Available online at www.elixirpublishers.com (Elixir International Journal)

Computer Science and Engineering

Elixir Comp. Sci. & Engg. 46 (2012) 8563-8566

ISSN: 2229-712X

Enhancing particle swarm optimization using chaotic operators for association rule mining

K.Indira[1] and S.Kanmani[2]

[1]Department of Computer Science and Engineering, Pondicherry Engineering College, Puducherry, India.
[2]Department of Information Technoogy, Pondicherry Engineering College, Puducherry, India.

ARTICLE INFO

Article history:
Received: 27 March 2012;
Received in revised form:
15 May 2012;
Accepted: 28 May 2012;

Keywords
Association Rule Mining,
Particle Swarm Optimization,
Chaotic Particle Swarm optimization,
Computational Accuracy.

ABSTRACT

Association Rule (AR) mining is a data mining task that discovers interesting relations between variables in databases. The main focus of research in association rule mining is on improving the computational efficiency of mining the rules. The conventional methods available for mining association rules depend on the threshold values of minimum support and minimum confidence. The setting up of these values needs great care and knowledge about the application domain. This paper deals with mining association rules using chaotic Particle swarm optimization (cPSO). Particle swarm optimization (PSO) is simple but powerful population based search technique for solving optimization problems. Chaotic PSO modifies the velocity updation function by introducing chaotic operators derived from chaotic maps. Both PSO and chaotic PSO generate AR with consistency and better computational accuracy when tested on three datasets. The range of attribute values in the dataset is found to affect the performance of the system.

Introduction

Wide availability of voluminous data and extracting useful information and knowledge from them has scored data mining a prominent place in the information industry. Data mining consists of several tasks depending on application domain and user interest. Association rule mining is one among the most widely used task in data mining [1, 2]. Association rule mining is discovery of interesting patterns or relations between variables in large databases. These relationships can be represented as IF–THEN statement. IF <some conditions are satisfied > THEN <predict some values of other attribute(s) >. The conditions associated in the IF part is termed as antecedent and those with the THEN part is called the consequent.

Given a set of items, objects, or binary variables $I = \{I_1, I_2, . . . , I_n\}$, an AR is formally defined[1] as an implication $X \rightarrow Y$ where $X,Y \subseteq I$ and $X \cap Y = \emptyset$. Both the antecedent (X) and the consequent (Y) are interpreted as a conjunction of the variables they contain.

Apriori algorithm has been the majorly used conservative method for mining association rule till recently. Many variations have been introduced in Apriori with focus on improving its efficiency and accuracy. Two parameters namely minimum support and minimum confidence set by the decision maker determines the efficiency of the system. This makes the algorithm lack in efficiency and effectiveness.

Evolutionary Algorithms are a subset of a larger set of heuristic methods. Heuristics are approximate methods attempting to find “very good” solutions in very large search spaces in reasonable time. Genetic Algorithm (GA) and Particle swarm optimization (PSO) are two population based evolutionary search methods and move from set of points (population) to another set of points in a single iteration with likely improvement using set of control operators. This study explores the application of PSO and Chaotic PSO for mining association rules.

The concept of particle swarm optimization was put forth by Kennedy and Eberhart [3,4,5]. It has been proved to be efficient at solving global optimization and engineering problems [6]. The advantages of PSO over many other optimization algorithms are its implementation simplicity and ability to reasonable convergence.

The remainder of the paper is organized as follows. Section 2 briefly discusses the literature review on particle swarm optimization. Section 3 is about the methodology applied for PSO and chaotic PSO to mine association rules in detail. The evaluation results and discussion are explored in section 4 followed by conclusion in section 5.

Literature Review

Association rules aim in extracting important correlation among the data items in the databases. Shichao Zhang et al [7] have given the association rule mining methods and the importance of rule interestingness measures. Association rule, basically extracts the patterns from the database based on the two measures such as minimum support and minimum confidence. To select the best measures for mining rules based on constraints such as multiple criteria is applied [8]. The support and confidence measures are described as stated in [9] for mining frequent itemset and association rule generation.

The PSO method is a member of the wide category of swarm intelligence methods. Particle swarm optimization incorporates swarming behaviors observed in flocks of birds, schools of fish, or swarms of bees, and even human social behavior, from which the idea is emerged [4, 10]. PSO is a population-based optimization tool, which could be implemented and applied easily to solve various optimization problems. As an algorithm, the main strength of PSO is its fast convergence, which compares favorably with many global

Tele:
E-mail addresses: induharini@gmail.com

optimization algorithms like Genetic Algorithms. PSO can be easily implemented and is computationally inexpensive since its memory and CPU speed requirements are low [11].

The fixing up of the best position [12) for particles after velocity updation by using Euclidean distance helps in generating the best particles. The chaotic operator based on Zaslavskii maps when used in velocity update equation [13] proved to enhance the efficiency of the method. The soft adaptive particle swarm optimization algorithm [14] exploits the self adaptation in improving the ability of PSO to overcome optimization problems with high dimensionality. The particle swarm optimization with self adaptive learning [15] aims in providing the user a tool for various optimization problems. The problem of getting struck at local optimum and hence premature convergence is overcome by adopting self adaptive PSO [16] where the diversity of population is maintained. This copes up with the deception of multiple local optima and reduces computational complexity.

Methodology

This section proposes a novel algorithm based on chaotic maps for mining association rules. The basis of association rule and the parameters related with association rule are discussed. The movement of particles from one generation to next generation is based on the fitness value. Calculation of the fitness value through the fitness function is described.

The accuracy measure for validating the mined rules is predictive accuracy. The method for measuring the predictive accuracy is defined followed by the particle swarm optimization algorithm for mining association rules. The proposed chaotic PSO method for mining AR is discussed.

Association Rules

Association rule mining finds interesting associations and/or correlation relationships among large set of data items. Association rules show attributes value conditions that occur frequently together in a given dataset. Each association rule has two quality measurements, support and confidence. Support implies frequency of occurring patterns, and confidence means the strength of implication and is defined as follows:

An itemset, X, in a transaction database, D, has a support, denoted as sup(X) or simply p(X), that is the ratio of transactions in D containing X. Or

$$\sup(x) = \frac{No.of\ transactions\ containing\ X}{Total\ No.of transactions} \quad (1)$$

The confidence of a rule X => Y, written as conf(X=>Y), is defined as

$$conf\ (X \rightarrow Y) = \frac{\sup(X \cup Y)}{\sup(X)} \quad (2)$$

Fitness Function

The importance of each particle is studied utilizing fitness function. Fitness value is evaluated using the fitness function. The objective of the fitness function is maximization. Equation 3 describes the fitness function.

$$Fitness(x) = conf(x) \times \log(\sup(x) \times length\ (x) + 1) \quad (3)$$

where fitness(x) is the fitness value of the association rule type x, sup(*x*) and conf(*x*) are as described in equation 1 and 2 and *length(x*) is length of the association rule type *x*. If the support and confidence factors are larger then, greater is the strength of the rule with more importance.

Predictive Accuracy

Predictive accuracy measures the effectiveness of the rules mined. The mined rules must have high predictive accuracy.

$$Predictive\ accuracy = \frac{|X\ \&\ Y|}{|X|} \quad (4)$$

where |X&Y| is the number of records that satisfy both the antecedent X and consequent Y, |X| is the number of rules satisfying the antecedent X.

Particle Swarm Optimization Algorithm

PSO simulates the behaviors of bird flocking. Suppose the following scenario: a group of birds are randomly searching food in an area. There is only one piece of food in the area being searched. All the birds do not know where the food is. But they know how far the food is in each iteration. The best strategy to find the food is to follow the bird which is nearest to the food.

PSO is initialized with a group of random particles (solutions) and then searches for optima by updating generations. During all iterations, each particle is updated by following the two "best" values. The first one is the best solution (fitness) it has achieved so far. The fitness value is also stored. This value is called "lbest" The other "best" value that is tracked by the particle swarm optimizer is the best value obtained so far by any particle in the population. This best value is a global best and is called "gbest" After finding the two best values, each particle updates its corresponding velocity and position with Equations 5 and 6 as follows

$$v_i^{new} = v_i^{old} + c_1 rand(\square)(lbest - x_i) + c_2 rand(\square)(gbest - x_i) \quad (5)$$

$$x_i^{new} = x_i^{old} + v_i^{new} \quad (6)$$

Where v_{id} is the particle velocity of the i^{th} particle, x_i is the current particle, i the particle number, rand ( ) is random number in (0,1), c1 the individual factor and c_2 the societal factor. The flowchart given in figure 1 shows the steps of PSO algorithm in detail.

Both c_1 and c_2 are usually set to be 2 in all literature works analyzed and hence the same adopted here. The velocities of particles are clamped to a maximum velocity V_{max}. If the sum of accelerations would cause the velocity to exceed V_{max}, which is a parameter specified by the user, then the velocity is limited to V_{max}.

Chaotic Particle Swarm Optimization

The canonical PSO tends to struck at local optima and thereby leading to premature convergence when applied for solving practical problems. To improve the global searching capability and escape local optima chaos is introduced in PSO [17]. Chaos is a deterministic dynamic system and is very sensitive and dependent on its initial conditions and parameters. The common method of generating chaotic behavior is based on Zaslavskii map[13]. This representation of map involves many variables. Setting right values for all these variables involved increases the complexity of the system and erroneous values might bring down the accuracy of the system involved. Logistic map and tent map are also the most frequently used chaotic behavior. The drawback of these maps is that the range of values generated by both the maps after some iteration becomes fixed to a particular range. To overcome this defect the tent map undisturbed by the logistic map [18] is introduced as the chaotic behavior. The new chaotic map model proposed with the following equation.

$$u_{k+1} = 4 * u_k(1 - u_k) \qquad 0 \le u_k \le 1 \quad (7)$$

$$v_{k+1} = \begin{cases} 1/1.001 * (2 * v_k + 0.001 * u_{k+1}) & 0 \le v_k \le 0.5 \\ 1/1.001 * (\square)2 * [(1 - v]_k) + 0.001 * u_{k+1} & 0.5 \le v_k \le 1 \end{cases}$$

The initial value of u_0 and v_0 are set to 0.1. The slight tuning of initial values of u0 and v_0 creates wide range of values with good distribution. The chaotic operator *chaotic_operator(k)* = v_k is designed therefore to generate different chaotic operators by tuning u_0 and x_0. The value u_0 is set to two different values for generating the chaotic operators 1 and 2.

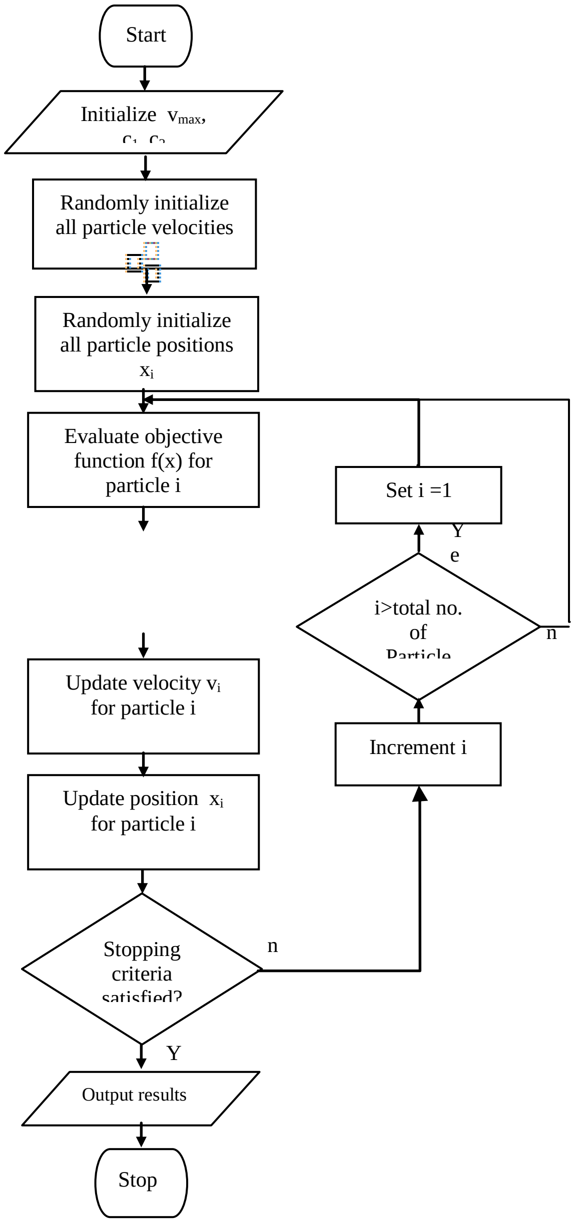

Figure 1. Flowchart of PSO

The velocity updation equation based on chaotic PSO is given in equation 8.

$$v_i^{new} = v_i^{old} + c_1 * Chaotic_{operator_1}(lbest - x_i) + c_2 * Chaotic_{operator_2}(gbest - x_i)$$

The algorithm of cPSO is depicted below

Evaluation Results and Discussion

Three datasets from University of California Irvine Machine Learning Repository namely Car Evaluation, Haberman's Survival and Lenses are taken up for evaluating both PSO and Chaotic PSO algorithm.

Car evaluation dataset contains 1728 instances of records with 6 attributes and the swarm size set was 700. The Haberman's Survival dataset contains 306 instances of records with 3 attributes and the swarm size set was 300 and the Lenses dataset contains 24 instances of records with 3 attributes and the swarm size set was 24. The experiment was conducted on Microsoft windows XP platform using Java as the developing environment. The initial velocity set was 0 for all the datasets and the learning factors c_1 and c_2 is 2. Maximum number of iterations carried out was 50.

Step1. Initialize the population - locations and velocities

Step 2. Evaluate the fitness of the individual particle (lBest)

Step 3. Keep track of the individual highest fitness (gBest)

Step 4. Generate chaotic operator1 and chaotic operaor2.

Step5. Modify velocities based on chaotic operator based velocity updation function

Step 6. Update the particles position

Step 7. Terminate if the condition is met

Step 7. Go to Step 2

The results of the experiment are shown in figures 2 to 4.

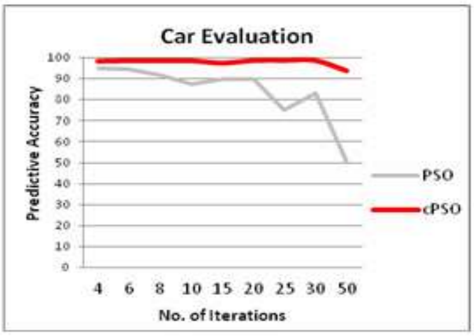

Figure 2. Predictive Accuracy for Car Evaluation Dataset

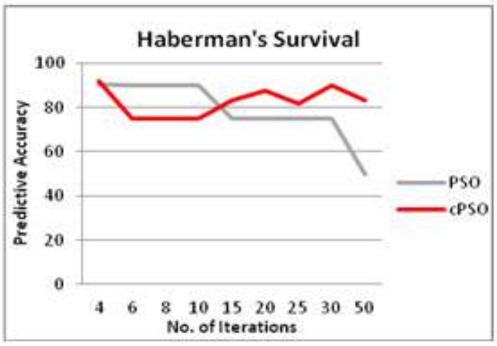

Figure 3. Predictive Accuracy for Lenses ataset

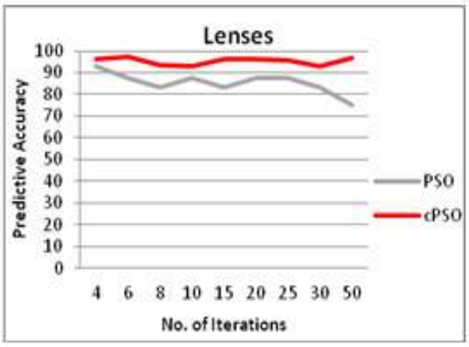

Figure 4. Predictive Accuracy for Haberman's Dataset

Discussion

The PSO and cPSO algorithm of rule mining was performed on three datasets. The maximum predictive accuracy achieved from the execution is plotted in the figures. Dataset taken up for experimenting varies from medicine to commercial datasets. The swarm size set depends on the dataset size. The chaotic operator is generated by varying the initial value u_0 alone. The PSO method generates consistent accuracy when compared to GA where the reproductive operators tend to diversify the population sometimes. The PSO based on chaotic operators generates rule which are highly consistent when compared to PSO. This could be noted clearly from figure 2 and 3 where the lines marked for cPSO are more linear than the PSO lines.

In figure 4 for the haberman's survival dataset the linearity of the lines is lost. The haberman's survival dataset contains age of the patient as one of the attributes. This attribute has values ranging from 30 to 83 while the lenses and car evaluation dataset has got categorical values ranging from 2 to 4 values alone. The distribution of values for the haberman's survival dataset is more when compared to lenses and car evaluation dataset. This difference brings the inconsistency in predictive accuracy of haberman's survival dataset when compared to Lenses and Car evaluation dataset where the predictive accuracy is consistent. The cPSO performs better than PSO in all the three datasets.

The convergence is arrived with minimum time for both Lenses and Car evaluation dataset but varies throughout the execution time for haberman's survival dataset.

Conclusion

Extracting meaningful information from large databases due to the huge development in information technology has become an important issue. So mining association rule needs more methods and techniques for meeting the requirements. This paper focuses on association rule mining carried out using Particle swarm optimization algorithm and chaotic PSO. The chaotic operators are introduced in velocity updation function of PSO and rules were mined using the chaotic PSO algorithm. When compared to other evolutionary methods PSO is found to be simple and the parameters involved in PSO could be configured with little difficulty.

Both PSO and Chaotic PSO produce consistent results when implemented on three different datasets. The values of the attributes involved and their range affects the consistency of the system. In future the same algorithm could be tried on other datasets with varying applications to analyze the Performance of cPSO in all domains.

References

Agrawal R, Imielinski T, Swami A, Mining Association Rules Between Sets of Items in Large Databases, *Proc. of the 1993 ACM SIGMOD International Conference on Management of Data*, New York, ACM Press, 1993, 207–216.

Piatetsky-Shapiro G, Frawley WJ, eds, Discovery, Analysis, and Presentation of Strong Rules, *Knowledge Discovery in Databases*, Cambridge, MA, AAAI/MIT Press, 1991.

J. Kennedy and R. C. Eberhart, Particle Swarm Optimization, *IEEE International Conference on Neural Networks.* Piscataway, NJ, 1995, pp. 1942-1948.

Kennedy and R. C. Eberhart, *Swarm intelligence,* Morgan Kaufmann, 2001.

A. P. Engelbrecht, *Fundamentals of Computational Swarm Intelligence*, John Wiley & Sons 2006.

Parsopoulos KE, Plagianakos VP, Magoulas GD and Vrahatis MN, Streching Technique for Obtaining Gloabal Minimizers Through Particle Swarm Optimization, *Proceedings Particle Swarm Optimization Workshop,* pp. 22–29, 2001.

Shichao Zhang and Xindong Wu, Fundamentals of Association Rules in Data Mining and Knowledge Discovery, *WIREs Data mining Knowledge Discovery*, vol. 1 March / April 2011.

Philippe Lenca, Patrick Meyer, Bonoit vaillant, Stephae lallich, On Selecting Interestingness Measures for Association Rules: User Oriented Description and Multiple Criteria Decision Aid, *European Journal of operation research* , Vol.184, pp. 610 – 626.

Jiawei Han and Micheline Kamber, *Data Mining: Concepts and Techniques*, 2nd edition. Elsevier publications.

Parsopoulos, K. E., and Vrahatis, M. N, On the Computation of all Global Minimizers through Particle Swarm Optimization, *IEEE Transactions on Evolutionary Computation*, 8(3), pp.211-224. 2004.

Eberhart R.C, Simpson P, Dobbins.R, *Computational Intelligence PC Tools*, Academic Press, 1996.

Kuo R.J, Chao C.M, Chiu Y.T, Application of Particle Swarm Optimization in Association Rule Mining, *Applied Soft Computing*, pp.323-336, 2011.

Bilal Atlas, Erhan Akin, Multi-objective Rule Mining using a Chaotic Particle Swarm Optimization Algorithms, *Knowledge Based System*, *Vol.*,23, pp. 455-460, 2009.

Yamina Mohammed Ben Ali, Soft Adaptive Particle Swarm Algorithm for Large Scale Optimization, *IEEE fifth International Conference on Bio Inspired Computing*, pp.1658-1662, 2010.

Yu Wang, Bin LI, Thomas Weise, Jianyu Wang, Bo Yun, Qiondjie Tian, Self-adaptive Learning Based on Particle Swarm Optimization, *Information Science*, 181, pp.4515-4538, 2011.

Feng Lu, Yanfen Ge, Liqun Gao, Self Adaptive Particle Swarm Optimization Algorithm for Global Optimization, *Sixth IEEE International Conference on Natural Computation*, pp.2692-2696, 2010.

W.J. Kong, W.J. Cheng, J.L. Ding, T.Y. Chai, A Reliable and Efficient Hybrid PSO Algorithm for Parameter Optimization of LS-SVM for Production Index Prediction Model, *Third International Symposium on Computational Intelligence and Design*, vol.2, pp.140-143,2010.