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 1, January 2017

A Review on PSO and Association Rule Mining for Item Set Generation

Sweta Mishra[1], Yamini Chouhan[2]

M.Tech Scholar, Dept. of C.S.E, Shri Shankracharya Group of Institutions, Bhilai(C.G), India[1]

Assistant Professor, Dept. of C.S.E, Shri Shankracharya Group of Institutions , Bhilai(C.G), India[2]

ABSTRACT: 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. Here we will use PSO algorithm for basket data analysis. We will try to analyse the buying habits of the customers, so that we can predict the customers need and help the seller to sell the items and to build a better relationship between them.

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

I. INTRODUCTION

The Process of analysing data from different perspectives and summarizing it into useful information is known as Data Mining. Data Mining is a type of analytical tool for analysing data. It can be used to analyse data from different angles or dimensions. Technically, it is the process of finding patterns or correlations among various fields in large relation database. Data mining helps to discover useful knowledge from large amount of data, this data can be stored in databases, data warehouses. 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].

A rule-based machine learning method for discovering interesting relations between different variables in large databases is Association Rule Mining. It is intended to identify strong rules discovered using some measures of interestingness. 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 patterns. That is, once the minimum support is given, the complete set of frequent patterns is determined [9]. In order to retrieve more correlations among items, users may specify a relatively lower minimum support [9].

One of the applications of association rule mining is the market-basket analysis. It is valuable for discovering business trends, sales promotion and direct marketing. It can be effectively used for store. It can also be used effectively for store layout, catalogue design and cross-sell. Association rule mining has important applications in other domains as well such as e-commerce applications, Web page personalization.

Association is transaction based, unlike data mining functions. In a transaction processing system, transaction such as market basket or web session is consisted in a case. The collection of items is an attribute of the transaction. Other attributes can be the date, time, location, user ID associated with the transaction. This attributes are referred as multi-record attributes. Transactional data is said to be in multi-record case format. Each case is associated with collection of items.

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 1, January 2017

case ID	attribute
TRANS_ID	ITEM_ID
11	B
11	D
11	E
12	A
12	B
12	C
12	E
13	B
13	C
13	D
13	E

Fig. 1: Transactional Data

A Heuristic global optimization method, originally proposed by James Kennedy and Russel C. Eberhart in 1995 is the Particle Swarm Optimization (PSO). It is one of the most commonly used computational method. It optimizes a problem by iteratively improving a candidate solution with regard to a given measure of quality. In order to solve a problem, PSO holds a population of candidate solution; say dubbed particles and moving these particles along the search space using some simple mathematical formulae over the particle's position and velocity. Each particle's movement is influenced by its local best known position and is also guided towards the best position in the search space. This are updated as better positions are found by other particles, thus moves the swarm towards the best solution.

II. LITERATURE SURVEY

K.N.V.D. Sarath, Vadla-mani Ravi [1] developed a binary particle swarm optimization (BPSO) based association rule miner method by formulating it as a combinatorial global optimization problem. The algorithm generates the best M rules from the given database, where M is a given number. The effectiveness of the algorithm as well as well known apriori algorithm and the FP-growth algorithm is tested on a real life bank dataset taken from commercial bank in India and three transactional datasets.

Waiswa P.P.W., Baryamureeba V. [2] extracted the association rules using Pareto-based multi-objective evolutionary algorithm rule mining method based on Genetic algorithm.

Nandhini M.,Janani M.,Sivanandham S.N [3] Proposed association rule mining algorithm using PSO and domain ontology. They concluded that combining PSO with domain ontology interactively, reduced the number of rules generated without compromising the quality of rules.

Jiawei Han at el, [4] proposed a novel frequent-pattern tree (FP-tree) structure, which is an extended prefix-tree structure for storing compressed, crucial information about frequent patterns, and develops an efficient FP-treebased mining method, FP-growth, for mining the complete set of frequent patterns by pattern fragment growth. Their performance study shows that the FP-growth method is efficient and scalable for mining both long and short frequent patterns, and is about an order of magnitude faster than the Apriori algorithm and also faster than some recently reported new frequent-pattern mining methods. Their suggested approach is to integrate the two algorithms and dynamically select the FP-tree-based and H-structbased algorithms based on the characteristics of current data distribution.

Shafiq Alam at el.,[5] makes two major contributions in this field. Firstly, provides a thorough literature overview focusing on some of the most cited techniques that have been used for PSO-based data clustering. Secondly, they analysed the reported results and highlight the performance of different techniques against contemporary clustering techniques. They also provide a brief overview of our PSO-based hierarchical clustering approach (HPSO-

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 1, January 2017

clustering) and compare the results with traditional hierarchical agglomerative clustering (HAC), K-means, and PSO clustering. They systematically surveyed the work, and presented the results of increasing trends in the literature of swarm intelligence, Particle Swarm Optimization and PSO-based data clustering. Their techniques are novel, collaboration and communication based and simple to implement. PSO has received prompt attention from optimization-based data mining researchers. PSO-based data clustering and hybrid PSO clustering techniques have outperformed many of the contemporary data clustering techniques. The approach has a tendency to be more accurate and to avoid getting trapped in local optima. PSO clustering, PSC clustering, EPSO clustering and HPSO-clustering are some of the popular techniques tested on benchmark datasets.

B. Minaei-bidgoli at el.,[6] proposes a multi-objective genetic algorithm approach for mining association rules for numerical data. Several measures are defined in order to determine more efficient rules. Three measures, confidence, interestingness, and comprehensibility have been used as different objectives for their multi objective optimization which is amplified with genetic algorithms approach. Finally, the best rules are obtained through pareto optimality. This method is based on the notion of rough patterns that use rough values defined with upper and lower intervals to represent a range or set of values. Mutation and crossover operators give powerful exploration ability to the method and allow it to find out the best intervals of existing numerical values. The experimental results show that the generated rules by this method are more appropriate – based on several different characteristics – than the similar approaches' results, and our method outperforms these methods.

Manish gupta [7] determination of the threshold values of support and confidence, affect the quality of association rule mining up to a great extent. Focus of my study is to apply weighted PSO for evaluating threshold values for support and confidence. The particle swarm optimization algorithm first searches for the optimum fitness value of each particle and then finds corresponding support and confidence as minimal threshold values after the data are transformed into binary values. The proposed method is verified by applying the food mart 2000 database of Microsoft sql server 2000.

Anshuman singh sadh at.el, [8] found the frequent patterns to know the effective patterns from the huge data. Then they find positive and negative rules. If we observe the above phenomena then we come to the point that the rule generation is also huge. In this paper they surveyed several aspects of optimization techniques by which we can optimize the association rules. So the main motivation of the survey was to minimize the rule generation or optimize rule generation larger size of rules can be minimized.

III. PROPOSED METHODOLOGY

PSO is a technique used to explore the search space of a given problem in order to find the parameters required to maximize a particular objective. To find the maximum and minimum value of a function or process, optimization mechanism is used.

PSO algorithm works in the following manner. At first it is initialized with a group of random particles (solutions). It then searches for optima by updating generation. Each particle is updated by two "best" value in every iteration. The first best value is called as "pbest", it is the best solution (fitness) it has achieved so far. Another best value is called as "gbest", it is achieved by the particle swarm optimizer obtained so far by any particle in the population. When a particle takes part of the population as its topological neighbours, the best value is a local best and is called as lbest. To evaluate the importance of each particle a fitness value is used. This fitness value for each particle is evaluated by a fitness function. The fitness function for a association rule A-B is given as:

$$Fitness = Support(A \rightarrow B) * Confidence(A \rightarrow B)$$

Where support is the percentage of transactions that contain both A and B whereas confidence is the probability of finding the right hand side of a rule in transactions under the condition that these transactions also contain left hand side. They are defined as follows:

$$Support(A) = \frac{the\ number\ of\ transactions\ that\ contain\ A}{total\ number\ of\ transaction}$$

$$Confidence(A \rightarrow B) = \frac{Support(\text{AUB})}{Support(A)}$$

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 1, January 2017

The overall PSO algorithm consists of three steps, which are repeated until some condition is met:

1. The fitness of each particle is evaluated.
2. Individual and global best fitness position is updated.
3. Velocity and position of each particle is updated.

Figure 2 shows the flow chart for our proposed methodology.

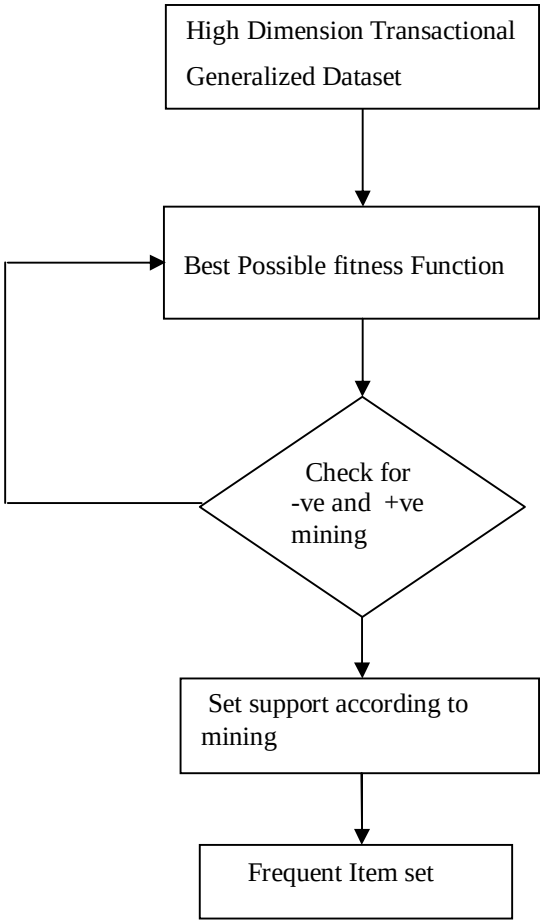

Fig. 2: Flowchart of proposed methodology

V. CONCLUSION

The Proposed methodology will help us for basket data analysis. We will try to analyze the buying habits of the customers, so that we can predict the customers need and help the sellers to sell their items. In our proposed methodology the input will be a multidimensional generalized data set that can be used everywhere and the output will be a data set that contains frequent item set.

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 1, January 2017

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.