



Association Rule Generation by Hybrid Algorithm based on Particle Swarm Optimization and Genetic Algorithm

Monika Verma^{#1}, Roopal Lakhwani^{*2}

^{#1,*2}Department of Computer Science and Engineering,
DIMAT Raipur (C.G.), INDIA,

Abstract— In data mining, association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. It analyzes and present strong rules discovered in databases using different measures of interestingness. The process of discovering interesting and unexpected rules from large data sets is known as association rule mining. This refers to a very general model that allows relationships to be found between items of a database. An association rule is an implication or if-then-rule which is supported by data. The association rules problem was first formulated in and was called the market-basket problem. In his paper trying to hybrid the algorithm, Genetic algorithm (GA) and Particle Swarm Optimization (PSO) is used to improve the fitness value of items by using the dataset.

Keywords—GA, PSO, Association Rule, Market basket, Data mining.

I. INTRODUCTION

In data mining, the process of discovering interesting and unexpected rules from large data sets in known as association rule mining. Association rule mining is a prominent and well researched method for discovering interesting relations between variables in large databases [1]. Agrawal et al. first proposed the issue of the mining association rule in 1993. Due to the rapid growth in the size and number of databases, there is a great need for discovering knowledge hidden in large databases. Knowledge discovery in databases is also known as data mining [2]. Through data mining, we can find useful patterns and rules from databases. These patterns and rules are very useful for decision making of an organization. Therefore, data mining has gained a lot of attentions recently. Specialists from different areas, including machine learning, statistics, artificial intelligence, and expert systems, have developed many powerful tools for data mining. An association rule is an implication or if-then-rule which is supported by data. The association rules problem was first formulated in and was called the market-basket problem [3]. Association rule mining has gained great deal of attention. Even today people use it for mining in KDD. An association rule is an expression $X \Rightarrow Y$, where X and Y are item sets[4].

A. Particle Swarm Optimization

The idea of PSO was initially proposed by Kennedy and Eberhart in 1995. Particle swarm optimization (PSO) is inspired by the social conduct observed in flocks of birds and schools of fish [5]. In PSO, a potential solution to the considered issue is represented by a particle, similar to the individuals in the bird and fish bunch. Each particle goes in the arrangement space and attempts to move toward a better solution by changing its direction and speed in view of its own past experience and the information from the current best particle of the swarm[6]. Recently, its performance was enhanced by using a multi-stage clustering procedure splitting the particles of the main swarm over various of sub-swarms based upon the values of objective functions and the particles positions using different positioning criteria to define three global bests of the swarm as well as employing fuzzy variables to assess the target capacity and imperatives of the problem employing an innovative method to pick the global and individual best positions to enhance the rate of convergence and diversity of solutions and using a self-clustering algorithm to divide the particle swarm into multiple tribes and choosing appropriate evolution techniques to update each particle [7].

B. Genetic Algorithm

The Genetic Algorithm was created by John Holland in 1970[3]. GA is stochastic search algorithm modeled on the procedure of common determination, which underlines natural evolution. GA has been successfully applied in many search, optimization, and machine learning issues[8]. GA lives up to expectations in an iterative manner by generating new populations of strings from old ones. Each string is the encoded paired, real etc., version of a candidate solution. An evaluation capacity relates a wellness measure to each string indicating its fitness for the problem. Standard GA apply genetic operators such selection, crossover and mutation at first random population in order to compute a whole generation of new strings. GA runs to produce answers for progressive generations. Genetic algorithms are a method of "breeding" PC projects and answers for enhancement or search issues by means of simulated evolution. Processes loosely based on natural selection, crossover, and mutation are repeatedly applied to a population of binary strings which represent potential solutions.

II. LITERATURE SURVEY

Numerous Researchers have proposed different methodologies for finding a best solution. **Ahmed A.A.Esmin et al [9]** in Hybrid Particle Swarm Optimization Algorithm with Genetic Mutation proposed a hybrid particle swarm optimization algorithm that uses the mutation procedure to enhance the standard particle swarm optimization (PSO) algorithm. Mutation operators are a necessary piece of developmental calculation strategies, preventing loss of diversity in a population of solutions, which permits a more prominent area of the inquiry space to be covered. The main idea of the HPSOM is to integrate the PSO with genetic algorithm mutation method. **Jiahua Xie and Jie Yang [10]** proposed a novel crossover operator for Particle swarm algorithm to enhance the performance of PSO. This proposed approach is called LPSO employs a Laplace Crossover operator (LC) to create great competitor arrangements. Laplace crossover operator is used to generate new offsprings and contend it with its guardian. After that the fitter particle is updated as the new current particle. **Millie Pant et all [11]** in Particle swarm optimization with crossover operator and its engineering applications proposed an algorithm named Quadratic Interpolation Particle Swarm Optimization (QIPSO) for solving global optimization problems. In this plan the routine system of PSO is modified by including a crossover operator to maintain the level of differing qualities in the swarm populace. Loss of assorted qualities generally takes place when the balance between two antagonists forms investigation (looking of the inquiry space) and exploitation (convergence towards the optimum) is disturbed. To overcome this problem the above algorithm is proposed in this paper. **Jong-Bae Park et all [12]** presented an efficient approach for settling the monetary dispatch (ED) issues with valve-point effects using a hybrid particle swarm optimization (PSO) system. This paper proposed a hybrid PSO (HPSO), which combines the conventional PSO technique with the crossover operation. The hybrid operation, which was generally utilized in the genetic algorithm (GA) methods to increase the exploration also, misuse capacity of the PSO system. The principle objective of ED problem is to minimize the total fuel cost of force plants subjected to the working limitations of a power system. **Payam Chiniforooshan and Shahrooz Shahparvari [13]** proposed a hybrid algorithm that combines differential evolution with particle swarm optimization, in particular HDEPSO. The primary objective of this paper is to achieve faster convergence rate and get better pareto ideal arrangements. **Zhi-Feng Hao et all [14]** in A particle swarm optimization algorithm with hybrid administrator, added to a system in which a hybrid step is added to the standard PSO. The crossover is taken between every molecule's individual best positions. After the crossover, the fitness of the individual best position is contrasted and that of the two off springs and the best one is taken as the new individual best position. In this paper the exploratory result shows that the modified algorithm increases the ability to break away from the nearby ideal. **Anjali Thareja and Dr. Archana Kumar [15]** proposed a modified approach to PSO with crossover operator. In this

paper, uniform hybrid administrator is taken as hybrid method. Real coded uniform crossover generates two offsprings from a couple of folks by consistently supplanting their components on each locus at certain probability.

III. PROBLEM IDENTIFY

A. Algorithmic Issues:

In the association rule for the mining field, the most of the researches were in the first place to improve the algorithmic performance and in to the second place are reduce to the output set by allowing all the possibility to express to constraints on the desired results. Over the past decade, a variety of the algorithms that address all these issues through the refinement of the search strategies, pruning techniques and the data structures have been developed. While most of the algorithms focus on the explicit discovery of all the rules that are satisfy minimal support and confidence constraints for a given dataset, the increasing consideration is being given to the specialized algorithms attempting to improve the processing time or facilitate user interpretation by reducing to the result set size and by incorporating in the domain knowledge.

B. Data Processing Model:

This issue addresses which are parts of the data streams are selected to apply association rule mining. The data streams consist of the ordered sequence of items. Each and every set of items are usually called "transaction". The issue of the data processing model here is to find a way to extract the transactions for association rule mining from the overall the data streams. Because of data streams come continuously and unboundedly, the extracted transactions are changing from time to time. According to the researches there are three stream data processing models, Landmark, Damped and Sliding Windows.

C. Memory Management:

The next fundamental issue we needed to consider as how to optimizing the memory space consumed when running in to the mining algorithm. This includes how to decide the information we must collect from data streams and how to choose a compact in-memory to the data structure that are allows to the information to be stored, updated and retrieved efficiently. Fully addressing these issues for mining algorithm can greatly improve its performance.

D. One Pass Algorithm to Generate Association Rules:

Another fundamental issue are selecting the right type of the mining algorithms Association rules can be found in two steps:

- Finding large item sets (support is \geq user specified support) for a given threshold support
- Generate to the desired association rules for the given confidence.

E. Application Dependent Issues:

Different data stream application situation may have diverse to needs for an association rule mining algorithm. Timeline Query Stream the data come constantly over time. In some applications, user may be interested in triumph association rules based on the data available during a convinced period of the time. Then the storage structure needs to be vigorously adjusted to reproduce the evolution of item set

frequencies over time. How to efficiently store the stream data with timeline and how to powerfully retrieve them during a assured time interval in response to user queries is another significant issue.

IV. METHODOLOGY

The proposed methodology is explained in. Figure 1 shows the working of proposed methodology. A Dataset is initialised firstly, the dataset can be analyzed by several methods. It is usually a good idea to try different approaches, compare their result, and then choose a model that suits the problem well. XLMiner provides a comprehensive set of analysis features based both on statistical and machine learning methods. Data may have quality problems that need to be addressed before applying a data mining technique. The Data may be irrelevant or duplicate, thus pre-processing is necessary. Pre-processing may be needed to make the data more suitable for data mining. There are a number of different tools and methods used for pre-processing. Initialisation of Swarm picks a random position inside the search space and compares the previous best set to the initial position. The fitness value in this study is utilized to evaluate the importance of each particle. The fitness value of each particle comes from the fitness function. The objective of this fitness function is maximization. The larger the particle support and confidence, the greater the strength of the association, meaning that it is an important association rule. The velocity calculation in the PSO algorithm due to pbest and gbest ,a particle a particle has three movement components; the inertia, cognitive, and social component. Mutation changes randomly the new offspring. For binary encoding we can switch a few randomly chosen bits from 1 to 0 . If the Termination Criteria is met, then show the Best run value, otherwise estimate the fitness value again.

V. RESULT AND DISCUSSION

The results obtained for the datasets with the devised algorithm has been compared with standard GA and standard BPSO. The GA, PSO and Hybrid PSO for association rule mining are implemented in MATLAB.

Since an evolutionary algorithm works on principle of random initialization hence each time when algorithm is run user gets different rules, so in particle it is run many times and then rules are collated by some mechanism. In the present research work each algorithm is run and in each run one best rule obtained is picked up and putted in the rule set. For each run the best rule obtained along with its support, confidence and fitness value is displayed.

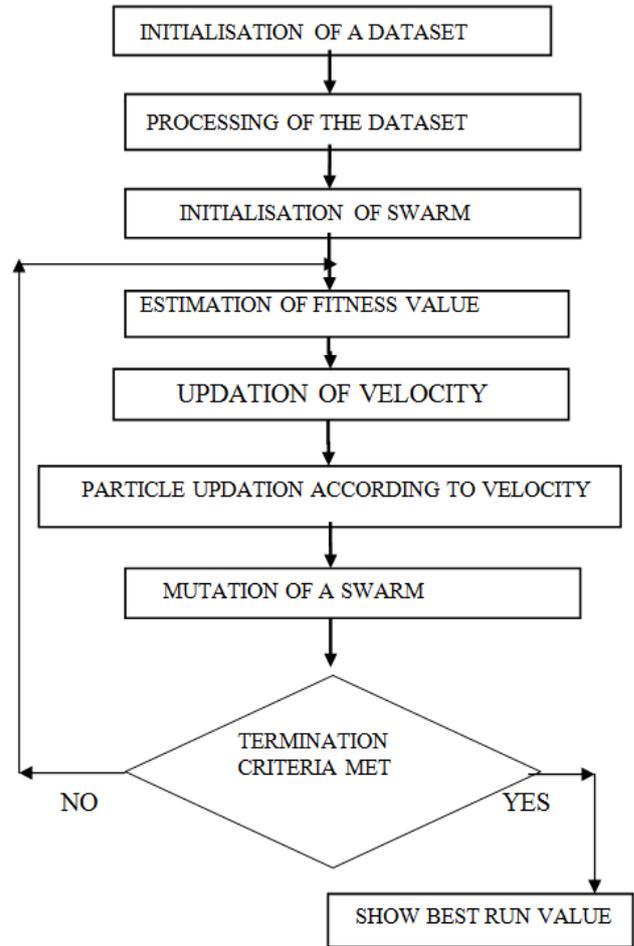


Figure 1: Flowchart of the Proposed Methodology

Table 1: Association Rule for Book dataset by using PSO

Rule No.	Antecedent	Consequent	Support	Confidence	Fitness
1	R, Firstpurch	Gender, F, DoltyBks	0.0105	0.1373	1.4416
2	ID,Gender,ChildBks,DoltyBks	F	0.0385	9.6250	0.3706
3	ID,ChildBks,CookBks	YouthBks	0.0180	0.5625	0.1602

Table 2: Association Rule for Book dataset by using GA

Rule No.	Antecedent	Consequent	Support	Confidence	Fitness
1	CookBks	M	0.1320	0.4681	0.06178
2	F	R, ChildBks	0.0645	0.3007	0.03903
3	DoltyBks	Gender, ChildBks	0.0110	0.0767	0.4070

Table 3: Association Rule for Book dataset by using Hybrid PSO and GA

Rule No.	Antecedent	Consequent	Support	Confidence	Fitness
1	F,ChildBks	ID, M	0.0615	2.1579	6.95314
2	M	Gender, F,CookBks,DoltyBks	0.0465	5.4706	4.37854
3	CookBks	F,YouthBks	0.0660	5.2800	4.20622

VI. CONCLUSION

In this paper, hybrid algorithm combines Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) based association rule miner method by formulating it as a combine a global optimization problem. In this paper, the method is used to bring out the balance between items(product), which will result in accurate prediction of the mined association rule and consistency in performance. A hybrid method combining both genetic algorithm and particle swarm optimisation methodology for mining association rule performs better than the individual performance of both GA and PSO in terms of predictive accuracy and consistency when tested on benchmark dataset.

REFERENCES

- [1] Manisha Gupta, "Application of Weighted Particle Swarm Optimization in Association Rule Mining," International Journal of Computer Science and Informatics (IJCSI) ISSN (PRINT): 2231-5292, Volume-1, Issue-3.
- [2] Suh-Ying Wur and Yungho Leu, "An Effective Boolean Algorithm for Mining Association Rules in Large Databases," National Taiwan University of Science and Technology fjyh.tammyg@cs.ntust.edu.tw.
- [3] Soumadip Ghosh+, Sushanta Biswas*, Debasree Sarkar*, "Partha Pratim Sarkar, Mining Frequent Itemsets Using Genetic Algorithm," International Journal of Artificial Intelligence & Applications (IJAA), Vol.1, No.4, October 2010.
- [4] Manish Saggarr Ashish Kumar Agrawal Abhimanyu Lad, "Optimization of Association Rule Mining using Improved Genetic Algorithms," 2004 IEEE International Conference on Systems, Man and Cybernetics.
- [5] James Kennedy, Russell Eberhart, "Particle Swarm Optimization", 0-7803-2768-3/95/\$4.00 © 1995 IEEE, pp. 1942-1948.
- [6] Zhi-fenghao, zhi-gang wang, hanhuang, "A particle swarm optimization algorithm with crossover operator", Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, 19-22 August 2007.
- [7] M. Andalib Sahnehsaraei, M.J. Mahmoodabadi, M. Taherkhorsandi, K.K. Castillo-Villar and S.M. Mortazavi Yazdi, "A Hybrid Global Optimization Algorithm: Particle Swarm Optimization in Association with a Genetic Algorithm", © Springer International Publishing Switzerland 2015.
- [8] Anubha Sharma and Nirupma Tivari, "A Survey of Association Rule Mining Using Genetic Algorithm", International Journal of Computer Applications & Information Technology Vol. I, Issue II, August 2012 (ISSN: 2278-7720).
- [9] Ahmed A.A. Esmen and Stan Matwin, "HPSOM: A Hybrid Particle Swarm Optimization Algorithm with Genetic Mutation", ICIC International ©2013 ISSN 1349-4198 pp 1919-1934.
- [10] Jiahua Xie, Jie Yang China, "A Novel Crossover Operator for Particle Swarm Algorithm", 2010 International Conference on Machine Vision and Human-machine Interface 978-0-7695-4009-2/10.
- [11] Millie Pant, Member, IAENG, Radha Thangaraj, Member, IAENG, and V. P. Singh, "Particle Swarm Optimization with Crossover Operator and its Engineering Applications.", May 2009..
- [12] Jong-Bae Park, Member, IEEE, Yun-Won Jeong, Joog-Rin Shin, Senior Member, Kwang Y. Lee, Fellow, and Jin-Ho Kim, "A Hybrid Particle Swarm Optimization Employing Crossover Operation for Economic Dispatch Problems with Valve-point Effects.", IEEE, 4 November 2007.
- [13] Jong-Bae Park, Member, IEEE, Yun-Won Jeong, Joog-Rin Shin, Senior Member, Kwang Y. Lee, Fellow, and Jin-Ho Kim, "A Hybrid Particle Swarm Optimization Employing Crossover Operation for Economic Dispatch Problems with Valve-point Effects.", IEEE, 4 November 2007.
- [14] San Francisco. USA. [6] Zhi-Feng Hao, Zhi-Gang Wang, Han Huang, "A Particle Swarm Optimization Algorithm with Crossover Operator.", International Conference on Machine Learning and Cybernetics, Hong Kong, 19-22 August 2007.
- [15] Anjali Thareja and Dr. Archana Kumar, International Journal of Computer Trends and Technology (IJCTT)-volume 14 number 1-Aug 2014