ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 9, September 2013

Analysis of MFGA based on Support Count

Mrinalini Rana[1] and P S Mann[2]

Research Scholar, D.A.V Institute of Engineering and Technology, Jalandhar, Punjab, India [1]

Asstt. Professor, D.A.V Institute of Engineering and Technology, Jalandhar, Punjab, India [2]

Abstract: This paper presents a Genetic algorithm based association rule mining in which multi fitness functions are used. Genetic algorithm is used for performing global search. This proposed algorithm generates intersecting association rules from dataset. A fitness function with parameter support is defined for generating frequent itemsets and then other parameters like confidence, lift, leverage etc are used for defining next fitness function for generating association rules. The proposed algorithm is compared with classical Apriori algorithm and also with existing Genetic algorithm for association rule mining on the basis of metrics Support Count, and comparisons are also made on different generations

Keywords: Multi-Fitness Function Genetic algorithm (MFGA), Apriori algorithm, Genetic Algorithm, Crossover Probability , Fitness function, Support count, Confidence, Lift, Leverage, Coverage.

1. INTRODUCTION

The evolutionary algorithms consist of many stochastic algorithms that based on Darwinian Theory. Genetic algorithm is one of the evolutionary algorithm can be used for association rule mining. Genetic algorithm is better than Apriori because Genetic algorithm performs global search which is not present in Apriori algorithm.

This paper presents proposed algorithm Multi-Fitness function Genetic algorithm (MFGA) based association rule mining. This proposed algorithm generates intersecting rules from dataset in two steps. The fitness function is defined for first step and also a different fitness function for second step. The first step is aimed to generate rules promptly and then second step is used to refine the rules generated in first step. Fitness function in second step includes some other interestingness measures than support and confidence to generate relevant rules.

Many algorithms are developed in recent years for mining association rules like Apriori, FP-Tree, Partition algorithm and etc. The subsequent papers [1][8][9] contributed by using multiple objectives like support, confidence, simplicity and etc.

Ashish Ghosh, Bhabesh Nath(2004) discussed: Multi-objective rule mining using genetic algorithms[1], used measures like support count, comprehensibility and interestingness for evaluating interesting rules as their different objectives for mining association rule problem.

Manish Saggar, Ashish Kumar Agrawal, Abhimanyu Lad (2004) discussed: Optimization of Association Rule Mining using Improved Genetic Algorithms [2], main objective is to use genetic algorithm in the discovery of high level prediction rules that perform a global search and perform better with attributes than greedy rule induction algorithms (used in data mining).

Bilal Alata¸ S. Erhan Akin (2005) discussed: An efficient genetic algorithm for automated mining of both positive and negative quantitative association rules [3], proposed a genetic algorithm as a search strategy for not only positive association rule but also for negative association rules. Proposed algorithm also not relies on minimum support and minimum confidence.

Virendra Kumar Shrivastava, Dr. Parveen Kumar, Dr. K. R. Pardasani(2010) discussed: Extraction of Interesting Association Rules using GA Optimization[4]. In this paper, they used genetic algorithm for extracting association rules. They used measures like support, confidence, interestingness, and completeness.

Amy H.L. Lim a, Chien-Sing Lee a, Murali Raman (2012) discussed: Hybrid genetic algorithm and association rules for mining workflow best practices [5], correlation measure instead of traditional support and confidence in genetic algorithm to driven data dynamically. Correlation fitness function is used to support upward closure in association rule.

Jesmin Nahar et al (2013) discussed: Association rule mining to detect factors which contribute to heart disease in males and females [6], investigates the sick and healthy factors related to heart diseases in males and females by using UCI Cleveland dataset, a biological database. They compared three rule generation algorithms Apriori, Predictive Apriori, Tertius.

Dong Gyu Lee et al (2013) discussed: Discovering Medical Knowledge using Association Rule Mining in Young Adults with Acute Myocardial Infarction[7 proposed association rule mining algorithm that can

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 9, September 2013

generate association rules related to hypertension and diabetes from AMI patients having age 45 years old or lesser. They also use measures like lift, leverage and etc

B. Minaei-Bidgoli, R. Barmaki, M. Nasiri(2013) discussed: Mining numerical association rules via multi-objective genetic algorithms[8], proposed a multi-objective genetic algorithm and used measures like confidence, interestingness, and comprehensibility as multiple objectives for genetic algorithm method.

Basheer Mohamad, Al-Maqaleh(2013) discussed: Discovering Interesting Association Rules: A Multi-objective Genetic Algorithm Approach[9], proposed a multi-objective genetic algorithm for generating association rules by using measures like support, confidence and simplicity/ comprehensibility.

Bettahally, N. Keshavamurthy, Asad M. Khan, Durga Toshniwal(2013) discussed: Privacy preserving association rule mining over distributed databases using genetic algorithm[10], compared traditional frequent pattern mining algorithm i.e. Apriori algorithm with proposed genetic algorithm in local search.

II. PROPOSED WORK

Proposed algorithm is Multi-fitness function Genetic algorithm for association rule mining. This proposed algorithm used two fitness functions for refining the rules. The proposed algorithm is two step algorithms. The first step is aimed to generate rules promptly and then second step is used to refine the rules generated in first step. To refine the rules different interestingness measures are used as fitness function parameters.

MULTI-FITNESS FUNCTION GENETIC ALGORITHM

Begin:
Step I; Initialize the population;
    For each individual; calculate fitness using support as parameter;
    For i=1 to maximum generations
        [Selection] Select the individuals using Roulette Wheel Selection;
        [Crossover] Crossover the parents to form new offspring's;
        [Mutation] Mutate new offspring at each locus;
    Place new offspring in the new generation;
    If fitness of new population> min_supp
    Select individuals for next generation;
    End If
Next i;
Step II; Initialize the population;
    For each individual; Calculate fitness using Confidence, Lift, Leverage, Coverage as parameter;
    For i=1 to maximum generations
        [Selection] Select the individuals using Roulette Wheel Selection;
        [Crossover] Crossover the parents to form new offspring's;
        [Mutation] Mutate new offspring at each locus;
    Place new offspring in the new generation;
    If fitness of new population> min_conf
    Select individuals for next generation;
    End If
Next i;
End

Operators in proposed algorithm:

I. **Selection Operator**: Roulette Wheel selection process is used for selecting the individual's w.r.t fitness function.

II. **Crossover operator:** Single-point crossover is used in proposed algorithm. For example,
    100000110101
    101010101011
Choose a random bit for crossover and resultant offspring is shown as:
    100000101011
    101010110101

III. **Mutation Operator**: This gives a chance that a gene within a chromosome will be flipped means '0' become '1 ' and '1' become '0'. For example:
    100000110101
Choose a random bit and resultant offspring is:
    100001110101

Fitness Function:

Confidence(A->B)= support(A->B)/support(A)
Lift(A->B)=confidence(A->B)/support(B)= support(A->B)/(support(A)* support(B))
Coverage(A->B) = support(A)
Leverage (A->B) = support(A->B) – support(A)*support(B) = confidence(A->B)- support(B)

Fitness Function = (w1*confidence(A->B) + w2 * coverage(A->B)) / (w3*lift(A->B) + w4*leverage(A->B))

where w1 is confidence weight,
w2 is coverage weight,
w3 is lift weight,
w4 is leverage weight.

For proposed algorithm w1=w3=w4=1 and w2=2.

III. RESULTS

To apply proposed algorithm on database to generate association rules:
As discussed earlier, the proposed algorithm implements a two stage rule mining. In the first stage simple rule mining is done with fitness function includes support and second

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 9, September 2013

stage refines the rules from step1 with fitness function with interestingness measures.

The experiments were performed on machine Intel® Core 2 Duo CPU 2.00GHz with 32bit operating system and software was MatLAB.

To analyse the performance of proposed algorithm with Apriori algorithm and existing Genetic Algorithm association rule mining below is the comparison between Apriori algorithm, Genetic algorithm, and Multi-Fitness Function Genetic algorithm based on support count:

Table 1 represents number of rules generated by various support count and Figure 1 represents the graph of comparison between by Apriori algorithm, Genetic algorithm, and Multi-Fitness Function Genetic algorithm.

Table1: Comparison between Apriori algorithm, Genetic algorithm, and Multi-Fitness Function Genetic Algorithm by number of rule generation with varies support count

Support Count	**Apriori Algorithm**	**Genetic Algorithm**	**Multi-Fitness FunctionGenetic Algorithm**
0.1	94698	122	48
0.2	94689	121	44
0.3	82602	127	41
0.4	82540	124	42
0.5	66368	129	46

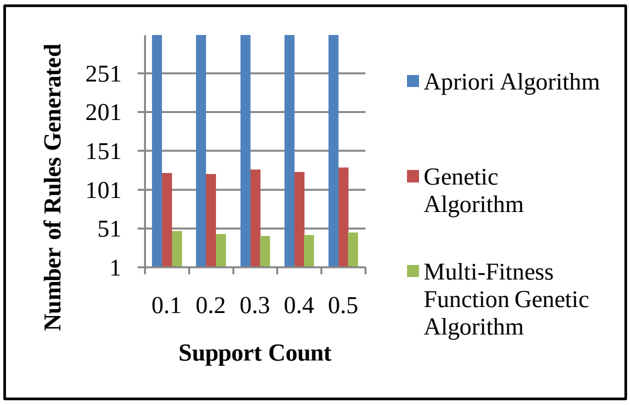

Figure 1: Comparison by various support count

RESULT ANALYSIS ON BASIS ON DIFFERENT MAXIMUM GENERATION:

Existing genetic algorithm and Multi- Fitness Function Genetic algorithm are evaluated on different generation value varies from 20 to 150 with crossover probability ranges from 0.6 – 0.9, as in the proposed algorithm single point crossover is used so crossover probability ranges from 0.6 – 0.9. Below are the graphs and the table representing number of rules generated with different generation size.

Table2: Existing Genetic Algorithm analysis based on number of Rules generated with different Crossover Probability, Support Count, and Different Maximum Generation

Maximum Generation	**Crossover Probability**	**Support Count**				
		0.001	**0.002**	**0.003**	**0.004**	**0.005**
20	0.6	140	129	104	112	93
50	0.6	138	129	114	122	83
100	0.6	127	131	106	106	86
150	0.6	133	124	108	111	80
20	0.7	124	134	105	108	83
50	0.7	122	120	121	109	74
100	0.7	123	141	113	104	82
150	0.7	131	138	104	104	82
20	0.8	123	129	108	106	78
50	0.8	123	136	110	107	89
100	0.8	141	119	117	117	85
150	0.8	132	135	109	109	86
20	0.9	130	135	99	116	77
50	0.9	132	132	124	103	83
100	0.9	124	124	106	123	76
150	0.9	123	123	110	101	78

Below graphs shows rules generated with crossover probability ranges from 0.6 – 0.9 with different generations varies from 20 – 150 and the results are of existing genetic algorithm.

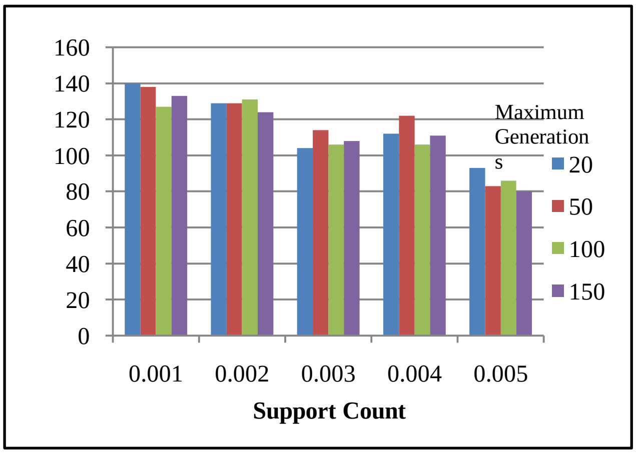

Figure 2: Number of rules generated with Crossover probability 0.6

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 9, September 2013

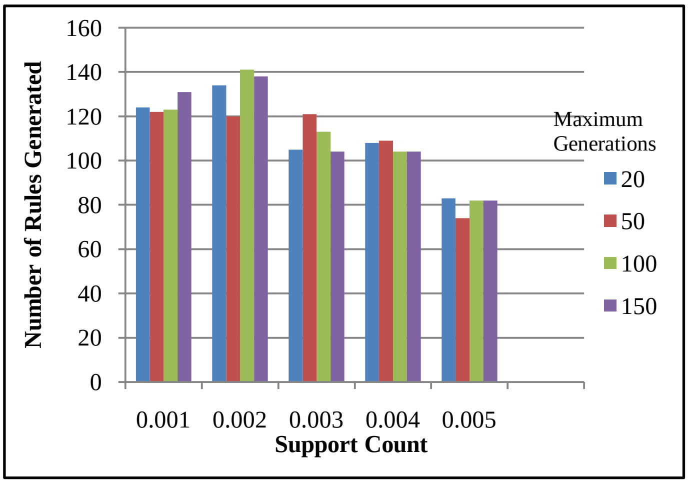

Figure 3 : Number of rules generated with Crossover probability 0.7

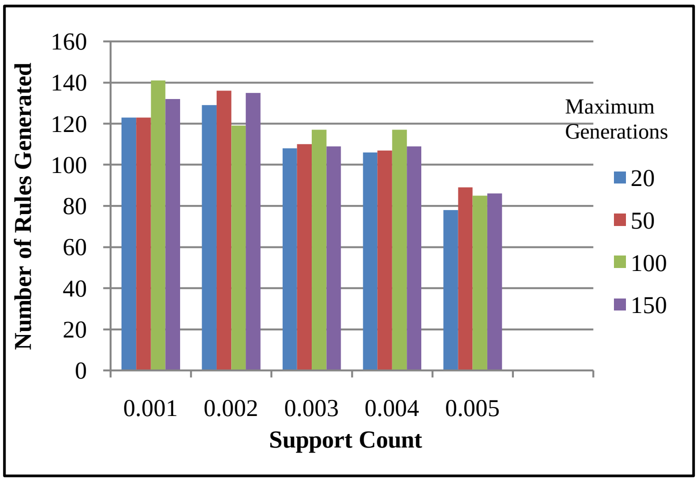

Figure 4: Number of rules generated with Crossover probability 0.8

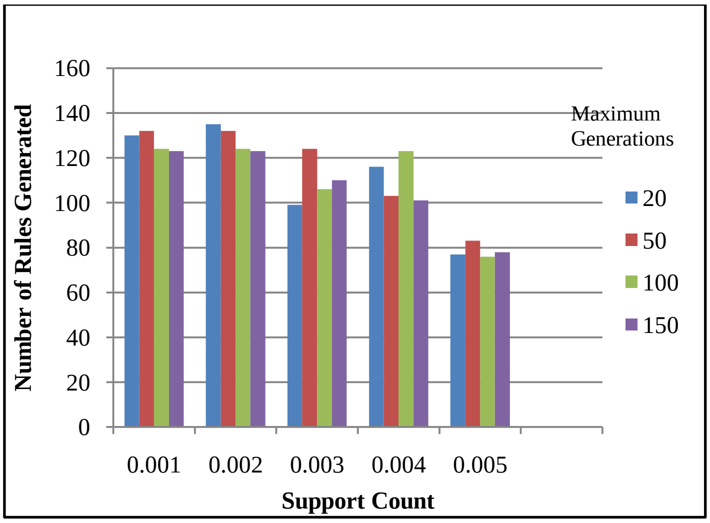

Figure 5: Number of rules generated with Crossover probability 0.9

Below graphs and table shows the results of proposed algorithm i.e. Multi- Fitness Function genetic algorithm.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 9, September 2013

Table3: Multi-Fitness Function Genetic Algorithm analysis based on number of Rules generated with different Crossover Probability, Support Count, and Different Maximum Generation

Maximum Generation	**Crossover Probability**	**Support Count**				
		0.001	**0.002**	**0.003**	**0.004**	**0.005**
20	0.6	43	43	31	34	28
50	0.6	42	46	35	36	31
100	0.6	47	49	33	35	28
150	0.6	40	47	42	40	27
20	0.7	45	45	41	40	31
50	0.7	44	42	36	40	31
100	0.7	47	44	43	37	24
150	0.7	41	43	33	41	30
20	0.8	40	45	43	37	28
50	0.8	38	46	34	32	31
100	0.8	39	46	36	35	32
150	0.8	47	55	39	39	33
20	0.9	42	47	37	33	28
50	0.9	40	46	35	36	28
100	0.9	39	44	39	38	18
150	0.9	39	46	33	37	22

Below graphs shows rules generated with crossover probability ranges from 0.6 – 0.9 with different generations varies from 20 – 150 and the results are of existing genetic algorithm.

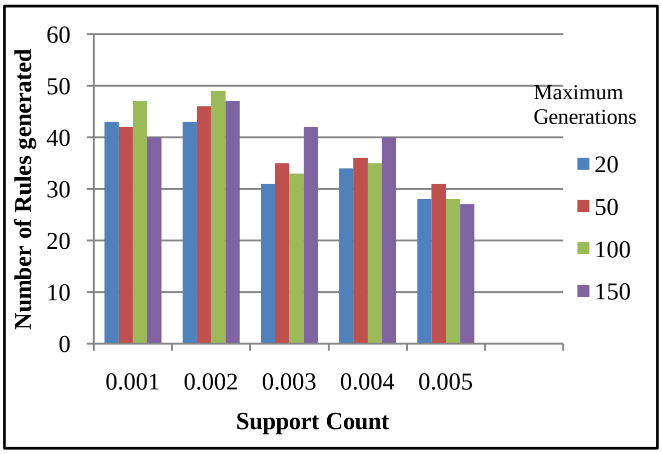

Figure 6: Number of rules generated with Crossover probability 0.6

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 9, September 2013

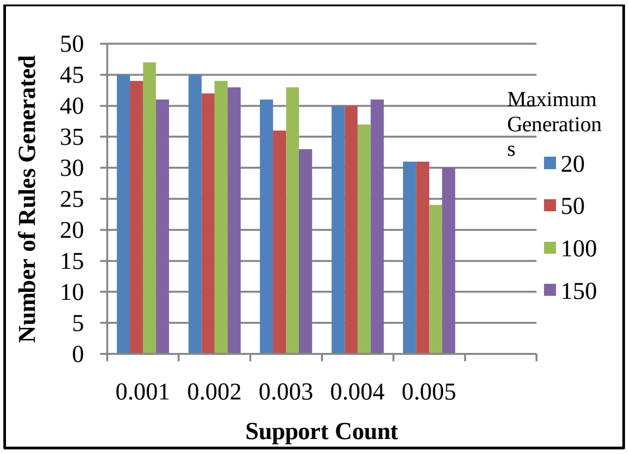

Figure 7: Number of rules generated with Crossover probability 0.7

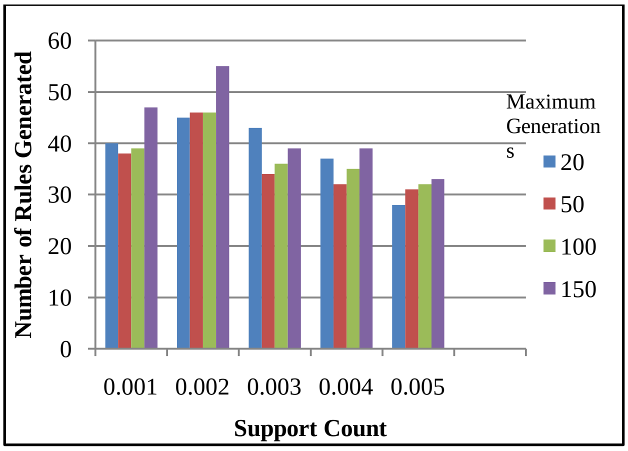

Figure 8: Number of rules generated with Crossover probability 0.8

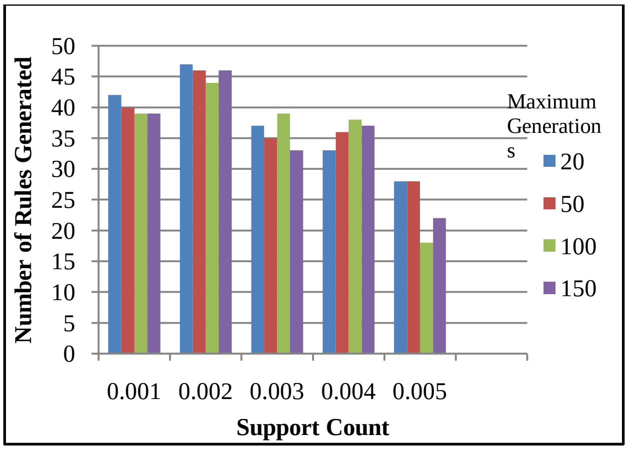

Figure 9: Number of rules generated with Crossover probability 0.9

IV. CONCLUSION

Simulation results shows that proposed genetic algorithm generates more efficient and effective rules than other association rule mining techniques such as Apriori and existing genetic algorithm. In the proposed algorithm termination condition involved was maximum generations. Analysis was also taken on the basis of different maximum generation and was also based on different crossover probabilities. As in the proposed algorithm single point crossover was used so the crossover probability varies from 0.6 to 0.9 so analysis was also based on this range. Proposed algorithm can be used for further more applications of data mining. In proposed algorithm different termination conditions can be also be used.

ACKNOWLEDGEMENT

The authors are greatly indebted to the Department of Computer Science and Engineering, D.A.V Institute of Engineering and Technology, Jalandhar for providing excellent lab facilities that make this work possible.

REFERENCES

[1] Ashish Ghosh, Bhabesh Nath," Multi-objective rule mining using genetic algorithms", *Information Sciences, Elsevier*, Vol.163, pp.123–133, 2004.

[2] Manish Saggar, Ashish Kumar Agrawal, Abhimanyu Lad, " Optimization of Association Rule Mining using Improved Genetic Algorithms", *IEEE International Conference on Systems, Man and Cybernetics, pp.3725-3729* , 2004.

[3] Bilal Alata¸ S· Erhan Akin," An efficient genetic algorithm for automated mining of both positive and negative quantitative association rules", *Soft Computing, Springer-Verlag*, Vol. 10, pp. 230–237, 2005.

[4] Virendra Kumar Shrivastava1 Dr. Parveen Kumar, Dr. K. R. Pardasani, *"*Extraction of Interesting Association Rules using GA Optimization*", Global Journal of Computer Science and Technology*, Vol. 10, Issue 5, Ver. 1.0, pp.81-84, July 2010.

[5] Amy H.L. Lim a, Chien-Sing Lee a, Murali Raman," Hybrid genetic algorithm and association rules for mining workflow best practices", *Expert Systems with Applications, Elsevier* , Vol. 39, pp. 10544–10551, 2012.

[6] Jesmin Nahar a, Tasadduq Imama, Kevin S. Tickle a, Yi-Ping Phoebe Chen," Association rule mining to detect factors which contribute to heart disease in males and females ",*Expert Systems with Applications, Elsevier* , Vol. 40,pp.1086–1093,2013.

[7] Dong Gyu Lee, Kwang Sun Ryu, Mohamed Bashir, Jang-WhanBae, Keun Ho Ryu, " Discovering Medical Knowledge using Association Rule Mining in Young Adults with Acute Myocardial Infarction", Journal of Medical Systems, *Springer*, Vol.37,Issue 2, 2013.

[8] B. Minaei-Bidgoli, R. Barmaki, M. Nasiri, " Mining numerical association rules via multi-objective genetic algorithms", *Information Sciences, Elsevier*, 2013.

[9] Basheer Mohamad, Al-Maqaleh, " Discovering Interesting Association Rules: A Multi-objective Genetic Algorithm Approach", *International Journal of Applied Information Systems*, Vol. 5, No.3, pp.47-52,February 2013.

[10] Bettahally, N. Keshavamurthy, Asad M. Khan ,Durga Toshniwal, "Privacy preserving association rule mining over distributed databases using genetic algorithm", *Neural Computing and Applications, Springer-Verlag*, 2013.

[11] J. Alcalá-Fdez1, A. Fernández1, J. Luengo1, J. Derrac1,S. García2, L. Sánchez3 And F. Herrera1," KEEL Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework",*J. of Mult.-Valued Logic & Soft Computing*, Vol. 17, pp. 255–287, 2011.

[12] By Jyothi Pillai, O.P.Vyas, "Encapsulation of Soft Computing Approaches within Itemset Mining – A Survey", *Global Journal of Computer Science and Technology Software & Data Engineering*, Vol. 12 Issue 15 Versions 1.0, 2012.

[13] Chengfeng Jian, Fang Li, "An Improved Virus Evolutionary Genetic Algorithm For Workflow Mining ", *Journal of Theoretical and Applied Information Technology*, Vol. 47 No.1, pp.406-411, 2013.

[14] Sanjeev Sharma, Vivek Bahde, Sudhir Sharma, "Optimization of Association Rules Using Genetic Algorithm", *International Journal of Soft Computing*, pp.75-79, 2007.

[15] Chad Creighton, Samir Hanash, "Mining gene expression databases for association rules", *Bioinformatics*, Vol. 19 no. 1, pp.79-86, 2003.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 9, September 2013

[16] Nicolò Flugy, Jesús Alcalá-Fdez, Andrea Bonarini, Francisco Herrera, " Evolutionary Extraction of Association Rules: A Preliminary Study on their Effectiveness", *Springer-Verlag*, pp. 646–653, 2009.
[17] Ming-Hsiung Ying, Shao-Hsuan Huang, Luen-Ruei Wu, "An Item Selection Strategy Based on Association Rules and Genetic Algorithms", *JOURNAL OF SOFTWARE*, Vol. 5, No. 12, pp. 1378-1383, 2010.
[18] Rahul Malhotra, Narinder Singh, Yaduvir Singh, "Genetic Algorithms: Concepts, Design for Optimization of Process Controllers", *Computer and Information Science*, Vol. 4, No. 2, pp.39-54, March 2011.
[19] Wenxiang Dou, Jinglu Hu, Kotaro Hirasawa and Gengfeng Wu, " Quick Response Data Mining Model Using Genetic Algorithm ", *SICE Annual Conference*, pp. 1214-1219, 2008.
[20] Rupali Haldulakar, Prof. Jitendra Agrawal, "Optimization of Association Rule Mining through Genetic Algorithm", *International Journal on Computer Science and Engineering*, Vol. 3 No. 3, pp.1252-1259, March 2011.
[21] Soumadip Ghosh, Sushanta Biswas, Debasree Sarkar, Partha Pratim Sarkar, "Mining Frequent Itemsets Using Genetic Algorithm", *International Journal of Artificial Intelligence & Applications*, Vol.1, No.4, pp.133-143 , October 2010.
[22] Anandhavalli M., Suraj Kumar Sudhanshu, Ayush Kumar, Ghose M.K., "Optimized association rule mining using genetic algorithm", *Advances in Information Mining*, Vol. 1, Issue 2 , pp. 01-04, 2009.
[23] Farah Hanna AL-Zawaidah, Yosef Hasan Jbara, Marwan AL-Abed Abu-Zanona, " An Improved Algorithm for Mining Association Rules in Large Databases", *World of Computer Science and Information Technology Journal*, Vol. 1, No. 7, pp.311-316, 2011.
[24] Pradeep Singh Raghav, Nikhil Singh, Satendra Singh Thakur, Priyank Jain, "Optimization Of Multiple Relational Classification Algorithms Using Genetic Algorithm", *International Journal of Computer Science and Communication*, Vol. 3, No. 1, pp. 215-217, January-June 2012.
[25] D. Kerana Hanirex, Dr. A. Kumaravel, " An Efficient Partition and Two Dimensional Approach For Mining Frequent Itemsets", *International Journal of Technological Synthesis and Analysis*, Vol. 1 Issue 1, pp. 14-17, December 2012.
[26] 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, pp. 1-4, January 2013.
[27] Huan Wu, Zhigang Lu, Lin Pan, Rongsheng Xu, "An Improved Apriori-based Algorithm for Association Rules Mining", *IEEE Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, pp.51-55 ,2009.
[28] D. Magdalene Delighta Angeline, I. Samuel Peter James, "Association Rule Generation Using Apriori Mend Algorithm for Student's Placement", *International Journal of Emerging Sciences*, Vol. 2(1), pp. 78-86, March 2012.
[29] R. Uday kiran, P Krishna Reddy, "An Improved Multiple Minimum Support Based Approach to Mine Rare Association Rules", *IEEE Symposium on Computational Intelligence and Data Mining*, 2009.