

April 2012

## Application of Weighted Particle Swarm Optimization in Association Rule Mining

Manisha Gupta

Shri Ram Institute of Technology, Jabalpur, Rajiv Gandhi Produyogiki Vishwavidhayalaya, Bhopal, India,  
manishagupta.cse@gmail.com

Follow this and additional works at: <https://www.interscience.in/ijcsi>



Part of the [Computer Engineering Commons](#), [Information Security Commons](#), and the [Systems and Communications Commons](#)

---

### Recommended Citation

Gupta, Manisha (2012) "Application of Weighted Particle Swarm Optimization in Association Rule Mining," *International Journal of Computer Science and Informatics*: Vol. 1 : Iss. 4 , Article 11.

DOI: 10.47893/IJCSI.2012.1053

Available at: <https://www.interscience.in/ijcsi/vol1/iss4/11>

This Article is brought to you for free and open access by the Interscience Journals at Interscience Research Network. It has been accepted for inclusion in International Journal of Computer Science and Informatics by an authorized editor of Interscience Research Network. For more information, please contact [sritampatnaik@gmail.com](mailto:sritampatnaik@gmail.com).

# Application of Weighted Particle Swarm Optimization in Association Rule Mining

Manisha Gupta

Shri Ram Institute of Technology, Jabalpur, Rajiv Gandhi Prodyogiki Vishwavidyalaya, Bhopal, India  
E-mail : manishagupta.cse@gmail.com

---

**Abstract** - 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. I am expecting that the particle swarm optimization algorithm will suggest suitable threshold values and obtain quality rules as per the previous works [1].

**Key words** - Particle Swarm Optimization, Association Rule Mining, Inertia Weight, Weighted PSO, Optimization, Support & Confidence.

---

## I. INTRODUCTION

In different kinds of information databases, such as scientific data, medical data, financial data, and marketing transaction data; analysis and finding critical hidden information has been a focused area for researchers of data mining. How to effectively analyze and apply these data and find the critical hidden information from these databases, data mining technique has been the most widely discussed and frequently applied tool from recent decades [2]. Although the data mining has been successfully applied in the areas of scientific analysis, business application, and medical research and its computational efficiency and accuracy are also improving, still manual works are required to complete the process of extraction.

Association rule mining model among data mining several models, including association rules, clustering and classification models, is the most widely applied method. The Apriori algorithm is the most representative algorithm for association rule mining [3]. It consists of many modified algorithms that focus on improving its efficiency and accuracy. However, two parameters, minimal support and confidence, are always determined by the decision-maker him/herself or through trial-and-error; and thus, the algorithm lacks both objectiveness and efficiency. Therefore, the main purpose of my study is to propose an improved algorithm that can provide feasible threshold values for minimal support and confidence.

For the purpose of simulation, I have employed the embedded database of Microsoft SQL Server 2000 to assess the proposed algorithm.

## II. LITERATURE REVIEW

*Association Rule Mining :*

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.

Agrawal et al. [4] first proposed the issue of the mining association rule in 1993. They pointed out that some hidden relationships exist between purchased items in transactional databases. Therefore, mining results can help decision-makers understand customers' purchasing behavior. An association rule is in the form of  $X \rightarrow Y$ , where X and Y represent Item set (I), or products, respectively and Item set includes all possible items  $\{i_1, i_2, \dots, i_m\}$ . The general transaction database  $(D = \{T_1, T_2, \dots, T_k\})$  can represent the possibility that a customer will buy product Y after buying product X .

Based on the concept of strong, rules, Agrawal et al., introduced association rules for discovering regularities between products in large scale transaction data recorded by point-of-sale (POS) systems in supermarkets. For example, the rule found in the sales

data of a supermarket would indicate that if a customer buys onions and potatoes together, he or she is likely to also buy burger. Such information can be used as the basis for decisions about marketing activities such as, e.g., promotional pricing or product placements. In addition to the above example from market basket analysis association rules are employed today in many application areas including Web usage mining, intrusion detection and bioinformatics.

### III. PARTICLE SWARM OPTIMIZATION

The concept of PSO was first suggested by Kennedy and Eberhart in 1995[5]. Particle swarm optimization (PSO) is inspired by the social behavior observed in flocks of birds and schools of fish. In nature, there is a leader who leads the bird or fish group to move, as illustrated in Fig. 1. Most members of the group follow the leader. In PSO, a potential solution to the considered problem is represented by a particle, similar to the individuals in the bird and fish group. Each particle travels in the solution space and attempts to move toward a better solution by changing its direction and speed based on its own past experience and the information from the current best particle of the swarm. [6]

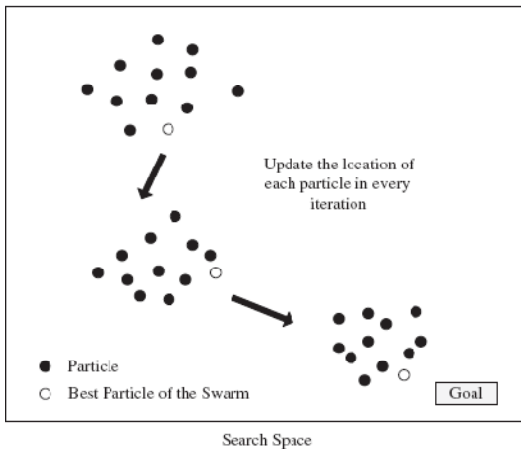


Fig. 1 : Swarm following the best particle to move to the goal

The procedure of PSO is described as follows:

#### A. Particle initialization:

An initial swarm of particles is generated in search space. Usually, the population size is decided by the dimension of problems.

#### B. Velocity and position update:

In each iteration, a new velocity value for each particle is calculated based on its current velocity, the distance from its previous best position, and the distance

from the global best position. The new velocity value is then used to calculate the next position of the particle in the search space. The particle's velocity and position are dynamically updated as follows:

$$V_{id}^{new} = w \times V_{id}^{old} + c_1 \times rand \times (P_{id} - x_{id}^{old}) + c_2 \times rand \times (P_{gd} - x_{id}^{old}),$$

$$x_{id}^{new} = x_{id}^{old} + V_{id}^{new}.$$

The new velocity of a particle,  $V_{new\_id}$ , is updated by taking into consideration of the particle's previous velocity,  $V_{old\_id}$ , and previous position,  $x_{old\_id}$ .  $w = [0.5 + rand/2]$  is an inertia weight and  $rand$  is a uniformly generated random number between 0 and 1. The cognition parameter,  $c_1$ , and social parameter,  $c_2$ , are acceleration coefficients that are conventionally set to a fixed value 0–2.  $P_{id}$  is the previous individual best position of this particle and  $P_{gd}$  is the current global best position then calculates the new position of the particle,  $x_{new\_id}$ . [6]

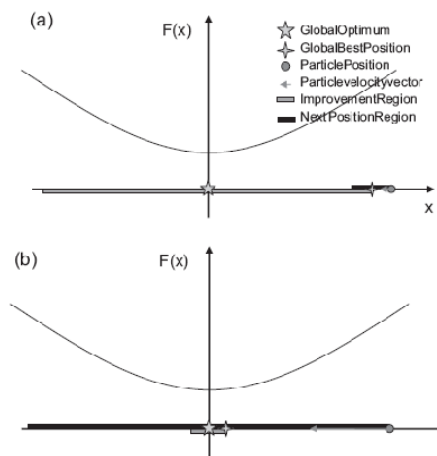


Fig. 2 : Two different situations in which a particle of PSO may fall during the search.

a) Both particle's position and the global best position are far from the optimum and the particle velocity is low compared to its distance to the optimum [7].

b) Global best position is close to the optimum and the particle position is far from them resulting in a small improvement region and a large next position region.

#### C. Evaluation and update of best locations:

The fitness value of each particle is calculated by the objective function. The values of  $P_{id}$  and  $P_{gd}$  are then evaluated and replaced if better particle best position or global best position is obtained.

*D. Termination:*

Steps (2) and (3) are repeated iteratively until the termination condition is met.

**IV. THE PROPOSED ASSOCIATION RULE MINING ALGORITHM**

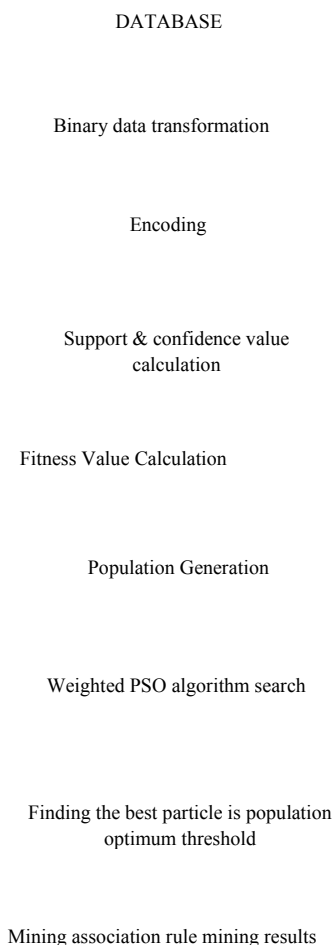


Fig. 3 : The proposed association rule mining algorithm

*The proposed algorithm:*

The proposed algorithm comprises two parts, preprocessing and mining. The first part provides procedures related to calculating the fitness values of the particle swarm. Thus, the data are transformed and stored in a binary format. In the second part of the algorithm, which is the main contribution of this study, the PSO algorithm is employed to mine the association

rules. First, we proceed with particle swarm encoding, this step is similar to chromosome encoding of genetic algorithms. The next step is to generate a population of particle swarms according to the calculated fitness value. Finally, the PSO searching procedure proceeds until the stop condition is reached, which means the best particle is found. The support and confidence of the best particle can represent the minimal support and minimal confidence. [1]

Thus, we can use this minimal support and minimal confidence for further association rule mining. Fig.3 illustrates the algorithm structure.

**A. Preprocessing of PSO association rule mining**

*Binary transformation :* In this we transform the transaction data into binary type data, each recorded and stored as either 0 or 1[8]. This approach can accelerate the database scanning operation, and it calculates support and confidence more easily and quickly. The transformation approach is explained by an example in Fig. 4.

In Fig. 4, there are five records, say T1 to T5, in the original data. Each of these records is transformed and stored as a binary type. For instance, there are a total of only four different products in the database, so four cells exist for each transaction. Take B4 as an example, this transaction only purchased products 2 and 3, so the values of cells 2 and 3 are both “1s,” whereas cells 1 and 4 are both “0s.”Conversion shown in figure 6.

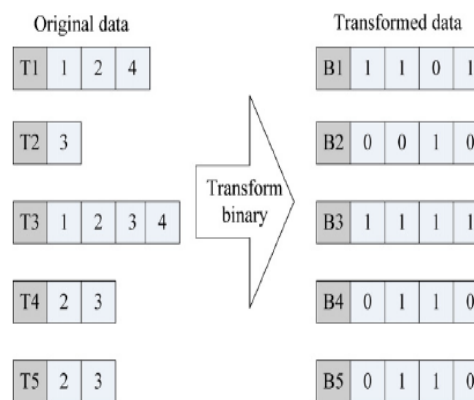


Fig. 4 : Data type transformation

**B. Application of PSO to association rule mining**

Applying PSO to association mining is the main part of this study. We use PSO as a module to mine best fitness value. The algorithmic process is quite similar to that of genetic algorithms, but the proposed procedures include only encoding, fitness value calculation, population generation, best particle search, and termination condition. Each of the steps in the PSO

algorithm and the process of generating association rules are explained as follows:

*Encoding* : According to the definition of association rule mining, the intersection of the association rule of item set X to item set Y ( $X \rightarrow Y$ ) must be empty. Items which appear in item set X do not appear in item set Y, and vice versa. Hence, both the front and back partition points must be given for the purpose of chromosome encoding. The item set before the front partition point is called “item set X,” while that between the front partition and back partition points is called “item set Y.”

*Fitness value calculation* : 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. Here, we employ the target function [9] to determine the fitness function value as shown in Eq.

$$Fitness(k) = confidence(k) \times \log(support(k) \times length(k) + 1)$$

Fitness (k) is the fitness value of association rule type k. Confidence (k) is the confidence of association rule type k.

Support (k) is the actual support of association rule type k.

Length (k) is the length of association rule type k.

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. Fitness value calculation is shown in figure 8.

*Population generation* : In order to apply the evolution process of the PSO algorithm, it is necessary to first generate the initial population. In this study, we select particles which have larger fitness values as the population. The particles in this population are called initial particles.

*Search the best particle* : First, the particle with the maximum fitness value in the population is selected as the “gbest.” We designed a method to constrain the search. The constrained method is to calculate the distance between the particle’s new position and all the possible particles inside the constrained range before the particle’s position is updated. Definitely, the particle with the smallest distance will be selected and treated as the particle’s new position. In the distance measuring function, we use traditional “Euclidean distance” as shown in Eq.

$$dist(x^n, y^m) = \sqrt{\sum_1^d (x_i^n - y_i^m)^2}$$

Where  $x_n$  is the position of the particle at nth update and  $y_m$  is the possible particle number min the constrained range. In addition, d is the dimension of the search space. The nearest possible particle is selected to be the target particle’s new position. This method can prevent a particle from falling beyond the search space when its position is updated.

*Termination condition* : To complete particle evolution, the design of a termination condition is necessary. In this study, the evolution terminates when the fitness values of all particles are the same. In other words, the positions of all particles are fixed. Another termination condition occurs after 100 iterations and the evolution of the particle swarm is completed.

Finally, after the best particle is found, its support and confidence are recommended as the value of minimal support and minimal confidence as shown in fig.10 & fig.11. These parameters are employed for association rule mining to extract valuable information [10].

## V. CONCLUSION

Implementation I have done till now has been encouraging and have depicted that the selection of Support and Confidence is done automatically and no manual processing is required. I have kept the population selection for the PSO as a user selection because by this user can test the details of the output on the basis of the results generated for the various population values. We can set its value to be a fixed value. 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.

From the graphs it is interpreted that the weighted PSO not only increases the support and confidence but also the run time taken is reduced considerably.

Further implementation of the Weighted PSO will be generating the results which will be analysed and compared with the results of the Non Weighted PSO.

## VI. MODEL EVALUATION RESULTS

This section will use the database provided by Microsoft SQLServer 2000 to verify the feasibility of the proposed algorithm.

### A. Experimental platform and database

This study’s experiment was conducted in the environment of Microsoft Windows XP using an IBM compatible computer with Intel Pentium IV 1.60GHz and 512MB RAM. The algorithm is coded using Java platform.



and discovery in knowledge-based databases), IEEE Transactions on Knowledge and Data Engineering 5 (6) (1993) 914–925 (Special issue).

- [4] R. Agrawal, T. Imielin' ski, A. Swami, "Mining association rules between sets of items in large databases", ACM SIGMOD Record 22 (2) (1993) 207–216.
- [5] J. Kennedy, and R. Eberhart, "Particle Swarm Optimization," Proceedings of the IEEE conference on neural networks – ICNN'95, vol. IV, Perth, Australia, 1995, pp. 1942-1948.
- [6] Chi-Yang Tsai, I-Wei Kao "Particle swarm optimization with selective particle regeneration for data clustering". Expert Systems with Applications 2010
- [7] Ahmad Nickabadi, Mohammad Mehdi Ebadzadeh, Reza Safabakhsh "A novel particle swarm optimization algorithm with adaptive inertia weight" Applied Soft Computing 2011
- [8] S.Y. Wur, Y. Leu, "An effective Boolean algorithm for mining association rules in large databases," in: Proceedings of the 6th International Conference on Database Systems for Advanced Applications, 1998, pp. 179–186.
- [9] S.H. Kung, "Applying Genetic Algorithm and Weight Item to Association Rule," Master Thesis, Department of Industrial Engineering and Management, YuanZe University, Taiwan, 2002.
- [10] G.jayalakshmi, Dr.K.Nageswara Rao "Mining association rules for large transactions using new support and confidence measures" Journal of Theoretical and Applied Information Technology 2005.

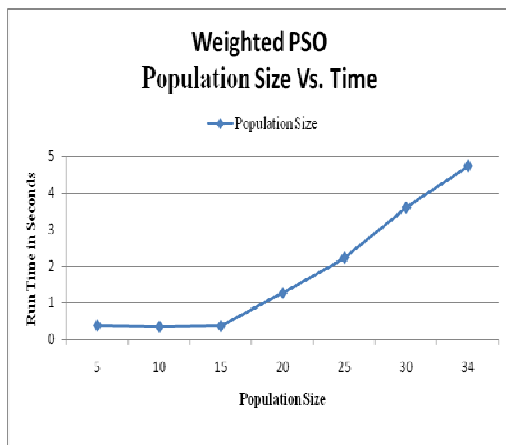


Fig. 9 : Weighted PSO Population Size Vs Time

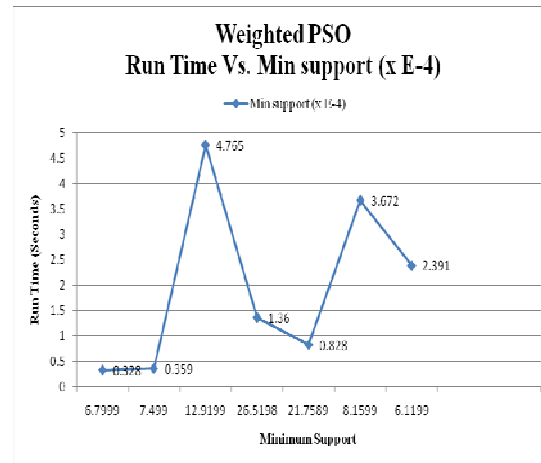


Fig. 10 : Weighted PSO Run Time Vs Minimum Support

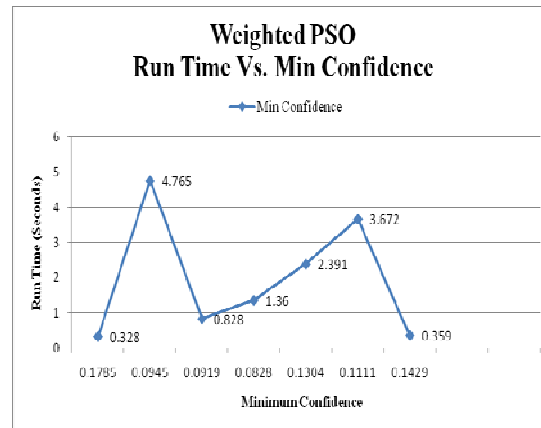


Fig. 11: Weighted PSO Run Time Vs Minimum Confidence

