ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

Frequent Item Set Generation Using Improved PSO over Generalized Transactional Dataset

Sweta Mishra[1], Yamini Chouhan[2]

M.Tech Scholar (Computer science and Engineering), Shri Shankracharya Technical Campus, Bhilai(C.G), India[1]

Assistant professor, Dept. of C.S.E, Shri Shankracharya Technical Campus, Bhilai(C.G), India[2]

ABSTRACT: The Frequent Item set generation using improved PSO over generalized dataset can be used in analyzing customer's buying habits, so that we can predict the customer's need and help the sellers to sell the items and build a better relationship between them. Here we will use PSO algorithm for basket data analysis. Data mining, as a discipline, is a group of techniques ranging from statistics, computer science, operation research and artificial intelligence, for efficient and automated discovery of previously unknown, valid, novel, actionable and understandable knowledge in large databases. Association rule mining is the data mining task employed to solve an important problem in marketing parlance viz., market basket analysis. This process analyses customer's buying habits by finding associations between the different items that customers place in their shopping baskets.

KEYWORDS: Association Rule Mining, PSO, Data analysis, Market basket analysis

I.INTRODUCTION

Data Mining

The computing process of discovering patterns in large data sets is termed as Data Mining. It is the process of extracting information from a data set and transforms it into an understandable structure for further use. It is a powerful technology with great potential to help companies/organization focus on the most important information in their data warehouses. To predict future trends and behaviours, allowing business to make proactive, Knowledge driven decisions, Data mining tools are used. The data warehouse supports Online Analytical Processing (OLAP), the function and performance requirement of which are quite different from those of OLTP applications traditionally supported by the operational database [Reddy, G et al 2010].The most important application of data mining is the association rule mining. Association Rule Mining was introduced by R.Agrawal and R.Srikant in 1993.

Association Rule Mining

A procedure which is meant to find frequent patterns, correlations, associations or casual structures from data sets found in various kinds of databases such as relational databases, transactional databases and other forms of data repositories is known as Association Rule Mining. Association is a data mining function that discovers the probability of the co-occurrences of items in a collection. The relationships between co-occurring items are expressed as association rules. By given a super specialized threshold, also known as minimum support, the mining of association rules can discover the complete set of frequent

Particle Swarm Optimization

James Kennedy and Russel C.Eberhart in 1995 proposed a heuristic global optimization method. It is one of the most commonly used computational methods known as "Particle Swarm Optimization" technique. It might sound complicated but it's a very simple algorithm. Over a number of iterations, group of variables have their values adjusted closer to the member whose value is closest to the target at any given point.

This algorithm keeps track of the three global variables:

1. Target value or condition.
2. Global best value (gBest) indicating which particle's data is currently closest to the target.
3. Stopping values indicating when the algorithm should stop if the target isn't found.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

Each particle consist of:

- A possible solution represented by a data.
- A velocity value indicating how much a data can be changed.
- A personal best (pBest) value indicating the closest the particle's data has ever come to the target.

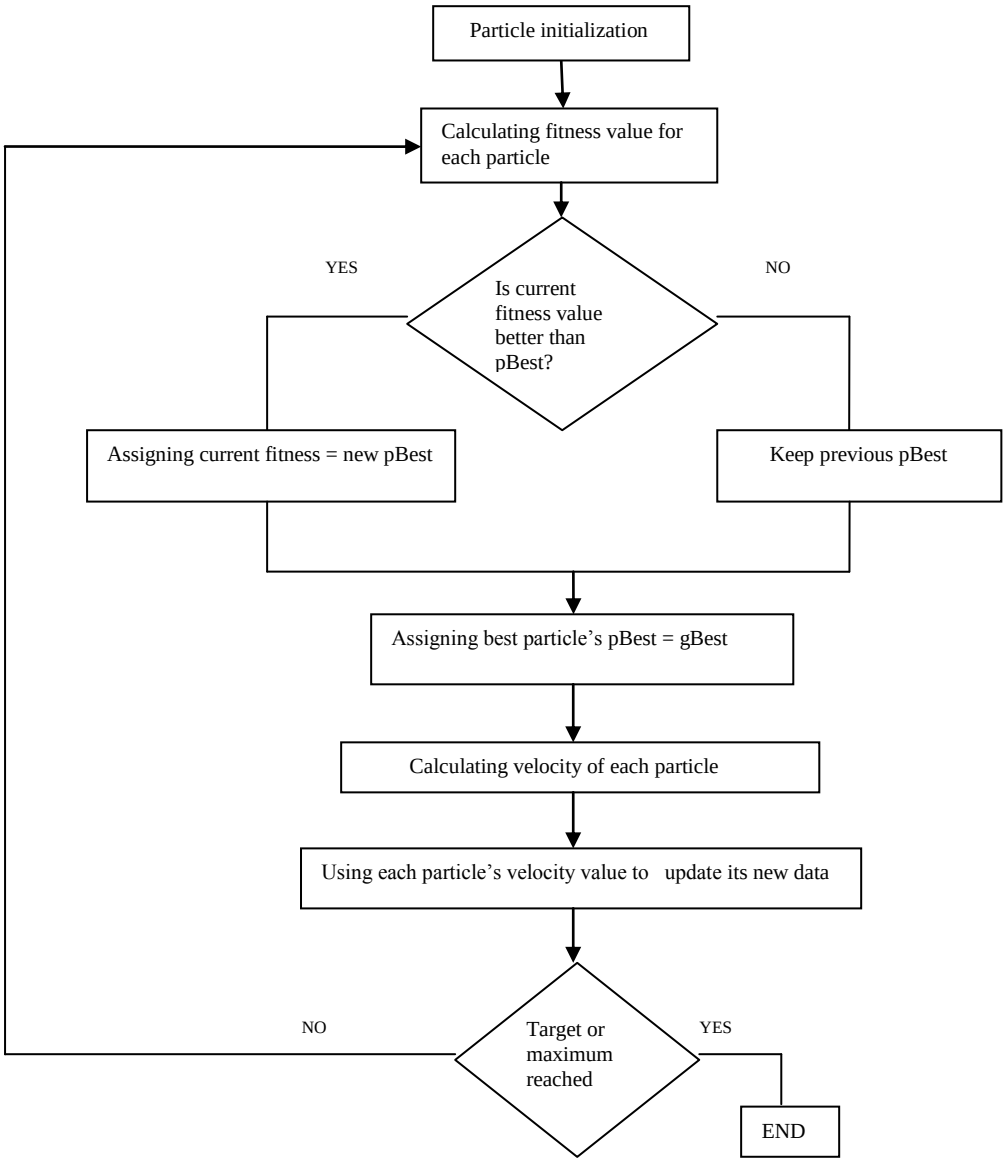

FIG 1: FLOW CHART SHOWING PSO ALGORITHM

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

III. **LITERATURE SURVEY**

S.No.	Author/Title/Publication/Year	Method Used	Description
1.	K.N.V.D. Sarath/ Association rule mining using binary particle swarm optimization / Elsevier 2013 [1]	PSO	Author developed a binary particle swarm optimization (BPSO) based association rule miner. Proposed BPSO based association rule miner generates the association rules from the transactional database by formulating a combinatorial global optimization problem, without specifying the minimum support and minimum confidence unlike the a priori algorithm.
2.	Anshuman Singh Sadh et. al./Apriori and Ant Colony Optimization of Association Rules/IJACR 2103 [8]	ACO	In this paper author applied apriori and ant colony optimization technique to achieve the positive and negative association rule optimization. By proposed approach author achieve better optimization.
3.	Manisha Gupta/Application of Weighted Particle Swarm Optimization in Association Rule Mining/ IJCSI 2014 [7]	Weighted PSO	This study has demonstrated that using the Weighted PSO algorithm can determine these two parameters quickly and objectively, thus enhancing mining performance for large databases by applying the FoodMart2000 database.
4.	Abdoljabbar Asadi et. al./New Binary PSO based Method for finding best thresholds in association rule mining/ Life Science Journal 2012	Binary PSO	Main goal of this paper is to presenting an optimal method to find suitable values of minimum threshold for support and confidence by means of Binary Particle Swarm Optimization. Data used for the paper is a 4000 random records sample from Foodmart 2000 Database. Implementation of the proposed method has been done using R2010b version of MATLAB software. Proposed algorithm improves the performance of association rule mining by automatically setting suitable values for minimum support and confidence thresholds.
5.	S.Deepa et. al./An Optimization of Association Rule Mining Algorithm using Weighted Quantum behaved	Weighted Quantum behaved PSO	Author propose Weighed Quantum behaved Particle Swarm Optimization (WQPSO) algorithm

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

	PSO/IJPCSC 2012		for improving the performance of association rule mining algorithm Apriori. It is a global convergence guaranteed algorithm, which outperforms original PSO algorithm and it has fewer parameters to control the search ability of PSO. Finding minimum support and minimum confidence values for mining association rules seriously affect the quality of association rule mining. In association rule mining, the minimum threshold values are always given by the user. But in this paper, WQPSO algorithm is used to determine suitable threshold values automatically and also it improves the computational efficiency of Apriori algorithm. First, the WQPSO algorithm is processed to find the minimum threshold values. In this algorithm which particle having the highest optimal fitness value, its support and confidence values are taken as the minimum threshold value to association rule algorithm.
6.	Ali Hadian et. al./Clustering Based Multi-Objective Rule Mining using Genetic Algorithm/IJDCTA 2010	Genetic Algorithm	In this article, authors have optimized the Multi objective Genetic approach to discover association rules from database. The proposed method is to cluster the data and ignore some dispensable comparisons with the database. Due to the fact that existence of a rule in some of the clusters is impossible, ignoring these clusters while counting support of rule's related item set helps the algorithm to avoid some redundant comparisons.

IV. METHODOLOGY

In this project we have proposed an algorithm which will overcome the bottle necks of existing algorithm which we have discussed in problem identification section. In proposed algorithm we need not to transform the indiscriminate transactional dataset into binary dataset, proposed algorithm can be directly applied on indiscriminate dataset. In earlier algorithms frequent item-sets are identified from all-possible item-sets called candidate sets by using a parameter called support count and a user-defined argument known as minimum support. Support count of an item set is defined by the number of records in the dataset that comprise all the items of that set. If there is value of minimum support is too high, then the number of frequent item sets spawned will goes less, and thus ensuing in generation of a limited rules. Over

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

again, if the value is too lesser, then nearly all possible item sets will converted frequent and consequently an enormous number of rules may be spawned. Picking better rules from them might be alternative problem. After detecting the frequent item-sets another user-defined parameter known as minimum confidence generates the rules.

Our algorithm

1. TransposedData Set as Input
2. Read the database to count the support of C1 to determine L1 using sum of rows.
3. L1= Frequent 1- itemsets and k: = 2
4. While (k-1 ≠ NULL set) do
   Begin
   Ck: = Call Gen_candidate_itemsets (Lk-1)
   Call Prune (Ck)
   for all itemsets i ∈ I do
   Calculate the support values using dot-multiplication of array;\
   Generate Random value of Support Rs
   if Rs is Optimal Check Using PSO
   Set Rs as minimum support
   End if
   Lk : = All candidates in Ck with a minimum support Rs;
   k:=k+1

   End
5. End of step-4

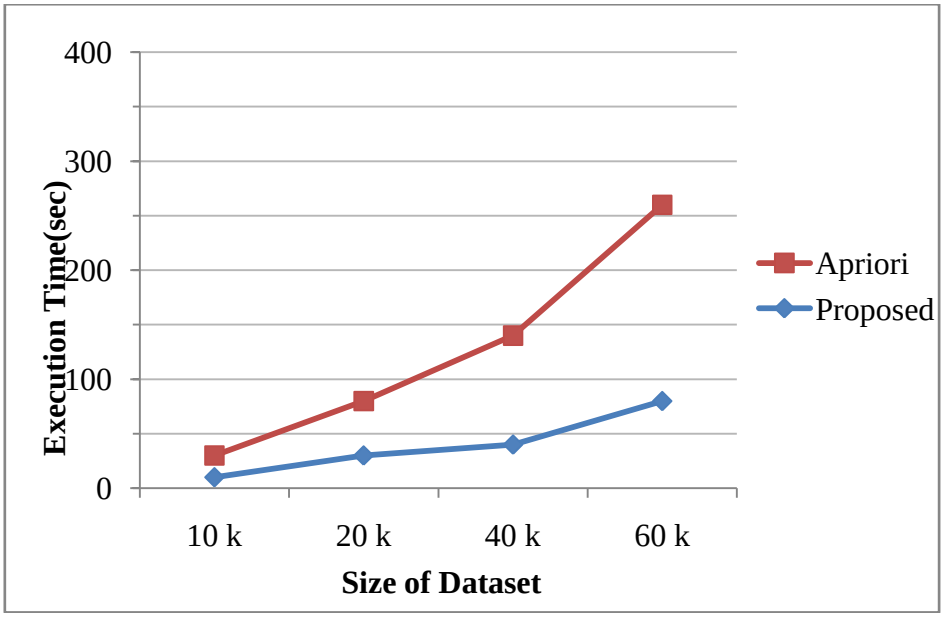

FIG 2 : **Performance Comparison of Apriori and our methodology**

Above Figure shows the performance comparisons between Apriori and proposed algorithm. It compares the execution time taken by the apriori algorithm and proposed algorithm for different datasets.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

V. CONCLUSION

Discovering association rules is an important class of data mining, and association rules have a wide area of usage. Although many efficient algorithm have been proposed up to now. Extracting association rules is still a computationally expensive operation in large data base.

Some bottle neck of Apriori algorithm overcame by earlier algorithms which are based on PSO (Particle Swarm Optimization), ACO (Ant Colony Optimization), Genetic algorithm etc. by getting optimal value of minimum support.

Most of the data mining research and study is experimented on market basket data i.e. available in generalized format, but Apriori and earlier algorithms takes input as Boolean dataset thus it requires conversion of generalized dataset to Boolean dataset, which takes much time.

Proposed algorithm works upon generalized dataset hence pre-processing time reduces and due to transposed data as input reduces time complexity for calculation of support count value and PSO provides optimal value of minimum support which helps to generate interesting frequent item sets.

REFERENCES

1. K.N.V.D. Sarath , Vadla-mani Ravi, "Association rule mining using binary particle swarm optimization", Engineering Applications of Artificial Intelligence, Elsevier, 1832–1840 , 26(2013).
2. Waiswa, P. P. W. , Baryamureeba, V. , "Extraction of interesting association rules using genetic algorithms", Int. J. Comput. ICTRes. 2(1), pp 26–33, 2008.
3. Nandhini,M.,Janani,M.,Sivanandham,S.N., "Associaiton rule mining using swarm intelligence and domain ontology", in :IEEE International Conference on Recent Trends in Information Technology, (ICRTIT),Coimbatore, pp.537–541, 2012..
4. JIAWEI HAN, "Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach", Data Mining and Knowledge Discovery, Kluwer Academic Publisher, pp 53–87,2004.
5. Shafiq Alam, Gillian Dobbie, Yun Sing Koh, Patricia Riddle, Saeed Ur Rehman, "Research on particle swarm optimization based clustering: A systematic review of literature and techniques, Swarm and Evolutionary Computation", Elsevier, Volume 17, Pages 1-13, August 2014, ISSN 2210-6502
6. B. Minaei-Bidgoli, R. Barmaki, M. Nasiri, Mining numerical association rules via multi-objective genetic algorithms, Information Sciences, Volume 233, Pages 15-24, 1 June 2013, ISSN 0020-0255.
7. Manisha Gupta, "Application of Weighted Particle Swarm Optimization in Association Rule Mining", International Journal of Computer Science and Informatics (IJCSI) Volume-1, Issue-3 ISSN (PRINT): 2231 –5292,
8. Anshuman Singh Sadh,Nitin Shukla,"Apriori and Ant Colony Optimization of Association Rules",International Journal of Advanced Computer Research , Volume-3 Number-2 Issue-10, pp - 35 (ISSN (print): 2249-7277 ISSN (online): 2277-7970), June-2013
9. Ms. Kumudbala Saxena, Dr. C.S. Satsangi, "A Non-Candidate Subset-Superset Dynamic Minimum Support approach for Sequential pattern Mining", International Journal of Advance Computer Research(IJACR), Volume-2, Number-4, Issue-6, December-2012.
10. Reddy, G., Srinivasu, R., Rao, M., and Reddy, S., "Datawarehousing, Datamining, OLAP and OLTP Technologies are essential elements to Support decision making in industries", International Journal Computer Science and Engineering, 9(2), ISSN: 0975-3397, pp. 2865-2871, 2010.
11. Pallavi Dubey ,"Association Rule Mining on Distributed Data", International Journal of Scientific & Engineering Research, Vol. 3, Issue 1, ISSN : 2229-5518, 2012.
12. Lee, W., Stolfo, S.J. and Mok, K.W., "A data mining framework for building intrusion detection models", IEEE Symposium on Security and Privacy, 1999.
13. Brijs, T., Swinnen, G., Vanhoof, K. and Wets.G, "Using Association Rules for Product Assortment Decisions: A Case Study", SIGKDD international conference on Knowledge discovery and data mining, pp – 254-260 , August 15 – 18, 1999, ISBN:1-58113-143-7 .
14. Brijs, T., Goethals, B., Swinnen, G., Vanhoof, K. and Wets. G, "A Data Mining Framework for Optimal Product Selection in Retail Supermarket Data: The Generalized PROFSET Model", SIGKDD international conference on Knowledge discovery and data mining, August 20 - 23, 2000.
15. Wang, K. and Su, M.Y., "Item Selection by Hub-Authority Profit Ranking", SIGKDD international conference on Knowledge discovery and data mining 2002 [SIGKDD], pp – 652-657, July 23-26, 2002, ISBN:1-58113-567-X.