

1
2
3
4
5
6
7
8

9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38

Original Research Article

Data Clustering using Hybridization Strategies of Continuous Ant Colony Optimization, Particle Swarm Optimization and Genetic Algorithm

ABSTRACT

Nowadays, clustering plays a critical role in most research areas such as engineering, medicine, biology, data mining, etc. Evolutionary algorithms, including continuous ant colony optimization, particle swarm optimization, and genetic algorithms, have been employed for data clustering. To improve searching skills, this paper examines four strategies, combining of continuous ant colony optimization and particle swarm optimization, and proposes a strategy which is a combination of these two algorithms with genetic algorithm. Available methods and the proposed method were implemented over several sets of benchmark data to assess the validity. Results were compared with the results of continuous ant colony optimization and particle swarm optimization. The high capacity and resistance of combined methods are obvious according to results.

Keywords: data clustering; continuous ant colony optimization; particle swarm optimization; genetic algorithm.

1. INTRODUCTION

Clustering is a process of grouping a set of objects into clusters. The internal members of each cluster look very similar to each other and are not similar to the members of other clusters. Clustering is employed in different sciences such as engineering, medicine, social sciences, and marketing. It is of data mining techniques, which extracts models from data without a pre-defined purpose.

To analyze clusters, objects are displayed in a n-dimensional space. Vectors show the properties of objects. The main purpose is to classify n objects into k clusters with objects looking very similar to each other. This paper uses Euclidean distance to minimize the deviation of internal points of clusters from a central point.

Various techniques and algorithms are employed to cluster data. Most problems relate to convex and non-linear clustering with local minimum answers. As a result, an optimal answer may not be found. In algorithms in which the primary answer is randomly selected, a local optimal answer may be found and prevent from reaching an optimal answer. Combining algorithms can improve the main algorithm and give a higher qualified answer. In evolutionary calculations, combination is necessary to improve the optimization algorithm. The performance of an alone algorithm is lower than combined algorithms. In the present algorithm, several new and effective methods have been developed by combining metaheuristic algorithms of continuous ant colony optimization [3], particle swarm optimization [2] and genetic algorithm [1].

The remainder of our paper is organized as follows: section 2 reminds basic concepts, including continuous ant colony optimization (ACOR), particle swarm optimization (PSO) and Genetic Algorithm (GA); section 3 explains combined systems ACOR-PSO, section 4 explains combined system ACOR-PSO-GA; section 5 presents the empirical results of UCI dataset, and conclusions are described in Sections 6.

2. PRELIMINARIES

2.1 CONTINUOUS ANT COLONY OPTIMIZATION

This algorithm was introduced by Sosha and Dorigo [4] to solve optimization problems for continuous functions. According to this method, any line of Pheromone table presents a solution from a collection of decision-making variables. Each solution includes a value from the target function. In this method, primary solutions are randomly found. The new ants in the next generation are produced by a Roulette Wheel probability achieved based on the target function of each solution of the pheromone table. The algorithm is detailed as follows:

1) The target function $f(s_i)$ is calculated for any solution s_i in the pheromone table. Available solutions in the table are arranged based on the target function. Thus, for a minimization problem, we have:

$f(s_1) \leq \dots \leq f(s_i) \leq \dots \leq f(s_k)$, Where k is the number of solution in pheromone table.

2) Weight (w) is calculated for i^{th} solution in the pheromone table as below:

$$w_i = \frac{1}{qk\sqrt{2\pi}} e^{-\frac{(i-1)^2}{2q^2K^2}} \quad (1)$$

In the above equation q shows the learning rate, which is a value between 0 and 1.

3) As to W_i for each available solution in the pheromone table, Roulette Wheel probability P_i is estimated as bellow:

$$p_i = \frac{w_i}{\sum_{j=1}^k w_j} \quad (2)$$

4) Stage 4 is iterated M times to produce M new ants ($M \leq K$): using normal distribution (μ_i^d, σ_i^d), a new ant is achieved for each variable. μ_i^d is a value selected from d^{th} variable and from i^{th} solution in the pheromone table with the probability p_i . σ_i^d is defined as follows:

$$\sigma_i^d = \tau \sum_{j=1}^k \frac{|x_j^d - x_i^d|}{K-1} \quad (3)$$

In the above equation, x_i^d is the d^{th} variable from the i^{th} solution. K is the size of the pheromone table and τ shows the rate of evaporation which is between 0 and 1.

5) M new ants are evaluated and less qualified solutions are replaced in the pheromone table by superior solutions and by M new ants.

2.2 Particle swarm optimization

It was developed by Kennedy and Eberhart [2] in 1995 and has been successfully employed in many scientific and applied areas. This is a population-based algorithm in which anyone is considered as a particle and any population consists of a number of these particles. In PSO, problem-solving space is regarded as a searching environment and anyplace is a problem-dependent solution. In this population, particles tries to find the best situation (best solution) in the searching place (solution space). All particles move according to their speed. The movement of particles in any iteration is calculated by the following formula:

$$v \quad (4)$$

$$x \quad (5)$$

In relations (4) and (5), x_i is the current position of i^{th} particle and v_i is the current speed of this particle. p_i is the best position where have been found by i^{th} , g_i is the best position which has been ever found by the population. W is the weight of inertia that gives a proportion of the previous speed. c_1 and c_2 are the acceleration coefficients and define the best effect of the position of each particle and the best overall position. Rand_1 and rand_2 are two random numbers between 0 and 1. Figure 1 presents the procedure of particle swarm optimization [5,6].

```

Initialize a population of particles with random positions and
velocities in the search space.
While(termination conditions are not met)
{
For each particle  $i$  do
Update the position of particle  $i$  according to equation
(4).
Update the velocity of particle  $i$  according to equation
(5).
Map the position of particle  $i$  in the solution space
and evaluate its fitness value according to the fitness
Function.
Update  $p_i$  and  $g_i$  if necessary.
End for
}
    
```

Fig. 1. Procedure of particle swarm optimization

2.3 Genetic algorithm

The genetic algorithm is an optimization technique and a random search method which has been produced by the concepts of natural selection theory and evolutionary processes. According to Darwin's evolutionary theory, in nature, some people of a population who are more qualified for living are survived and others perish in a competition for living [1]. Two basic operators of genetic algorithm are crossover and mutation. The applied cutting operator is a two-point crossover and the act of mutation occurs accidentally over a gene. Figure 2 displays the procedure of genetic algorithm.

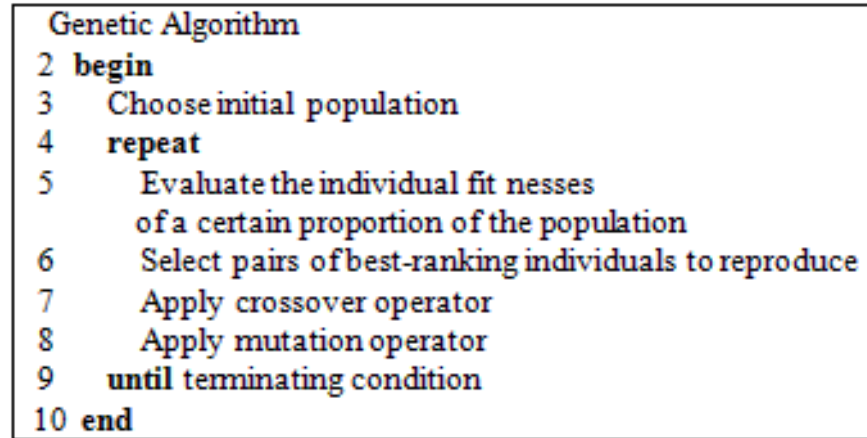


Fig. 2.Procedure of genetic algorithm

100
101
102
103
104
105
106
107
108
109

2.4 Solution structure and objective function

In continuous ant colony optimization, solutions, called pheromones, are recognized in particle swarm optimization as particles and in genetic algorithm as chromosome, which have similar structures. One vector solution is from real numbers in dimensions of $k \times d$. k is the number of clusters and d is the data dimensions, which should be clustered. Figure 3 presents a sample of this solution. All variables are standardized before clustering. In this study, each variable is mapped in $[0,1]$ by linear normalization. Relation (6) is used in this regard.

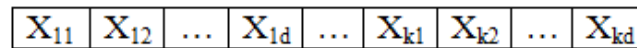


Fig. 3. Structure of a solution

110

$$x^{new} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (6)$$

111
112
113
114
115

In relation 6, x^{new} is the normalized value for x and $\min(x)$ and $\max(x)$ are the smallest and largest data respectively. The target function which is used to evaluate the fitness of clustering is defined as follows:

$$F \quad (7)$$

116
117
118
119
120
121
122
123
124
125
126
127

N_c = number of center

N_s = sample size

$\min\|x_i - c_k\|^2$ = distance between sample i to center k

$d(C_k, C_j)$ = distance between center k and center j

2.5 Data Clustering

Clustering in division method is not under the supervision of data. It is used to divide a set of data into classes of similar data objects. Swarm intelligence, including ant colony optimization for continuous functions and particle swarm optimization, is used in clustering. Ant-based clustering was first introduced by Deneubourg et al [7]. Particle swarm optimization and genetic algorithms have been used in many research studies [8-14]. Studies disclosed that by advancing in this area, these methods can prevent from falling into local optimum and have better performance than some other traditional clustering algorithm.

3. Hybridization Strategies of Continuous Ant Colony Optimization and Particle Swarm Optimization

This section explains four combined models created by combination of the Continuous Ant Colony Optimization and particle swarm optimization.

3.1 Sequential Approach

In this method, ACOR and PSO share a set of answers with each other, which is called “pheromone-particle table”. This set of answers includes pheromone in ACOR and current solution in PSO. Based on pheromone-particles table, PSO produces newer particles and replaces lower particles and solutions in pheromone-particles table with solutions superior to new particles. According to pheromone- particles table updated by PSO, ACOR produces new ants and replaces lower particles and solutions in pheromone-particles with solutions superior to new ants. The features of the sequential method include: (1) superior solutions produced by PSO and ACOR can be kept in pheromone table; (2) ACOR produces new ants and replaces low solutions in pheromone table. This brings variety in the pheromone table and prevents from local optimization. Figure 4 shows the main stages of the sequential approach [15].

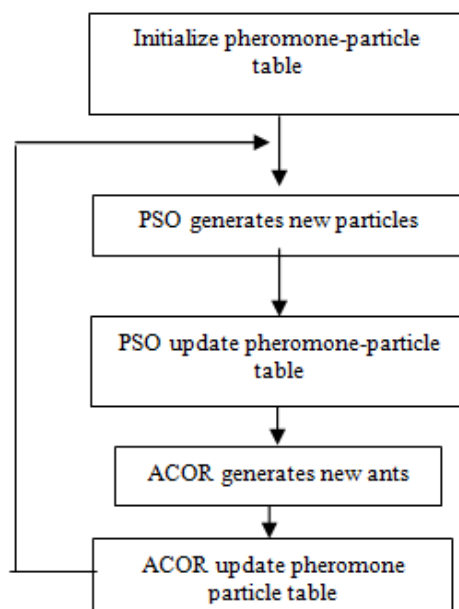


Fig. 4. Sequential approach

3.2 Parallel Approach

158 In this method, According to pheromone-particles table, PSO produces new particles and ACOR
 159 produces new ants. Underneath solutions in pheromone-particles table are replaced by solutions
 160 superior of K new particles and M new ants. Figure 5 shows the main stages of the Parallel approach
 161 [15].
 162

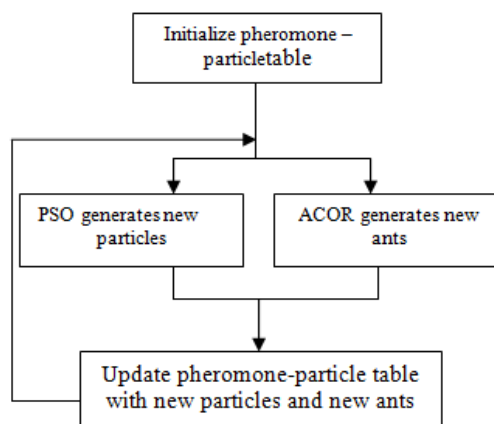


Fig. 5. Parallel approach

163 3.3 Sequential Approach with the enlarged pheromone-particle table

164
 165 In this method, PSO produces k new particles based on pheromone-particles table. These particles are
 166 combined with pheromone-particles table and form an extended table with 2k in size. In PSO, Pbest
 167 and Gbest are achieved from updated pheromone-particles table. In ACOR, ants are extended based
 168 on pheromone tables with varied production. This method prevents from local optimum. Figure 6
 169 shows the main stages of this approach [15].
 170

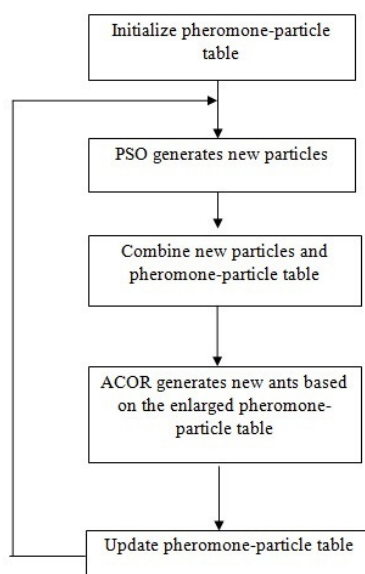


Fig. 6. Sequential approach with the enlarged pheromone-particle table

171 3.4 Global best exchange

172

173 In global best exchange method, PSO produces new particles based on its particle table. ACOR
174 produces new ants based on its pheromone table. Two models exchange their best solutions. This
175 approach keeps the main features of PSO and ACOR. Figure 7 shows the main stages of this approach
176 [15].
177

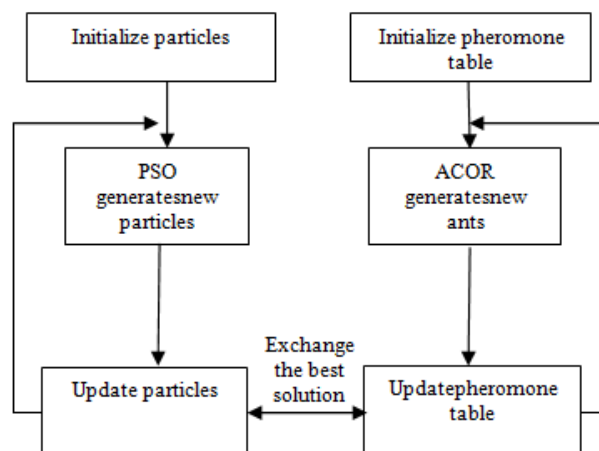


Figure7. Global best exchange approach

178 4. Hybridization Strategies of Continuous Ant Colony Optimization, Particle Swarm 179 Optimization and Genetic algorithm (The proposed model) 180

181
182 In such combined strategy, ACOR, PSO and genetic algorithm share a set of answers with each other
183 which is called pheromone-particles-chromosome table. This set of answers includes pheromone in
184 ACOR, current solutions in PSO and chromosomes in genetic algorithm. Based on pheromone-
185 particles-chromosome, PSO produces new particles and replaces lower particles and solutions in
186 pheromone-particles-chromosome table with solutions superior to new particles. According to
187 pheromone- particles-chromosome table updated by PSO, ACOR produces new ants and replaces
188 lower particles and solutions in pheromone-particles-chromosome table with solutions superior to new
189 ants. Then, based on this table, genetic algorithm produces new chromosome by ACOR. Lower
190 solutions in pheromone-particles-chromosome table are then replaced by superior solutions of new
191 chromosomes. The features of this method include: (1) superior solutions produced by PSO, ACOR
192 and GA can be kept in solution table; (2) ACOR produces new ants and replaces low solutions in
193 solution table; (3) GA produces new chromosomes and replaces low solutions in solution table. This
194 brings variety in the solution table and prevents from local optimization. Figure 8 shows the main
195 stages of this approach.
196

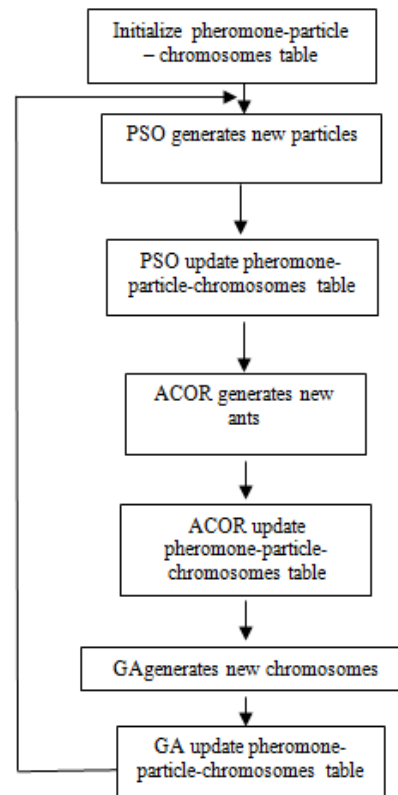


Fig. 8. Hybridization Strategies of Continuous Ant Colony Optimization, Particle Swarm Optimization and Genetic algorithm

5. Experiments

Five combined models, independent PSO and independent ACOR with four dataset from UCI dataset for clustering. In table 1 show dataset of UCI. Relation (7) displays the efficiency of clustering. Tables 3 indicate the average answer of current algorithms over Iris dataset and the average answer of current algorithms over Wine dataset, and Tables 4 indicate the average answer of current algorithms over Contraceptive Method Choice(CMC) dataset and the average answer of current algorithms over Liver Disorders dataset. To summarize, the sequential method was introduced as hybrid I, parallel method as hybrid II, the sequence with enlarged pheromone table as hybrid III, global best exchange as hybrid IV and ACOR-PSO-GA as hybrid V.

The proposed implementation was executed on the Matlab9 platform, with an Intel Core 2 Duo CPU running at 3.0GHz and 4GB RAM. Table 2 presents parameters required to implement algorithms. Empirical results show that combined methods are effective and hybrid V is more efficient than other combined methods. In this model, superior solutions in iterations are kept. Since three different algorithms are used in this method, there is a wide variety of population. As it is improbable for a global optimization to occur, hybrid III is better than three other combined methods. The reason is population variation which happens because pheromone-particles table is doubled and prevents from local optimization. Empirical results also show that ACOR is more efficient than PSO and combined methods are preferred to ACOR and PSO.

Table 1. Datasets from the UCI repository

No	Names	#Instances	Numeric features
1	Iris	150	3
2	Wine	178	13
3	Contraceptive method choice (CMC)	1473	9
4	Liver Disorders	354	6

Table 2. System parameter setting

Parameter	Hybrid I Hybrid II Hybrid III Hybrid IV	PSO	ACOR	Hybrid V
Learning rate C1	1.5	1.5		1.5
Learning rate C2	1.5	1.5		1.5
Learning rate q	1.5		1.5	0.95
Evaporation rate	1.5		0.05	0.05
Number of particles	20	20		20
Size of Pheromone-particle table	15			15
Number of ants	5		20	5
Number of iterations	100	100	100	100
Number of chromosomes				20
Size of Pheromone-particle- chromosomes table				30
Mutation rate				0.4
Crossover rate				0.05

Table 3. The average answer in four runs over Iris dataset and wine dataset

Evolutionary Algorithms	The average cost of the best solution in four runs over Iris dataset	The average cost of the best solution in four runs over Wine dataset	Average cpu time in four runs over Iris dataset	Average cpu time in four runs over Wine dataset
PSO	5.215725	9.82015	1 min	1.2 min
ACOR	2.906325	8.28645	1.2 min	1.3 min
Hybrid I	1.4843	5.364175	2.25 min	2.3 min
Hybrid II	1.6556	6.912975	2.26 min	2.28 min
Hybrid III	0.53655	0.512925	2.4 min	2.51 min
Hybrid IV	1.5965	2.074675	1.25 min	1.31 min
Hybrid V	0.2057	0.5119	2.42 min	2.52 min

Table 4. The average answer in four runs over CMC dataset and Liver Disorders dataset

Evolutionary Algorithms	The average cost of the best solution in four runs over CMC dataset	The average cost of the best solution in four runs over Liver Disorders dataset	Average cpu time in four runs over CMC dataset	Average cpu time in four runs over Liver Disorders dataset
PSO	31.8046	13.9330	2.4 min	1.4 min
ACOR	25.2554	8.7916	2.46 min	1.45min
Hybrid I	11.1978	6.3008	3.45 min	2.45 min
Hybrid II	10.4297	5.2128	3.55 min	2.50 min
Hybrid III	7.6416	1.7849	4.02 min	2.55 min
Hybrid IV	11.1487	5.1094	2.5 min	1.52 min
Hybrid V	4.3351	0.9739	4.10 min	3 min

6. Conclusions

To improve data clustering, four combined strategies were used. These strategies were created by combining continuous ant colony optimization and particle swarm optimization. Another combined method was also proposed, which is a combination of these two algorithms with genetic algorithm. Simulation results revealed that the proposed method is preferred to other combined methods, PSO and ACOR.

REFERENCES

- [1] Goldberg, David E, "Genetic Algorithms in Search", Optimization and Machine Learning, Kluwer Academic Publishers, Boston, MA. 1989.
- [2] Kennedy J. and Eberhart R.. "Particle Swarm Optimization". Proceedings of the IEEE International Conference on Neural Networks. Perth, Australia. 1995; pp.1942-1945.
- [3] Eberhart R, Kennedy J. "A new optimizer using particle swarm theory". in: Proceedings 6th International Symposium on Micro Machine and Human Science. Nagoya, Japan. 1995; pp. 39-43.
- [4] Socha K, Dorigo M. "Ant colony optimization for continuous domains". European Journal of Operational Research. 2008; 185(3): 1155-1173.

- 259 [5] J. Han, M. Kamber, "Data Mining Concepts and Techniques (2nd ed.)", Morgan
260 Kaufmann Publishers. San Francisco, USA. 2006.
261
- 262 [6] Merwe D.W, Engelbrecht A.P, "Data clustering using particle swarm optimization". in:
263 Proceedings of IEEE Congress on Evolutionary Computation. Canberra, Australia. 2003;
264 pp.215-220.
265
- 266 [7] Deneubourg J, Goss S, Franks N, Sendova-Franks A, Detrain C, Chretien L. "The
267 dynamics of collective sorting robot-like ants and ant-like robots". in: Proceedings of the 1st
268 International Conference on Simulation of Adaptive Behavior: From Animals to Animats.
269 Paris, France. 1991; pp.356-363.
270
- 271 [8] Omran M, Engelbrecht A.P, Salman A. "Particle swarm optimization method for image
272 clustering". International Journal of Pattern Recognition and Artificial Intelligence. 2005;
273 19(3): 297-322.
274
- 275 [9] Merwe D.W, Engelbrecht A.P. "Data clustering using particle swarm optimization". in:
276 Proceedings of IEEE Congress on Evolutionary Computation. Canberra, Australia. 2003;
277 pp.215-220.
278
- 279 [10] Cui X, Potok T.E. "Document clustering analysis based on hybrid PSO+K-means
280 algorithm". Journal of Computer Science. 2005; 1(3): 27-33.
281
- 282 [11] Zhenkui P, Xia H, Jinfeng H. "The clustering algorithm based on particle swarm
283 optimization algorithm". in: Proceedings of International Conference on Intelligent
284 Computation Technology and Automation. Changsha, Hunan, China. 2008; pp.148-151.
285
- 286 [12] Ahmadyfard A, Modares H. "Combining PSO and K-means to enhance data clustering".
287 in: Proceedings of the 4th IEEE International Symposium on Telecommunications. Tehran,
288 Iran. 2008; pp.688-691.
289
- 290 [13] Tsai C.Y, Kao I.W. "Particle swarm optimization with selective particle regeneration for
291 data clustering". Expert Systems with Applications. 2011; 38(6): 6565-6576.
292
- 293 [14] Maulik U, Bandyopadhyay S. "Genetic algorithm-based clustering technique". Pattern
294 Recognition. 2000; 33: 1455-1465.
295
- 296 [15] Huang C.L, Huang W.C, Chang H-Y, Yeh Y.C, Tsai C.Y. "Hybridization Strategies for
297 Continuous Ant Colony Optimization and Particle Swarm Optimization Applied to Data Clustering".
298 Applied Soft Computing. 2013.
299