issue 7, July-2012
- [12] S.A. Abaya "Association Rule Mining Based on Apriori Algorithm in Minimizing Candidate Generation" International Journal of Scientific & Engineering Research, vol-3, issue 7, 2012
- [13] S. Vijayarani M. Sathiyaa Prabha, "Association Rule Mining Using Artificial Bee Colony Algorithm", International Journal of Computer Application (1975-8887), Volume 33-No.2, November 2011.
- [14] Farah Hanna AL-Zawaidah, Yosef Hasan Jbara and Marwan AL-Abed Abu-zanana, "An improved Algorithm for mining Association rule in large database" World of Computer and Information technology, Vol. 1, no. 7, 2011, pp 311-316
- [15] Zhuang Chen, Shibang Cai, Quilin Song Chonglai Zhu, "An Improved Apriori Algorithm based on pruning Optimization and transaction reduction" IEEE transaction on evolution computation, 2011, pp 1908-1911.
- [16] Wei Yang Quing, Yang Ren Hua, Liu Pei Yu "An improved apriori algorithm for association rule mining" IEEE International Symposium, Pg 942-946, Aug 2009.
- [17] Huan Wu, Zhigang Lu, Lin Pan, Rong Seng XU and Wenbao jiang "An improved Apriori based algorithm for association rule mining" IEEE Sixth international conference on fuzzy systems and knowledge discovery, pp 51-55, 2009.
- [18] Dongme Sun, Shaohua Teng, Wei Zhang, Haibin Zhu "An Algorithm to Improve the Effectiveness of Apriori" IEEE Conference, pp 385-390, Aug 2007.
- [19] Pratibha Mandave, Megha Mane, Prof. Sharada Patil, "Data mining using Association rule based on APRIORI algorithm and Improved approach with illustration", International Journal of Latest Trends in Engineering and Technology (IJLTET).
- [20] PEI-WEI TSAL, JENG-SHYANG PAN, "Enhanced Artificial Bee Colony Optimization", International Journal of Innovative Computing, Information and control, VOL 5, 12-Dec-2009.