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

Improved Rule Mining Classification Algorithm Based on Particle Swarm Optimization

D.Senthil Kumar, Dr.N.Jayaveeran

Research Scholar, P.G & Research Dept. of Computer Science, KhadirMohideen College, Adirampattinam, Thanjavur District. India.

Associate Professor and Head, P.G & Research Dept. of Computer Science, KhadirMohideen College, Adirampattinam. Thanjavur, India.

ABSTRACT: Rule mining Classification is an important research problem in the emerging field of data mining aimed at finding a small set of rules from the training data set with the optimal obstructed path and predetermined targets. This paper proposes an improved rule mining classification algorithm based on particle swarm optimization (IRMCPSO). The experimental results demonstrate and prove the efficiency of the proposed algorithm, achieve higher predictive accuracy, stronger global optimum search and also confine much smaller rule list than other classification algorithm.

KEYWORDS: Data mining, Classification, PSO

I. INTRODUCTION

The clustering analysis is a process to categorize the sample set of similar characteristics. It's not only an important means of obtaining knowledge from a great set of samples, but also a general technique used in the data mining [12, 13]. There are variety of clustering algorithms such as clustering algorithms based on the model, the layer, the flat surface partition, the density, the mesh and sub-space and so on [14]. Commonly there are two problems in most clustering algorithms: first, some parameters have to be given in advance under the condition of no prior knowledge which is very hard to determine those parameters. The second is hard to assure the time and space efficiency of clustering analysis to a big sample set or a high dimensions sample set.

Particle Swarm Optimization (PSO) incorporates swarming behaviours observed in flocks of birds, schools of fish, or swarms of bees, and even human social behaviour, from which the idea is emerged. PSO is a population-based optimization tool, which could be implemented and applied easily to solve various function optimization problems, or the problems that can be transformed to functional optimization problems. As an algorithm, the main strength of PSO is its fast convergence that compares favourably with many global optimization algorithms like Genetic Algorithms (GA), Simulated Annealing (SA) and other global optimization algorithms. Population-based heuristics are more costly because of their dependency directly upon function values rather than derivative information. They are however susceptible to premature convergence, which is especially the case when there are many decision variables or dimensions to be optimized.

The most optimist solution can be worked out in particle swarm optimization algorithm by the cooperation of each individual. The particle without quality and volume serves as each individual, and the simple behavioural pattern is regulated for each particle to show the complexity of the whole particle swarm.

In PSO, the potential solution called particles fly through the problem space by following the current optimum particles. Each particles keeps tracks of its coordinates in the problem space that are associated with the best solution (fitness) achieved so far. This value is called as pbest. Another best value that is tracked by the particle swarm

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

optimizer is the best value, obtained so far by any particle in the neighbors of the particle. This value is called lbest. When a particle takes all the population as its topological neighbors, the best value is a global best and is called gbest. The particle swarm optimization concept consists of, at each time step, changing the velocity of (accelerating) each particle toward its pbest and lbest (for lbest version). Acceleration is weighted by random term, with separate random numbers being generated for acceleration towards pbest and lbest locations. After finding the best values, the particle updates its velocity and positions with following equations.

Mobile Ad Hoc Networks (MANETs) consists of a collection of mobile nodes which are not bounded in any infrastructure.Nodes in MANET can communicate with each other and can move anywhere without restriction. This non-restricted mobility and easy deployment characteristics of MANETs make them very popular and highly suitable for emergencies, natural disaster and military operations.

Nodes in MANET have limited battery power and these batteries cannot be replaced or recharged in complex scenarios. To prolong or maximize the network lifetime these batteries should be used efficiently. The energy consumption of each node varies according to its communication state: transmitting, receiving, listening or sleeping modes. Researchers and industries both are working on the mechanism to prolong the lifetime of the node's battery. But routing algorithms plays an important role in energy efficiency because routing algorithm will decide which node has to be selected for communication.

The main purpose of energy efficient algorithm is to maximize the network lifetime. These algorithms are not just related to maximize the total energy consumption of the route but also to maximize the life time of each node in the network to increase the network lifetime. Energy efficient algorithms can be based on the two metrics: i) Minimizing total transmission energy ii) maximizing network lifetime. The first metric focuses on the total transmission energy used to send the packets from source to destination by selecting the large number of hops criteria. Second metric focuses on the residual batter energy level of entire network or individual battery energy of a node [1].

II. REVIEW OF LITERATURE

Researchers have proposed several numbers of methodologies for finding a best solution.

The idea of Particle Swarm Optimization (PSO) is a population-based optimization method initially proposed by Kennedy and Eberhart in 1985[10]. PSO is inspired by the social conduct observed in flocks of birds and schools of fish; 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. The information from the current best particle of the swarm, updates the current generation of particles (each particle is a candidate solution to the problem) using the information about the best solution obtained by each particle and the entire population [11].

Monika Verma and RoopalLakhwani proposed a 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 is prediction of the mined association rule and consistency in performance [1].

GPSO partitions the population into two subpopulations based on the fitness values. Evolution of the upper half is performed through GA and lower half through PSO. The subpopulations are combined at the end of each generation and the process repeated. GA helps in maintaining the exploration concept and PSO takes care of exploitation thereby maintaining the balance between exploration and exploitation avoiding premature convergence. Combining the subpopulations after each iteration preserves the desirable datasets thereby resulting in consistency of results over all iterations is proposed in K.Indira and S.Kanmani[2].

Jiang et al.(2011) proposed to improve the performance of PSO an enhanced EA based on the characteristics of PSO, multi-parent crossover algorithm and differential evolution (DE)[3].

Tang and Wang proposed a hybrid multi-objective EA incorporating the concepts of personal best and global best in PSO and multiple crossover operators to update the population, maintains a non-dominated archive of personal best [4].

Ykhlef proposed the application of association rule mining to time series using GA by discretising the segment time series to a number of shapes enhanced the prediction of time series[5].

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

Rodriguez, M. et al. presents an Efficient Distributed Genetic Algorithm for classification Rule extraction in data mining (EDGAR), which promotes a new method of data distribution[6].
Hong-yuan Shen et al proposed a mountain clustering method constructs a mountain function according to the density of the sample, but it is not easy to find all peaks of the mountain function. The improved PSO algorithm is used to find peaks of the mountain function [7].
YannisMarinakis et al proposed Hybrid PSO-GRASP algorithm for the solution of the clustering problem is a two phase algorithm which combines a PSO algorithm for the solution of the feature selection problem and a GRASP for the solution of the clustering problem. Due to the nature of stochastic and population-based search [8].
Xueping Zhang et al present the process of doing so, we first use PSO algorithm via MAKLINK graphic to get the optimal obstructed path, and then we developed PSO K-Medoids SCOC (PKSCOC) algorithm to cluster spatial data with obstacles constraints [9].

The disadvantages of these algorithms are:

- The above said methodologies failed to give constant result
- Finds the clustering center of the sample data is slow and inaccurate.
- Finds most of local optimization of the mountains function, and is hard and worthless.

The proposed method, Improved Rule Mining Classification Algorithm Based on Particle Swarm Optimization is a consequence to overcome the above said drawbacks.

III. PROPOSED ALGORITHM

Now a day most of the researchers concentrate on improving the performance of the existing standard algorithm in the field of association rule mining. All the possible output sets are tuned to reduce in size and the desired result express the constraints. Different numbers of algorithms show all these issues through this refinement of the search approaches, pruning methods and developing the data structure. Over the past decade, huge numbers of algorithms are explicitly focused on finding association rules. These rules fulfil minimal support and confidence constraints for a given dataset. The specialized algorithms which improves the domain knowledge are combined and reduced the result set size by facilitate the user or the processing time.

It addresses the problem of finding selected data streams to apply association rule mining. The item sets are extracted in the ordered sequence from data streams found above. These item sets are typically called as transaction. The overall data streams are used to extract the transactions for association rule mining. Changing transactions are extracted frequently because of data streams come absolutely and continuously.

Association rule mining algorithm is another fundamental problem in optimizing the memory space consumed. It includes the data stream that provides the decision making information in a suitable data structure which allows the information to be retrieved, updated and stored effectively in compact format of memory storage.

The challenge here is to choose the accurate type of the association rule mining algorithms. It can be found in two steps:

a) Discover large item sets (support is ≥ user specified support) for a given threshold support
b) Producing the cause in terms of association rules for the given confidence level set

Application Dependent Issues:

Each association rule mining algorithm is different in nature and it may use different data stream application. Over the period of time the data come in a constant timeline through query stream. Most of the association rule algorithms are used to achieve the user interest based on the data available during a convinced period of the time. The evolution of frequent item sets needs to be vigorously adjusted with the storage structure. The significant issue here is how to store the stream data efficiently with timeline and how to retrieve eloquently during a assured time interval to respond the user queries.

IV. METHODOLOGY

The proposed algorithm is expressed as follows and shown in figure fig.1. The initial task is to prepare data; transform raw transaction data into the required structure, load the transformed data into the mining database. Quality of data must be checked before applying data mining technique. The Data may be irrelevant or duplicate. Hence the data pre-processing is becoming a mandatory task. It makes the data more suitable for data mining. In general, the transaction data mining rules are created by using the training data set by eliminating the unwanted fields/data (like student name, reg.no, DOB, contact no. etc.) from transaction database. Today, different methods and tools are available to handle data pre-processing. Swarm technique elites the search space

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

inside from a random position and associates the best dataset to the initial position. Next, evaluate the importance of each particle utilized and fitness value in this study.

The fitness function comes to each particle using the fitness value and the main objective of the fitness function is maximization. The strength and an important of the association rule is the particle support and confidence. The PSO algorithm used to calculate velocity of pbest and gbest. There are three particle components available which are inertia, cognitive, and social component. Randomly changes the new offspring. The binary encoding switch randomly chooses bits from 1 to 0. Lastly, the knowledge extract the best run value, otherwise estimate the fitness value again.

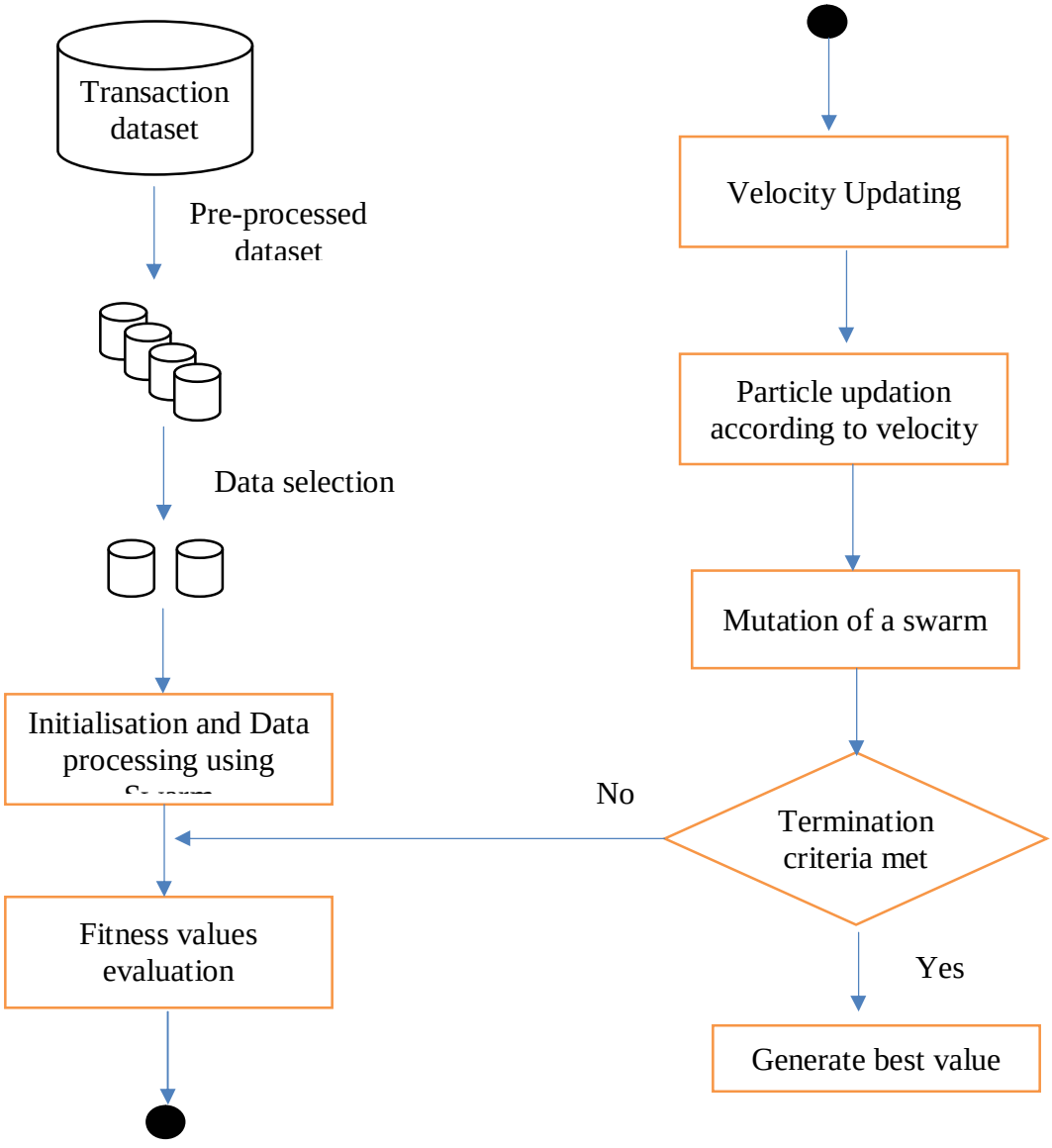

Fig.1 Proposed Method IRMCPSO

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. RESULT AND DISCUSSION

The results obtained from the hostel inmate health care datasets with the developed algorithm has been compared with standard GA and standard PSO algorithm. The GA, PSO and IRMCPSO algorithms for association rule mining are implemented in MATLAB. The performance of PSO, GA and IRMCPSO are given in below tables Table1, Table2 and Table3.

Table 1: Association Rule for hostel inmate health care dataset by using PSO

Rule No.	Antecedent	Consequent	Support	Confidence	Fitness
1	1 to 3 litres, hot drinks, tawa fry, vegetarian	Acidity	0.03	9.62	0.37
2	3 to 5 litres, beverages, tawa fry, non-vegetarian	Obesity	0.01	0.56	0.16
3	below 1 litre , soft drinks, tawa fry, vegetarian	Ulcer	0.01	0.13	1.44

Table 2: Association Rule for hostel inmate health care dataset by using GA

Rule No.	Antecedent	Consequent	Support	Confidence	Fitness
1	1 to 3 litres, hot drinks, tawa fry, vegetarian	Acidity	0.01	0.07	0.40
2	3 to 5 litres, beverages, tawa fry, non-vegetarian	Obesity	0.13	0.46	0.06
3	below 1 litre , soft drinks, tawa fry, vegetarian	Ulcer	0.06	0.30	0.03

Table 3: Association Rule for hostel inmate health care dataset by using IRMCPSO

Rule No.	Antecedent	Consequent	Support	Confidence	Fitness
1	1 to 3 litres, hot drinks, tawa fry, vegetarian	Acidity	0.05	1.15	6.95
2	3 to 5 litres, beverages, tawa fry, non-vegetarian	Obesity	0.05	1.47	4.37
3	below 1 litre , soft drinks, tawa fry, vegetarian	Ulcer	0.15	2.46	4.20

VI. CONCLUSION

The proposed algorithm improved rule mining classification algorithm based on particle swarm optimization (IRMCPSO) that combines Genetic Algorithm (GA) and Particle Swarm Optimization (PSO). This proposed method brings out the health problems of students concerning uncertain habitations and the daily water consumption. It predicts the accurate result through mined association rule with consistency in performance as well. The proposed method IRMCPSO by combining Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) provide better performance than the individual performance of both GA and PSO in terms of predictive accuracy and consistency when tested on benchmark dataset.

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

REFERENCES

1. Monika Verma and RoopalLakhwani, Association Rule Generation by Hybrid Algorithm based on Particle Swarm Optimization and Genetic Algorithm, IJCSET, Vol 5, Issue 12, pp. 423-426, 2015.
2. K. Indira and S. Kanmani, Mining association rules using hybrid genetic algorithm and particle swarm optimisation algorithm, Int. J. Data Analysis Techniques and Strategies, Vol. 7, No. 1, 2015, pp. 59 – 76.
3. Jiang, D., Zeng, S., Wang, H. and Wu, Z. (2011) 'Particle swarm optimizer with hybrid multi-parent crossover and discrete recombination', International Journal of Intelligent Information and Database Systems, Vol. 5, No. 6, pp.597–617.
4. Tang, L. and Wang, X. (2012) 'A hybrid multi-objective evolutionary algorithm for multi-objective optimization problems', IEEE Transactions on Evolutionary Computation, Vol. 17, No. 1, pp.20–45.
5. Ykhlef, M. (2012) 'Association mining of dependency between time series using genetic algorithm and discretisation', International Journal of Business Intelligence and Data Mining, Vol. 6, No. 1, pp.55–70.
6. Rodriguez, M., Escalante, D.M. and Peregrin, A. (2011) 'Efficient distributed genetic algorithm for rule extraction', Applied Soft Computing, Vol. 11, No. 11, pp.733–743.
7. Hong-yuan Shen, Xiao-qi Peng, Jun-nian Wang, and Zhi-kun Hu (2005), 'A Mountain Clustering Based on Improved PSO Algorithm', Springer-Verlag Berlin Heidelberg, LNCS 3612, pp. 477 – 481.
8. YannisMarinakis, Magdalene Marinaki, and Nikolaos Matsatsinis, 'A Hybrid Particle Swarm Optimization Algorithm for Clustering Analysis', Springer-Verlag Berlin Heidelberg, LNCS4654, pp. 241 – 250, 2007
9. Xueping Zhang, Jiayao Wang, Zhongshan Fan, and Xiaoqing Li, 'A Particle Swarm Optimization Method for Spatial Clustering with Obstacles Constraints', Springer-Verlag Berlin Heidelberg, LNAI 4682, pp. 367–376, 2007.
10. James Kennedy, Russell Eberhart, "Particle Swarm Optimization", 0-7803-2768-3/95/$4.00 0 1995 IEEE, pp. 1942-1948.
11. 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.
12. David Hand, HeikkiMannila, Padhraic Smyth: Principles of Data Mining, The MIT Press,2001.
13. Jain A K, Dubes R C: A1gorithms for Clustering Data. Englewood Cliffs, NJ: Prentice Hall,1988.
14. Xing xiaoshuai, Jiao Licheng: Clustering Method in the Field of Data Mining, Journal ofCircuits and Systems, 8(1), 2003.

BIOGRAPHY

Dr.N.Jayaveeranhas pursued his Ph.d from Bharathidasan university, Tiruchirappali-24, Tamil Nadu,and he is currently working as Associate Professor and Head, PG and Research Department of Computer Science, KhadirMohideen College, Adirampattinam, Tamil Nadu. He has published several research articles in international and national journals. His area of research includes Roughset Theory, Data mining, Network Security, Bigdata analysis and cloud computing. He has 29 years of teaching experience and he is currently guiding eight research scholars.

Mr. D. Senthil Kumarpursued Master Degree of Computer Science in 1999 from Bharathidasan University, Tiruchirapalli - 24, Tamil Nadu and Master of Philosophy in 2007 from Bharathidasan University, Tiruchirappali-24, Tamil Nadu. He is currently pursuing PhD in Bharathidasan University and he is currently working as Lecturer in the Department of Computer Science in Bharathidasan University Model College, Aranthangi, Tamil Nadu. He has published several papers in various conference and his research work is mainly focused on Data mining and Cloud computing. He has 14 years of teaching experience.