

# Improved optimization of numerical association rule mining using hybrid particle swarm optimization and cauchy distribution

Imam Tahyudin<sup>1</sup> and Hidetaka Nambo<sup>2</sup>

<sup>1,2</sup>Artificial Intelligence Laboratory, Graduate School of Natural Science and Technology, Division of Electrical Engineering and Computer Science, Kanazawa University, Japan

<sup>1</sup>Department of Information System, STMIK AMIKOM Purwokerto, Indonesia

---

## Article Info

### Article history:

Received Sep 7, 2017

Revised Sep 10, 2018

Accepted Sep 16, 2018

### Keywords:

Numerical data

ARM

PSO

Cauchy distribution

Multi-objective functions

PARCD

---

## ABSTRACT

Particle Swarm Optimization (PSO) has been applied to solve optimization problems in various fields, such as Association Rule Mining (ARM) of numerical problems. However, PSO often becomes trapped in local optima. Consequently, the results do not represent the overall optimum solutions. To address this limitation, this study aims to combine PSO with the Cauchy distribution (PARCD), which is expected to increase the global optimal value of the expanded search space. Furthermore, this study uses multiple objective functions, i.e., support, confidence, comprehensibility, interestingness and amplitude. In addition, the proposed method was evaluated using benchmark datasets, such as the Quake, Basket ball, Body fat, Pollution, and Bolt datasets. Evaluation results were compared to the results obtained by previous studies. The results indicate that the overall values of the objective functions obtained using the proposed PARCD approach are satisfactory.

Copyright © 2019 Institute of Advanced Engineering and Science.

All rights reserved.

---

## Corresponding Author:

Imam Tahyudin,

Artificial Intelligence Laboratory,

Graduate School of Natural Science and Technology,

Electrical Engineering and Computer Science, Kanazawa University,

Kakumamachi, Kanazawa, Ishikawa, Japan.

Tel.: +81-76-234-4835

Fax: +81-76-234-4900

Email: imam@blitz.ec.t.kanazawa-u.ac.jp

---

## 1. INTRODUCTION

The ARM or association analysis method is used to find associations or relationships between variables, which often arise simultaneously in a dataset [1]. In other words, association analysis builds a rule for several variables in a dataset that can be distinguished as an antecedent or a consequent. The Apriori and Frequent Pattern (FP) growth methods are widely employed in association analysis. These methods are suitable for categorical or binary data, such as gender data, i.e., males can be represented by 0 and females by 1 [2]. Furthermore, if the data are numeric, such as age, weight or length, these methods process the data by transforming numerical data into categorical data (i.e., a discretization process). This transformation process requires more time and can miss a significant amount of important information because data transformation does not maintain the main meaning of the original data [3], [4], [5]. For example, if age data represents a 35 years old person and is transformed to 1, this obscures the original meaning of the age information. In addition, both methods require manual intervention to determine the minimum support (attribute coverage) and confidence (accuracy) values. Note that this step is subjective in some cases; thus, the results will not be optimal [6], [7].

To resolve this problem, some researchers have proposed solutions that employ optimization approaches, e.g., particle swarm optimization (PSO) [4], fuzzy logic [8], and genetic algorithm (GA) [3], [7]. Regarding of the PSO approach which has multiple objective functions for solving association analysis of numerical data without a discretization process. This research produced the better result than other previous optimization methods. It has optimum value automatically without determining the minimum support and minimum confidence. However, this method can also become trapped in local optima. When iterations are complete and the number of iterations tends toward infinity, the velocity value of a particle approaches 0 (the weight value of the velocity function is between 0 and 1). Therefore, the search is terminated because the PSO method can not find the optimal value when the velocity value is 0. Thus, PSO often fails to seek the overall optimal value [4], [9], [10].

We proposed a method that can address the premature searching and the limitations of traditional methods that it does not use a discretization process. In other word, the original data are processed directly using the concept of the Michigan or Pittsburgh approaches. Furthermore, support and confidence threshold values are determined automatically using the Pareto optimality concept. One solution to this problem is by combining PSO with the Cauchy distribution. This combination increases the size of the search space and is expected to produce a better optimal value. Yao et al (1999) reported that combining a function with the Cauchy distribution will result in a wider coverage area; thus, when the Cauchy distribution is combined with the function of the PSO method, the optimal value will increase [10].

Therefore, the purpose of this study is to find the optimal value of the numerical data in association analysis problems by combining PSO with the Cauchy distribution (PARCD). Furthermore, we determine the value of several objective functions such as support, confidence, comprehensibility, interestingness, and amplitude, as a parameter to evaluate the performance of the proposed method.

Problem solving in numerical data association analysis is generally performed using several approaches, including discretization, distribution and optimization. That the discretization is performed using partitioning and combining, clustering [11], [12] and fuzzy [8] methods, and the optimization approach is solved using the optimized association rule [13], differential evolution [14], GA [3], [7] and PSO [4], [15] as shown in Figure 1.

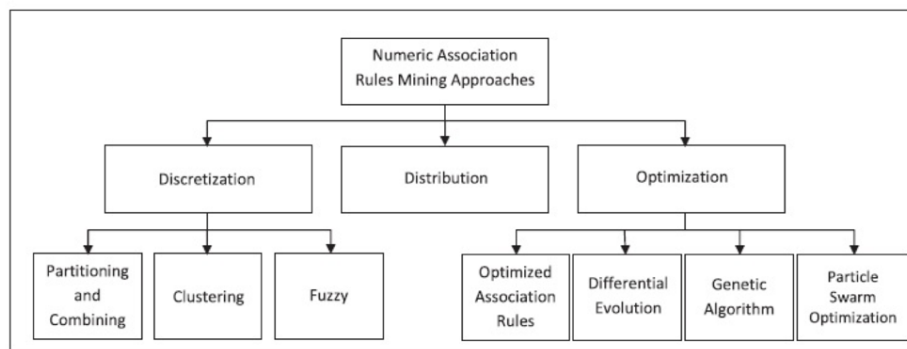


Figure 1. Numeric association analysis rule mining

We focus to solve the problem of association analysis of numerical data by optimization. The previous research from optimization approach is known as the GAR method. It has been attempted to find the optimal item set with the best support value without using a discretization process [13]. And then, the differential evolution optimization approach includes the generation of the initial population, as well as mutation, crossover and selection operations. The multi-objective functions are optimized using the Pareto optimality theory. This method is known as MODENAR [14]. Furthermore, a study of numerical association rule mining using the genetic algorithm approach (ARMGA). It successfully solved association analysis of numerical data problems without determining the values of the minimum support or minimum confidence manually. In addition, this method can extract the best rule that has the best relationship between the support and confidence values [7]. Another study of GA approach has been used MOGAR method. It presented that using MOGAR method was faster than using conventional methods, such as Apriori and FP-growth algorithms, because the time complexity of the MOGAR method tends to be simpler, and follows quadratic distribution. On the other hand, the Apriori

algorithm follows an exponential distribution, which requires more time for computation [3].

Next, the optimization method has been used PSO for solving numerical ARM problem. Some authors who performed PSO method such as they used ARM to investigate the association of frequent and repeated dysfunction in the production process. The result obtained a faster and more effective optimization employed PSO, which resulted in a faster and more effective optimization process than the other optimization methods [16]. In addition, the PSO approach was used to improved the computational efficiency of ARM problems such that appropriate support and confidence values could be determined automatically [17]. In 2012, the development of PSO for ARM problems was performed by weighting the item set. This weighting is very important for very large data because such data often contain important information that appears infrequently. For example, in medical data, if there is a rule {stiff neck, fever, aversion to light}  $\rightarrow$  {meningitis} that rarely appears but this rule is very important because in fact this condition is often happen [18]. In 2013, Sarath and Ravi introduced binary PSO (BPSO) to generate association rules in a transaction database. This method is similar to the Apriori and FP growth algorithms; however, BPSO can determine optimum rules without specifying the minimum support and confidence values [19]. In 2014, Beiranvand et al. studied numerical data association analysis using the PSO method. They stated that the employed method could effectively analyze numerical data association analysis problems without using a discretization process. This research employs four objective functions, i.e., support, confidence, comprehensibility and interestingness. This method is referred to as MOPAR [4]. In 2014, Indira and Kanmani conducted research using a PSO approach; however, they attempted to improve results and analysis time using an adaptive parameter determination process to determine various parameters, such the constant and weight value in a velocity equation. They developed the Apriori algorithm using a PSO approach (APSO), and the results demonstrated that this approach was faster and better compared to using only an Apriori method [15]. In addition, the combination of PSO and GSA has been conducted for solving optimal reactive power dispatch problem in power system. The problem has successfully accomplished on basis of efficient and reliable technique. And then, the result were found satisfactorily to a large extent that of reported earlier [20]. Verma and Lakhwani examined ARM problems by combining PSO and a GA. The results showed better accuracy and consistency compared to individual PSO or a GA method [21].

There are many developments of PSO method. i.e. the papers; "the implementation of PSO in distributed generation sizing" [22], "improved canny edges using cellular based PSO technique in digital images" [23], and the hybrid method. One of hybrid methods is the hybrid PSO with the Cauchy distribution [24]. This method provides better results compared to using only PSO. In 2011, this combined method was retested for SVM parameter selection [25-27]. The combined approach was also used to improve performance weaknesses in a process to identify a watermark image based on discrete cosine transform (DCT). The results demonstrated that combining PSO with the Cauchy distribution outperforms the compared method [28]. In 2014, an empirical study demonstrated that combining PSO with the Cauchy distribution provided. The results show that the use of PSO with Cauchy distribution higher than using only PSO [29].

To the best of our knowledge, combining PSO with the Cauchy distribution has not been applied to ARM problems that involve numerical data. This research has important contribution for optimization approach of numerical ARM problem.

The reminder of this paper is organized as follows. Research method is discussed in Section 2. This section describes the design of the multiple objective functions and the development of the proposed PARCD method. Section 3 exposes the experimental result and discussion of proposed method which was tested using a dataset benchmark. This section also provides a comparison of the results obtained by the proposed PARCD method and existing methods. Conclusions and suggestions for future work are provided in Section 4.

## 2. RESEARCH METHOD

### 2.1. Objective Design

This study uses multiple objective functions, i.e., support, confidence, comprehensibility, interestingness and amplitude. First, the support criterion determines the ratio of transactions for item X to the total transaction (D), i.e.,  $\text{support}(X) = |X| / |D|$ . Then, if A is the antecedent of the transaction dataset as a precondition then C is consequence as the conclusion of a transaction dataset. The support value of if A then C ( $A \rightarrow C$ ) is computed as follows:

$$\text{Support}(A \cup C) = \frac{|A \cup C|}{|D|} \quad (1)$$

where  $|A \cup C|$  is the number of transaction which contain A and C.

The minimum support value is closely linked to the number of items covered to determine the referenced rule. If the threshold value is low, the support covers many items and vice versa. The support measurement is used to determine the confidence measurement criteria, i.e., the criteria used to measure the quality or accuracy of the rule derived from the total transactions. Such rules are often developed for each transaction to better demonstrate quality or accuracy [4]. Confidence can be expressed as follows,

$$Confidence(A \cup C) = \frac{Support(A \cup C)}{Support(A)} \quad (2)$$

However, these criteria are not guaranteed to produce appropriate rules. Thus, for a given rule to be considered reliable and to provide overall coverage, the result must also satisfy the comprehensibility and interestingness criteria. Gosh and Nath (2004), stated that less number of attributes in antecedent component of a rule show that the rule is comprehensible [30]. The comprehensibility measurement criteria can be expressed as follows:

$$Comprehensibility(A \cup C) = \frac{\log(1 + |C|)}{\log(1 + |A \cup C|)} \quad (3)$$

where  $|C|$  is the number of consequence item and  $|A \cup C|$  is the rule number of if A then C ( $A \rightarrow C$ ).

Next, the interestingness criteria are used to generate hidden information by extracting some interesting rule or unique rule. This criterion is based on the support value and is expressed as follows:

$$Interestingness(A \cup C) = \left[ \frac{Supp(A \cup C)}{Supp(A)} \right] \left[ \frac{Supp(A \cup C)}{Supp(C)} \right] \left[ 1 - \frac{Supp(A \cup C)}{|D|} \right] \quad (4)$$

The right side of Eq. (4) consists of three components. The first component shows the generation probability of the rule that is based on the antecedent attribute. The second is based on the consequence attributes and the third is based on the total dataset. There is a negative correlation between interestingness and support. When the support value is high, the interestingness value is low because the number of frequent items covered is small [4].

The last criterion is the amplitude interval. The amplitude interval, which is a measure of a minimization function, differs from support, confidence and comprehensibility measures, which are maximization functions. The amplitude interval is expressed as follows:

$$Amplitude(A \cup C) = 1 - \frac{1}{m} \sum_{i=1}^m \left[ \frac{u_i - l_i}{\max(A_i) - \min(A_i)} \right] \quad (5)$$

Here, m is the number of attributes in the item set ( $|A \cup C|$ ),  $u_i$  and  $l_i$  are the upper and lower bounds encoded in the item sets corresponding to attribute i.  $\max(A_i)$  and  $\min(A_i)$  are the allowable limits of the intervals corresponding to attribute i. Thus, rules with smaller intervals are intended to be generated [14].

## 2.2. PSO

PSO, which was first introduced by Kennedy and Eberhart (1995), is an evolutionary method inspired by animal behavior, e.g., flocks of birds, school of fish, or swarms of bees [31]. PSO begins with a set of random particles. Then, a search process attempts to find the optimal value by performing an update generation process. During each iteration, each particle is updated by following two best values. The first is the best solution (fitness) achieved to this point. This value is called pBest. The other best value tracked by the swarm particle optimizer is the best value obtained by each particle in the population. The value is called gBest. After finding pBest and gBest, each particle's velocity and corresponding position are updated [15].

Each particle p in some iteration t has a position  $x(t)$  and displacement speed  $v(t)$ . The finest particles (pBest) and best global positioning (gBest) are stored in memory. The speed and position are updated using Eqs. 6 and 7, respectively [15].

$$V_{i,new} = \omega V_{i,old} + C1rand()(pBest - X_i) + C2rand()(gBest - X_i) \quad (6)$$

$$X_{i, new} = X_{i, old} + V_{i, new} \quad (7)$$

Here  $\omega$  is the inertia weight;  $V_{i, old}$  is the velocity of the  $i$ -th particle before updating;  $V_{i, new}$  is the velocity of the  $i$ th particle after updating;  $X_i$  is the  $i$ -th, or current particle;  $i$  is the number of particles;  $rand()$  is a random number in the range (0, 1);  $C1$  is the cognitive component;  $C2$  is the social component;  $pBest$  is the particle best or local optima in some iterations on every running;  $gBest$  is the global best or global optima in some iterations on every running. Particle velocities in each dimension are restricted to maximum velocity  $V_{max}$  [32].

### 2.3. Cauchy Distribution

Yao et al. (1999) used a Cauchy distribution to implement a wider mutation scale [10]. A general formula for the probability density function is expressed as follows.

$$f(x) = \frac{1}{s\pi(1 + ((x - t)/s)^2)} \quad (8)$$

A Cauchy random variable is calculated as follows. For any random variable  $X$  with distribution function  $F$ . The random variable  $Y=F(X)$  has a uniform distribution in the range [0,1]. Consequently, if  $F$  is inverted, the random variable can use a uniform density to simulate random variable  $X$  because  $X = F^{-1}(Y)$ . Therefore, the cumulative distribution function of Cauchy distribution is expressed as follows

$$F(x) = \frac{1}{\pi} \arctan(x) + 0.5 \quad (9)$$

Therefore if

$$y = \frac{1}{\pi} \arctan(x) + 0.5 \quad (10)$$

by inverting its function, the Cauchy random variable can be expressed as follows

$$x = \tan(\pi(y - 0.5)) \quad (11)$$

This function can be expressed by Eq. (12) because  $y$  has a uniform distribution in the range (0,1]. Thus, we obtain the following,

$$x = \tan(\pi/2 \cdot rand[0, 1]) \quad (12)$$

### 2.4. PSO for Numerical Association Rule Mining with Cauchy Distribution

PARCD is an extension of the MOPAR methods that combines PSO and the Cauchy distribution to solve problems that occur in the association analysis of numerical data [33]. The goal is to find the optimal value of amateurs and avoid being trapped in local optima. Essentially, this method uses the concept of PSO but modifies the velocity equation by including the Cauchy distribution. The velocity function is expressed as follows,

$$V_i(t+1) = \omega(t)V_i(t) + C1rand()(pBest - X_i(t)) + C2rand()(gBest - X_i(t)) \quad (13)$$

The next step is normalization by using  $V_i(t+1)$  value (Eq. 13), which makes the vector length 1. The variant of the Cauchy distribution is infinite and the objective function scales are 1 [10].

$$U_i(t+1) = \frac{V_i(t+1)}{\sqrt{V_{i1}(t+1)^2 + V_{i2}(t+1)^2 \dots + V_{iK}(t+1)^2}} \quad (14)$$

The result of the normalization process is multiplied by the Cauchy random variable as follows.

$$S_i(t+1) = U_i(t+1) \cdot \tan\left(\frac{\pi}{2} \cdot rand[0, 1]\right) \quad (15)$$

Then, the result of Eq. (15) which is a combination of the velocity value and the Cauchy distribution, is used to determine the new position of a particle.

$$Xi(t + 1) = Xi(t) + Si(t + 1) \quad (16)$$

## 2.5. PARCD Pseudo code and Flowchart

The PARCD pseudocode as shown in Figure 2 and flowchart as shown in Figure 3 show that the algorithm begins by initializing the velocity vector and position randomly. The algorithm calculates the multi-objective functions as the current fitness. Then, it executes looping iterations to seek pBest until it finds the gBest value as the optimal solution.

---

Pseudocode of PARCD:

---

```

Procedure: Combination of PSO and Cauchy distribution for Numerical ARM
Input: PARCD parameters
Output: multi objective function value
1. begin
   t ← 0
2. initialize  $X_i(t)$  by encoding routine;
3. calculate the multi objectives for all particles // current fitness
4. evaluate  $X_i(t)$  by decoding routine and keep the best solution;
5. while (not terminating condition) do
6.   for each particle  $X_i$  in swarm do
7.     update velocity  $V_i(t+1)$  // using (13)
8.     update position  $X_i(t+1)$  //using (16)
9.     calculate  $U_i(t+1)$  and  $S_i(t+1)$  //By Cauchy distribution using (14) and (15)
10.    evaluate  $X_i(t+1)$  //using (16)
11.    if  $f(X_i(t+1)) < f(pBest_i(t))$  then //pBest(t) : historical best position
12.      update  $pBest_i(t) = X_i(t+1)$ ; //update the best local position
13.    end;
14.     $gBest(t+1) = \text{argmin}\{f(pBest_k(t)), gBest(t)\}$  //update the global best position
15.    t ← t+1;
16.  end;
17.  out put: the best solution gBest;
18. end;

```

---

Figure 2. PARCD pseudocode

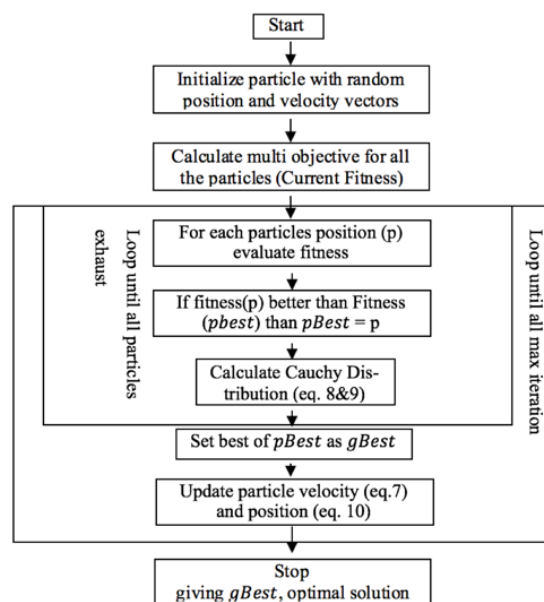


Figure 3. PSO flowchart

### 3. RESULT AND DISCUSSION

#### 3.1. Experimental Setup

We conducted an experiment using the Quake, Basketball, Body fat, Pollution, and Bolt benchmark datasets in Table 1. from the Bilkent University Function Approximation Repository. The experiment was performed using a computer with an Intel Core i5 processor with 8 GB main memory running Windows 7. The algorithms were implemented using MATLAB.

For the proposed algorithm, we set the population size, external repository size, number of iterations, C1 and C2,  $\omega$ , velocity limit and xRank parameters in Table 2. to 40, 100, 2000, 2, 0.63, 3.83, and 13.33 respectively.

Table 1. Dataset Properties

Dataset	No. of Records	No. of Attributes
Quake	2178	4
Basketball	96	5
Body fat	252	15
Pollution	60	16
Bolt	40	8

Table 2. Parameters

Parameter	Size	External Repository Size	Number of iteration	C1,C2	$\omega$	Velocity Limit	xRank
Average	40	100	2000	2	0.63	3.83	13.33

#### 3.2. Experiments

Association rule analysis comprises two steps. The first step is to determine the frequent itemset that includes the antecedents or consequences of each attribute. The second step is to implement the proposed algorithm.

##### 3.2.1. Output Rules of the PARCD Results

This experiment shows the 20<sup>th</sup> run time where each running contains 2000 rules. We presented three datasets of output rules i.e. Body fat, Bolt, and Pollution datasets. Table 3 shows the results obtained with the Body fat dataset. For Rule 1, there are eight antecedent attributes and three consequent attributes. For Rule 2, the number of antecedent and consequent attributes are the same as Rule 1. For the last rule, the number of antecedent and consequent attributes are six and two, respectively.

The antecedent attributes of Rule 1 are case number, percent body fat (Siri's equation), density, age, adiposity index, chest circumference, abdomen circumference, and thigh circumference. The consequent attributes are percent body fat (Brozek's equation), height, and hip circumference. For Rule 2, the antecedent and consequent attributes are the same as Rule 1. Thus, Rules 1 and 2 can be expressed as follows: if (att1, att3, att4, att5, att8, att11, att12, att14) then (att2, att7, att13). For Rule 2000, the antecedent attributes are Percent body fat using Brozek's equation, Percent body fat using Siri's equation, density, height, neck circumference and knee circumference, and the consequent attributes are case number and weight. Therefore, Rule 2000 is if (att2, att3, att4, att7, att10, att15) then (att1, att6).

Table 4 shows the results obtained with the Bolt dataset, which has eight attributes; (run, speed, total, speed2, number2, Sens, time and T20Bolt). As can be seen, the first two rules the same results for both antecedent and consequent attributes. The antecedent attributes are total and time, and the consequent attributes are run and speed1. Therefore, the rule is if (total, time) then (run, speed1). The rule 2000 shows that the antecedent

attributes are run and speed2. However, the consequent attribute is unknown. Thus, this rule cannot be declared clearly because it does not have a conclusion.

Table 5 shows the rule results for the pollution dataset obtained using the proposed particle representation PARCD method. The results for the first and second rules are the same. Here, the antecedent attributes are JANT, EDUC, NONW, and WWDRK, and the consequent attributes are PREC, JULT, OVR65, DENS and HUMID. Thus, the rule is if (JANT, EDUC, NONW, WWDRK) then (PREC, JULT, OVR65, DENS, HUMID).

The Rule 2000 has an ACN result that differs from the first and second attributes. The antecedent attributes of Rule 2000 are JANT, OVR65, HOUS, POOR, HC and HUMID and its consequent attributes are POPN, EDUC, DENS, NOX, and SO@. Thus, the final rule is if (JANT, OVR65, HOUS, POOR, HC) then (POPN, EDUC, DENS, NOX, SO@).

Table 3. ACN Rules (the Body fat dataset)

Rules	ACN	LB < Attribute < UB
Rule 1	Antecedent	1.096724 < Att1 < 1.108900 57.988435 < Att3 < 69.574945 309.987803 < Att4 < 314.218245 55.294719 < Att5 < 66.896106 136.234441 < Att8 < 138.744999 40.927433 < Att11 < 41.562953 20.266071 < Att12 < 20.586850 22.220988 < Att14 < 23.180185
	Consequence	35.426088 < Att2 < 42.169776 113.825926 < Att7 < 122.261793 32.375620 < Att13 < 33.596051
Rule 2	Antecedent	1.096724 < Att1 < 1.108900 57.988435 < Att3 < 69.574945 309.987803 < Att4 < 314.218245 55.294719 < Att5 < 66.896106 136.234441 < Att8 < 138.744999 40.927433 < Att11 < 41.562953 20.266071 < Att12 < 20.586850 22.220988 < Att14 < 23.180185
	Consequence	35.426088 < Att2 < 42.169776 113.825926 < Att7 < 122.261793 32.375620 < Att13 < 33.596051
.....		
.....		
Rule 2000	Antecedent	12.402089 < Att2 < 18.144187 56.221481 < Att3 < 65.667791 139.024098 < Att4 < 289.982951 94.156397 < Att7 < 136.200000 57.669974 < Att10 < 87.300000 18.798957 < Att15 < 19.060978
	Consequence	1.054478 < Att1 < 1.108900 31.100000 < Att15 < 40.883823
Note :		
	Att1	: Case Number
	Att2	:Percentage using Brozek's equation
	Att3	:Percentage using Siri's equation
	Att4	:Density
	Att5	:Age (years)
	Att6	:Weight (lbs)
	Att7	:Height (inches)(target)
	Att8	:Adiposity index
	Att9	:Fat Free Weight
	Att10	:Neck circumference (cm)
	Att11	:Chest circumference (cm)
	Att12	:Abdomen circumference (cm)
	Att13	:Hip circumference (cm)
	Att14	:Thigh circumference (cm)
	Att15	:Knee circumference (cm)
	Att16	:Ankle circumference (cm)
	Att17	:Extended biceps circumference (cm)
	Att18	:Forearm circumference (cm)
	Att19	:Wrist circumference (cm)

Table 4. ACN Rules (the Bolt dataset)

Rules	ACN	LB < Attribute < UB
Rule 1	Antecedent	11.911616 < Att3 < 16.259242 62.782669 < Att7 < 65.562550
	Consequence	23.688468 < Att1 < 31.295955 5.928943 < Att2 < 6.000000
Rule 2	Antecedent	11.911616 < Att3 < 16.259242 62.782669 < Att7 < 65.562550
	Consequence	23.688468 < Att1 < 31.295955 5.928943 < Att2 < 6.000000
.....		
.....		
Rule 2000	Antecedent	13.621221 < Att1 < 29.817232 1.761097 < Att4 < 2.325029
	Consequence	None
Note :		
	Att1	:RUN
	Att2	:SPEED1
	Att3	:TOTAL
	Att4	:SPEED2
	Att5	:NUMBER2
	Att6	:SENS
	Att7	:TIME
	Att8	:T2OBOLT



Table 5. ACN Rules (the Pollution dataset)

Rules	ACN	LB < Attribute < UB
Rule 1	Antecedent	42.431841 < Att2 < 46.441110 9.675301 < Att6 < 10.303791 24.171326 < Att9 < 27.345700 42.882070 < Att10 < 44.054696
	Consequence	21.695266 < Att1 < 22.757671 77.760994 < Att3 < 80.221960 6.698662 < Att4 < 7.071898 7436.549761 < Att8 < 7801.004046 58.816363 < Att15 < 63.240005
Rule 2	Antecedent	42.431841 < Att2 < 46.441110 9.675301 < Att6 < 10.303791 24.171326 < Att9 < 27.345700 42.882070 < Att10 < 44.054696
	Consequence	21.695266 < Att1 < 22.757671 77.760994 < Att3 < 80.221960 6.698662 < Att4 < 7.071898 7436.549761 < Att8 < 7801.004046 58.816363 < Att15 < 63.240005
.....		
.....		
Rule 2000	Antecedent	39.363260 < Att2 < 46.455909 8.721294 < Att4 < 9.206407 89.212389 < Att7 < 90.700000 21.796671 < Att11 < 23.231486 606.938956 < Att12 < 648.000000 67.768113 < Att15 < 73.000000
	Consequence	2.956662 < Att5 < 3.005372 9.450171 < Att6 < 10.068287 9345.537477 < Att8 < 9699.000000 225.061313 < Att13 < 288.274133 242.720468 < Att14 < 250.733264
Note :	Att1	:PREC Average annual precipitation in inches
	Att2	:JANT Average January temperature in degrees F
	Att3	:JULT Average July temperature in degrees F
	Att4	:OVR65 SMSA population aged 65 or older
	Att5	:POPN Average household size
	Att6	:EDUC Median school years completed by those over 22
	Att7	:HOUS of housing units which are sound and with all facilities
	Att8	:DENS Population per sq. mile in urbanized areas, 1960
	Att9	:NONW non-white population in urbanized areas, 1960
	Att10	:WDRK employed in white collar occupations
	Att11	:POOR poor of families with income ; USD3000
	Att12	:HC Relative hydrocarbon pollution potential
	Att13	:NOX Same as nitric oxides
	Att14	:SO@ Same as Sulphur dioxide
	Att15	:HUMID Annual average, relative humidity at 1 pm
	Att16	:MORT Total age-adjusted mortality rate per 100,000

### 3.2.2. Output of multi-objective function and correlation of PARCD methods

The basic concept of association analysis comprises two steps, i.e., the first step is the determination rules which in every rule contain antecedent and consequent and the second step is the implementation of the algorithm (i.e., the proposed method). This method begins with the initialization process, which as the start of the algorithm starts with the determine the multi-objective function value and calculates the particle velocity and positioning at  $i$ . Then, an iterative process is performed to search for pBest and gBest as the optimal solution.

Table 6 shows the results of the multi-objective function of the PARCD method. Here, there are four parameters i.e., support, confidence, comprehensibility and interestingness. Then, the method is examined using five datasets i.e., quake, basketball, body fat, bolt, and pollution. Generally, the Bolt dataset is the dominant data set and has the highest value for each parameter (except comprehensibility). Conversely, the least dominant dataset is quake (with the exception of the confidence parameter).

Table 6. The Output of PARCD Method

Dataset	Support (%)	Confidence (%)	Comprehensibility	Interestingness (%)
Quakes	22.97	86.73 ± 25.88	785.2 ± 37.72	2.34 ± 9.30
Basket Ball	61.04	92.69 ± 17.87	545.80 ± 167.74	6.56 ± 21.16
Body fat	73.94	81.26 ± 30.67	333.49 ± 218.95	10.61 ± 21.03
Pollution	250.84	96.88 ± 9.49	231.08 ± 168.35	43.43 ± 39.68
Bolt	60.45	34.96 ± 43.91	110.63 ± 165.76	9.51 ± 18.61

The first parameter, i.e., support, showed a higher value with the Bolt dataset (250.84%) and the lowest with the quake dataset (22.97%). The average was approximately 90%. The highest confidence value was similar to the support value. The highest confidence value was obtained with the Bolt dataset (96.88%) with a deviation of approximately 10. The lowest confidence value was obtained with the pollution dataset (34.96%) with a very high deviation of just under 45. The average confidence value was approximately 80%. The highest comprehensibility value was obtained with the Quake dataset (approximately 785). The lowest comprehensibility value was obtained with the pollution dataset (approximately 110 with a deviation, well over 165). The average comprehensibility value was approximately 400. The final parameter, i.e., interestingness, obtained the highest value with the bolt dataset (approximately 43% with a deviation of just under 40). The lowest interestingness value was obtained with the quake dataset (2.34% with a deviation of just under 10). The average interestingness value was approximately 15%. This demonstrates that the support and confidence values, i.e., 90% and 80% respectively, were satisfactory. Moreover, the comprehensibility value was four times better; however, the interestingness value was not satisfactory (approximately 15%).

The correlation values between each objective function are shown in Table 7 and Figure 4. The results show one objective function with another are significant association either be positive or negative. The correlation value of all objective functions to amplitude was always close to zero. In other words, the correlation to the amplitude function was low. This proves the opinion given by Alatas et al. (2008), i.e., the amplitude function differs from other functions because it attempts to minimize while the other functions attempt to maximize their values.

Table 7. Correlation of Multi-Objective Function

		Support	Confidence	Comprehensibility	Interestingness	Amplitude
Quake	Support	1	<b>0.8076</b>	0.2112	<b>0.9999</b>	0.0000
	confidence	0.8076	1	0.3971	<b>0.8077</b>	0.0000
	comprehensibility	0.2112	0.3971	1	0.2113	0.0000
	interestingness	0.9999	0.8077	0.2113	1	0.0000
	amplitude	0.0000	0.0000	0.0000	0.0000	1
Basket ball	Support	1	0.4360	<b>-0.7437</b>	<b>0.9750</b>	0.0000
	confidence	0.4360	1	0.1646	0.5716	0.0000
	comprehensibility	-0.7437	0.1646	1	<b>-0.6350</b>	0.0000
	interestingness	0.9750	0.5716	-0.6350	1	0.0000
	amplitude	0.0000	0.0000	0.0000	0.0000	1
Body fat	Support	1	<b>0.8137</b>	<b>-0.8340</b>	<b>0.8555</b>	0.0000
	confidence	0.8137	1	<b>0.9917</b>	<b>0.9469</b>	0.0000
	comprehensibility	0.8340	0.9917	1	<b>0.9575</b>	0.0000
	interestingness	0.8555	0.9469	0.9575	1	0.0000
	amplitude	0.0000	0.0000	0.0000	0.0000	1

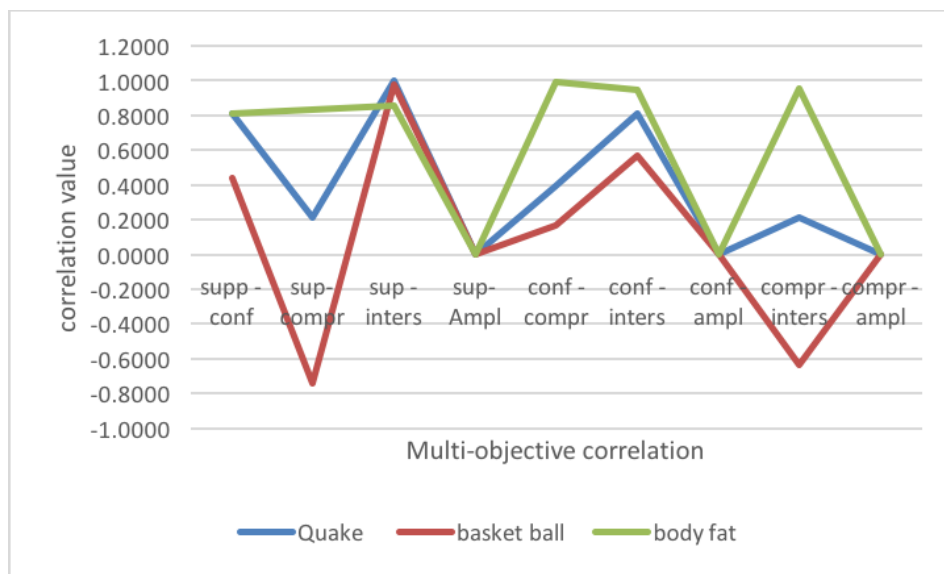


Figure 4. The correlation of objectives

### 3.2.3. The comparison of multiobjective function between PARCD and other methods

Table 8 shows a comparison of the support value obtained by the proposed PARCD method and five previous methods (i.e., the MOPAR, MODENAR, GAR, MOGAR, and RPSOA methods). Generally, the support percentage obtained by the PARCD method was better than that obtained by the other methods. The support value obtained by the PARCD method with the Quake dataset was the lowest (22.97%). The highest value was obtained by the MOPAR method (46.26%). The support value of the remaining methods was just over 35% on average. The support values obtained with the basketball and body fat dataset were the highest, i.e., 61.04% and 73.94%, respectively. The second highest support value was obtained by the MOGAR method with the basketball dataset (50.82%). The average support value of all other methods was well over 35%. The lowest support value for the body fat data set is MOPAR method (22.95%), and the average value was approximately 65%.

The comparison of number of rules and confidence values are shown in table 9. The proposed PARCD method demonstrates a nearly similar number of rules compared to other methods. The greatest number of rules obtained with the quake dataset was achieved by the MODENAR method (55 rules). The PARCD method obtained the greatest number of rules with the basketball (78 rules); however, with the body fat dataset, the PARCD method obtained the lowest number of rules (32). The MOGAR method obtained the greatest number of rules with the basketball dataset. The confidence values obtained by the PARCD, MOPAR, and MOGAR methods were approximately the same (just over 80%). Generally, the MOPAR method showed the highest confidence value with all datasets, with the exception of the body fat dataset, with which the MOGAR method obtained the highest confidence value. Then, the second position is PARCD method.

Tables 8 and 9 show that the support and confidence values were correlated with the number of rules, i.e., significant negative correlation was observed. Note that, if the support and confidence values were high, then the number of rules was low (and vice versa). This condition occurs because the high support and confidence values effectively filter the number of rules selectively.

Table 10 shows the size value and amplitude percentage obtained by the proposed PARCD and existing methods. Generally, the size value of the body fat dataset was the highest with all methods, e.g., the GAR method obtained a size value of approximately 7.5. On the other hand, the size value of the Quake dataset with the MODENAR method was the lowest. The PARCD method obtained the best amplitude value with the Basketball dataset (approximately 2%), while the opposite value is also using PARCD method which Quake dataset gain around 65%. The amplitude value obtained by the MOPAR method was fairly good. The amplitude value obtained by the MOPAR method with the Body fat dataset was approximately 4%, and that obtained by the

MOPAR with the quake dataset result was less than that obtained by the PARCD method, which was just over 50%. In addition, the MODENAR, MOGAR, and GAR methods outperformed both the PARCD and MOPAR methods. Their amplitude results were approximately 17% to 29% for all dataset.

Table 8. The Comparison of Support Value

Dataset	Support (%)					
	PARCD	MOPAR	MODENAR	GAR	MOGAR	RPSOA
Quake	22.97	31.97	39.86	38.65	30.12	38.74
Basketball	61.04	30.76	37.20	36.69	50.82	36.44
Body fat	73.94	22.95	65.22	65.26	57.22	65.22

Table 9. The Comparison of Number of Rules and Confidence Values

Dataset	Number of Rules (%)				
	PARCD	MOPAR	MODENAR	MOGAR	RPSOA
Quake	51	54.1	55	45	46
Basketball	78	69.75	48	50	34
Body fat	32	70.8	52	84	46

Dataset	Confidence (%)				
	PARCD	MOPAR	MODENAR	MOGAR	RPSOA
Quake	86.73 ± 25.88	89.32 ± 0.18	63 ± 2.8	82	63 ± 2.8
Basketball	92.69 ± 17.87	95 ± 0.12	61 ± 2.1	83	60 ± 2.8
Body fat	81.26 ± 30.67	81.8 ± 0.27	62 ± 3.2	85	61 ± 1.8

Table 10. The Comparison of Size and Amplitude Value

Dataset	Size (%)				
	PARCD	MOPAR	MODENAR	GAR	RPSOA
Quake	2.28	2.20	2.03	2.33	2.22
Basketball	2.14	2.55	3.21	3.38	3.21
Body fat	5.79	2.26	6.87	7.45	6.94

Dataset	Amplitude (%)				
	PARCD	MOPAR	MODENAR	GAR	RPSOA
Quake	64.71	53.85	17	25	17
Basketball	2.14	30.0	19	25	19
Body fat	10.80	3.61	25	29	25

The overall results indicate that proposed PARCD method can reach wider compared to the existing methods when searching for an optimal value. These results also indicate the proposed method may be robust for problems in others fields, such as the numerical association rule mining optimization problem.

#### 4. CONCLUSION

This study has proved that combining the PSO with Cauchy distribution can solve the numerical ARM problem. The problems of local minimum and premature convergence with large datasets can be solved using the proposed method. The experimental results demonstrate that the proposed PARCD method outperforms existing methods (i.e., MOPAR, MODENAR, GAR, and RPSOA) relative to all multi-objective functions, such as the support, confidence, comprehensibility, interestingness and amplitude functions. In future, the numerical problem of ARM problem can be further improved by developing or combining other methods, such as time series or deep learning method.

#### ACKNOWLEDGEMENT

This research was supported by various parties. We would like to thank the scholarship program of KU - DIKTI (Kanazawa University, Japan and Directorate of Higher Education (DIKTI), Indonesia). We also thank to STMIK AMIKOM Purwokerto, Indonesia. We express our gratitude to the anonymous reviewers who provided thoughtful input and suggestions to improve this paper.

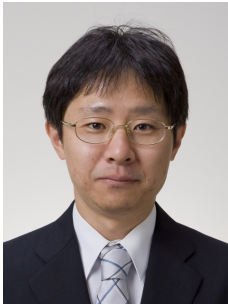
#### REFERENCES

- [1] H. Jiawei, K. Micheline, and P. Jian, *DATA MINING (Concept and Techniques)*, vol. 3, no. 13. 2012.
- [2] I. H. Witten, F. Eibe, and H. Mark A., *Data Mining (Practical Machine Learning Tools and Techniques)*, vol. 3, no. 9. Elsevier, 2011.
- [3] B. Minaei-Bidgoli, R. Barmaki, and M. Nasiri, Mining numerical association rules via multi-objective genetic algorithms, *Inf. Sci. (Ny)*, vol. 233, pp. 1524, Jun. 2013.
- [4] V. Beiranvand, M. Mobasher-Kashani, and A. Abu Bakar, Multi-objective PSO algorithm for mining numerical association rules without a priori discretization, *Expert Syst. Appl.*, vol. 41, no. 9, pp. 42594273, Jul. 2014.
- [5] V. Pachón Álvarez and J. Mata Vázquez, An evolutionary algorithm to discover quantitative association rules from huge databases without the need for an a priori discretization, *Expert Syst. Appl.*, vol. 39, no. 1, pp. 585593, Jan. 2012.
- [6] X. Yan, C. Zhang, and S. Zhang, Genetic algorithm-based strategy for identifying association rules without specifying actual minimum support, *Expert Syst. Appl.*, vol. 36, no. 2, pp. 30663076, Mar. 2009.
- [7] H. R. Qodmanan, M. Nasiri, and B. Minaei-Bidgoli, Multi objective association rule mining with genetic algorithm without specifying minimum support and minimum confidence, *Expert Syst. Appl.*, vol. 38, no. 1, pp. 288298, Jan. 2011.
- [8] D. Arotaritei and M. G. Negoita, *An Optimization of Data Mining Algorithms Used in Fuzzy Association Rules Algorithms for Fuzzy Association Rules*, V. Palade, R.J. Howlett, L.C. Jain KES 2003, LNAI 2774, pp. 980985, 2003.
- [9] C. Li, Y. Liu, A. Zhou, L. Kang, and H. Wang, *A Fast Particle Swarm Optimization Algorithm with Cauchy Mutation and Natural Selection Strategy*, Isica 2007, pp. 334343, 2007.
- [10] M. Gen; L. Lin; H.Owada, *Multiobjective Hybrid Evolutionary Algorithms for Data Mining: Survey and Case Study of Clustering*, Proc. Soc. Plant Eng. Japan 2015, Autumn Conf., 2015.
- [11] R. . c Alhaji and M. . Kaya, Multi-objective genetic algorithms based automated clustering for fuzzy association rules mining, *J. Intell. Inf. Syst.*, vol. 31, no. 3, pp. 243264, 2008.
- [12] M. Almasi and M. S. Abadeh, Rare-PEARs: A new multi objective evolutionary algorithm to mine rare and non-redundant quantitative association rules, *Knowledge-Based Syst.*, vol. 89, pp. 366384, Jul. 2015.
- [13] J. Mata, J. Alvarez, and J. Riquelme, Discovering numeric association rules via evolutionary algorithm, *Adv. Knowl. Discov.*, pp. 4051, 2002.
- [14] B. Alatas, E. Akin, and A. Karci, MODENAR: Multi-objective differential evolution algorithm for mining numeric association rules, *Appl. Soft Comput.*, vol. 8, no. 1, pp. 646656, Jan. 2008.
- [15] K. Indira and S. Kanmani, Association rule mining through adaptive parameter control in particle swarm optimization, *Comput. Stat.*, vol. 30, no. 1, pp. 251277, 2014.
- [16] H.-S. Wang, W.-C. Yeh, P.-C. Huang, and W.-W. Chang, Using association rules and particle swarm optimization approach for part change, *Expert Syst. Appl.*, vol. 36, no. 4, pp. 81788184, 2009.

- [17] R. J. Kuo, C. M. Chao, and Y. T. Chiu, Application of particle swarm optimization to association rule mining, *Appl. Soft Comput.*, vol. 11, no. 1, pp. 326336, Jan. 2011.
- [18] P. Sehrawat, *Weighted Association Rule Mining Using Particle Swarm Optimization*, PAKDD 2011 Workshop, LNAI 7104, vol. 2, no. 1, pp. 113, 2012.
- [19] K. N. V. D. Sarath and V. Ravi, Association rule mining using binary particle swarm optimization, *Eng. Appl. Artif. Intell.*, vol. 26, no. 8, pp. 18321840, Sep. 2013.
- [20] J. . V. U. Jithendranath, A. S. V. U. Babu, and G. D. V. U. Sukumar, Hybrid PSO-GSA Method of Solving ORPD Problem with Voltage Stability Constraint, *TELKOMNIKA Indones. J. Electr. Eng.*, vol. 12, no. 8, pp. 58075813, 2014.
- [21] M. Verma and R. Lakhwani, Association Rule Generation by Hybrid A lgorithm based on Particle Swarm Optimization and Genetic Algorithm, *IJCSET*, vol. 5, no. 12, pp. 423426, 2015.
- [22] J. J. Jamian, M. W. Mustafa, H. Mokhlis, and M. A. Baharudin, Implimentation of Evolutionary Particle Swarm Optimization in Distributed Generation Sizing, *IJECE*, vol. 2, no. 1, pp. 137146, 2012.
- [23] M. Krishnaveni, P. Subashini, and T. T. Dhivyaprabha, Improved Canny Edges Using Cellular Based Particle Swarm Optimization Technique for Tamil Sign Digital Images, *IJECE*, vol. 6, no. 5, pp. 21582166, 2016.
- [24] H. Wang, S. Zeng, Y. Liu, and W. Wang, Re-diversification Based Particle Swarm Algorithm with Cauchy Mutation, *2nd Int. Symp. Intell. Comput. Appl.*, vol. 2, no. 1, pp. 362371, 2007.
- [25] Q. Wu, Hybrid forecasting model based on support vector machine and particle swarm optimization with adaptive and Cauchy mutation, *Expert Syst. Appl.*, vol. 38, no. 8, pp. 90709075, 2011.
- [26] Q. Wu and R. Law, Cauchy mutation based on objective variable of Gaussian particle swarm optimization for parameters selection of SVM, *Expert Syst. Appl.*, vol. 38, no. 6, pp. 64056411, 2011.
- [27] Q. Wu, Cauchy mutation for decision-making variable of Gaussian particle swarm optimization applied to parameters selection of SVM, *Expert Syst. Appl.*, vol. 38, no. 5, pp. 49294934, 2011.
- [28] S. M. Gharghory, Hybrid of Particle Swarm Optimization With Evolutionary Operators To Fragile Image Watermarking Based DCT, *IJCSIT*, vol. 3, no. 3, pp. 141157, 2011.
- [29] V. Jancauskas, Empirical Study of Particle Swarm Optimization Mutation Operators, *Balt. J. Mod. Comput.*, vol. 2, no. 4, p. 199, 2014.
- [30] A. Ghosh and B. Nath, Multi-objective rule mining using genetic algorithms, *Inf. Sci. (Ny)*, vol. 163, no. 13, pp. 123133, Jun. 2004.
- [31] J. Kennedy and R. Eberhart, *Particle swarm optimization*, Neural Networks, 1995. Proceedings., IEEE Int. Conf., vol. 4, pp. 19421948 vol.4, 1995.
- [32] Y. Xinjie and M. Gen, Introduction to Evolutionary Algorithms. *Springer London Dordrecht Heidelberg, New York*, 2010.
- [33] I. Tahyudin and H. Nambo, *The Combination of Evolutionary Algorithm Method for Numerical Association Rule Mining Optimization*, in The Tenth International Conference on Management Science and Engineering Management, 2016, p. 1.

**BIOGRAPHY OF AUTHORS**

**Imam Tahyudin** was born in Indramayu, West Java, Indonesia. He is a Ph.D student of Graduate school of NST, division of Electrical Engineering and Computer Science, Kanazawa University, Japan. He was a Lecturer in STMIK AMIKOM Purwokerto, Central Java, Indonesia since 2009. He is a member of the International Society of Management Science and Engineering Management (ISMSEM), Institute of Advanced Engineering and Science (IAES), Association of Computing and Informatics Institutions Indonesia (APTIKOM), Indonesian Computer, Electronics, Instrumentation Support Society (IndoCEISS), Association of Information System (AIS) and Association of Information System for Indonesia (AISINDO).



**Hidetaka Nambo** received his Ph.D degree in 1999 from Kanazawa University. He was a Research Associate in 1999 in the Department of Electrical Information, Engineering Faculty of Kanazawa University. He was a lecturer of Graduate School of Natural Science and Technology since 2015. He is an associate professor of Collage of Science and Engineering. He is engaged in research on the monitoring system by living plants and data mining. He is a member of the IEEE, the IEE of Japan, the IEICE of Japan, the IPSJ of Japan. He can be reached at [nambo@blitz.ec.t.kanazawa-u.ac.jp](mailto:nambo@blitz.ec.t.kanazawa-u.ac.jp).