e-ISSN: 2349-9745
p-ISSN: 2393-8161

Scientific Journal Impact Factor (SJIF): 1.711

International Journal of Modern Trends in Engineering and Research

www.ijmter.com

Study and Analysis of Cluster Optimization Algorithms: Particle Swarm Optimization and Cohort Intelligence

M. S. Bhosale[1], R. V. Mane[2]

[1]*Department of Technology, Shivaji University, Kolhapur*
[2]*Department of Technology, Shivaji University, Kolhapur*

Abstract— Clustering is one of the most fundamental tasks in data mining. The process of clustering partitions the given data points into a group based on some similarity measures like Euclidean metric. However finding the best clustering result is difficult. There are various cluster optimization techniques presented in literature to improve result of clustering algorithm. The recent trend in cluster optimization uses Swarm Intelligence (SI) such as ant colony optimization, honey bee mating, particle swarm optimization etc. which tries to find better clustering results. The SI is influenced by natural tendency of swarms in environment. Another approach successfully applied for cluster optimization is Cohort Intelligence (CI). The CI also uses natural and society tendency of individuals learning from each other while living in particular group. This paper presents the theoretical analysis between Particle Swarm Optimization (PSO) and Cohort Intelligence (CI).

Keywords- Clustering, Cohort Intelligence, Data mining, Particle Swarm Optimization, Swarm Intelligence.

I. INTRODUCTION

Clustering is the process of dividing input data set into separate groups based on some common properties between the objects [1]. K-means is most famous clustering algorithm presented 50 years before and evolved later by making different modifications in it. The followers of K-means includes K-means ++, K medoids, PAM, CLARA and many more [18] [20]. The clusters obtained using these algorithms may not always results in best solution because they might suffers from problems of local optima and slow or premature convergence. In order to address the local minima problem of clustering algorithms many heuristic algorithms are proposed. These algorithms are based on concept of combinatorial optimization. Basically optimization is a mathematical discipline that consists of the finding of minimum and maximum of functions by choosing input values within allowed interval. Optimization is applied in many fields including Operations Research, artificial intelligence and computer science. The optimization can be a single objective or multiobjective. In single objective optimization solution is achieved based on single objectives whereas multiobjective optimization focuses on more than one objective.

To generate clusters from given set of data points and optimizing these clusters could be achieved by many methods. Some advanced optimization techniques such as Genetic Algorithm (GA), Simulated Annealing (SA), Ant Colony Optimization, Honey-bee mating are used in cluster optimization process. The figure 1 shows the classification of cluster optimization algorithms. The algorithms can be classified into three groups; one is Simulated Annealing (SA) where we can achieve single objective evolution criteria; second is Evolutionary Algorithm (EA) which includes the Genetic Algorithm (GA) and Differential Evolution (DE) and they are based on population of solution instead of single solution; and third one is Swarm Intelligence (SI) which is based on collective behavior of self organized system of swarms in environment.

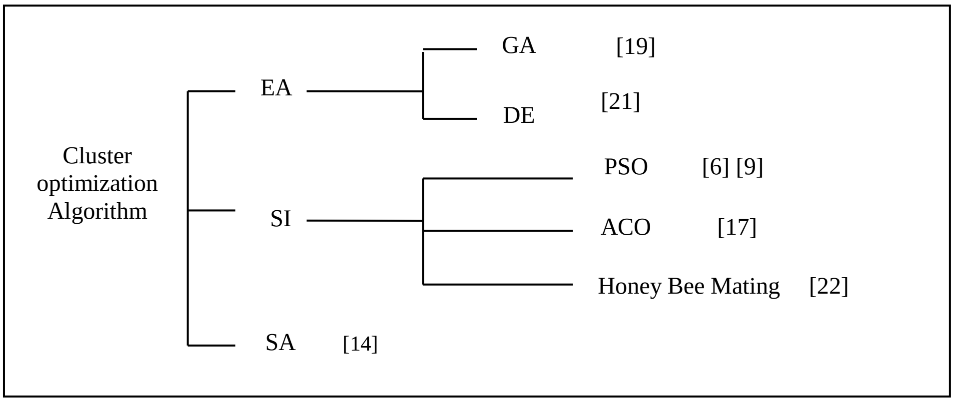

Figure 1. Classification of Cluster Optimization Algorithm

The main focus of this work is to compare the two algorithms PSO and CI and find out their strengths and weakness.

In 1995 Eberhart and Kennedy presented a stochastic population based optimization approach called as Particle Swarm Optimization (PSO) [9]. The word "swarm" refers to irregular movements of the particles in the problem space. It considers the population of particles that moves through the problem hyperspace with given velocities [5]. Over a number of iterations, the velocities of the individual particles are stochastically adjusted according to the previous best known position for the particle itself and the neighborhood best position. Both the particle best and the neighborhood best are derived from a user defined fitness function [10], [11]. The optimal solution for each particle obtained naturally by their movements.

Recently Anand Kulkarni et al proposed a novel methodology of Cohort Intelligence (CI) [2]. The Cohort means group of such a candidate (objects) who interacting and competing with one another to achieve some individual goal which is common to all candidates. Every candidate in the cohort tries to improve its own behavior by observing behavior of other candidates. This process repeats until overall group get saturated with same behavior.

The Reminder of this paper is organized as follows. Section 2 presents the related work regarding cluster optimization. The details of PSO algorithm is presented in section 3. In section 4 idea of Cohort intelligence is explained. Section 5 explains the comparison between CI and PSO. At last section 6 present the concluding remark.

II. RELATED WORK

To perform cluster optimization presented algorithms are used as it is or could combine with other state of art algorithms to improve their results. In [13] k-means is combined with the genetic algorithm to create optimized clusters. Selim and Alsultan proposed a method of cluster analysis based on simulated annealing [14]. In [15], a honey-bee mating optimization was applied for solving clustering problems. An ant colony optimization (ACO) for clustering problems is proposed by Shelokar, Jayaraman, and Kulkarni [17]. Taher Niknam et.al in [16] suggests the hybrid method of cluster analysis by combining ACO and SA. PSO is combined with some other methods and has some forms like Chaos PSO (CPSO) [8] which improves the searching behavior of PSO, Robust Hybrid PSO (RHPSO) [7] based on based on piecewise linear chaotic map (PWLCM) and sequential quadratic programming. Linearly Decreasing Weight PSO (LDWPSO) [6] effectively balances out the global and local search abilities from the swarm.

III. PARTICLE SWARM OPTIMIZATION

Particle Swarm Optimization (PSO) is an evolutionary algorithm resembles the social behavior of swarm like bird flocking and fish schooling [9]. The idea of algorithm is very simple where a flock of birds circling over an area where they can smell a hidden source of food. The one who is closest to

the food chirps the loudest and the other birds swing around in his direction. If any of the other circling birds comes closer to the target than the first, it chirps louder and the others veer over toward him. In PSO a number of swarm (particles) are placed in the search space and each evaluates the objective function for its current position. Like swarm each particle is moving with some velocity in search space toward its best position and it is achieved by combining some previous information, with its current information. The particles iteratively evaluate the fitness of the candidate solutions and remember the location where they had their best success. The individual's best solution is called the particle best (local best). Each particle makes this information available to their neighbors. A particle then adjusts its position according to own experience as well as the experience of neighboring particles.
Each object in particle swarm is represented by three vectors namely current position , previous best position and the velocity of the particle [6]. The current position of particle is represented by its coordinates and at each iteration next best position is calculated. The value of best position is stored in vector and it will be used to compare values in later iterations. The velocity vector defines next position for the particle and adjusts accordingly when particle is moving towards the optimal results.

Algorithm for PSO

1. Initialize particles (population).
2. Loop
3. Calculate fitness value for each particle and initialize best fitness value as $pbest_i$ among population.
4. Compare current fitness value with $pbest_i$. If current value is better than $pbest_i$ then update $pbest_i$ with current value. Otherwise continue with $pbest_i$.
5. Find the particle in the neighborhood with the best fitness value so far, and consider this value as a global best $gbest_i$.
6. Change the velocity and position of the particle according to the following equation.

$$v_i^{n+1} = w\,v_i^n + C1 \cdot rand(\cdot) \times (p_i - c_i) + C2 \cdot Rand(\cdot) \times (pbest_i - c_i) \quad (1)$$

$$c_i = c_i + v_i \quad (2)$$

7. Stop if stopping criteria is satisfied.
8. End Loop.

W is the inertia weight [12], C1 and C2 are positive acceleration coefficient; Rand () and rand () are two independent random variables ranges between 0 and 1; Superscript represent the iteration number; represents last known best value for pi, is the global best solution for the particle.

IV. COHORT INTELLIGENCE

Cohort analysis is a branch of behavioral analytics that takes the data from given platform and instead of considering all users into one unit it breaks them into separate related groups for analysis purpose. These related groups generally referred as cohorts, which usually share common characteristics or behavior within a defined time interval. By observing these patterns (behavior) of time, decisions are taken in data mining application. Cohort analysis plays important role in big data analysis and business analytics.

In 2013 Anand Kulkarni et. al. proposed a new cluster optimization technique based on cohort intelligence [2]. The idea of the algorithm is based on natural and society characteristics of the individuals living in particular group. As all individuals have some qualities different than others; and they always try to improve their qualities by interacting with their neighbors. The behavior of the candidate is defined by the qualities associated with it. Cohort intelligence mimics the same scenario where each candidate in the group makes competition and interaction to all other members of the group to achieve some common goal for all candidates.

Notations:

- $f(X^c)$ Objective function (behavior of each individual) where c= (1, 2, . . . C)
- C Number of candidates in cohort.
- $X = [x_1, \dots, x_i, \dots x_m]$ set of qualities for candidate X.
- Si Sampling interval for each quality given by $x_i^{min} \leq x_i \leq x_i^{max}$.
- ϵ Convergence Parameter.

Algorithm for CI

1. Initialize all parameters like number of candidates C, Convergence Parameter ϵ, number of iterations n, sampling interval for each quality Si, sampling interval reduction factor $r \in [0,1]$ and number of variation t.
2. Calculate the probability of every candidate c to being selected by using

$$p^c = \frac{\frac{1}{f*(X^c)}}{\sum_{c=1}^{C} \frac{1}{f*(X^c)}}$$

3. Generate random number using rand $r \in [0,1]$ and apply roulette wheel approach to decide to follow the behavior and qualities of the candidate.

4. Every candidate minimizes the sampling interval for its every feature to its local neighborhood.

5. Samples qualities from updated interval and find the associated behavior.

6. If there is no change in behavior of each candidate, consider cohort is saturated.

7. Accept any behavior from available as a final objective function and stop if number of iteration equals to or cohort is saturated.

5.1 Advantages and disadvantage of PSO

5.1.1 Advantages:

1) PSO can be applied into both scientific research and optimization problems of different engineering fields.
2) PSO doesn't have overlapping and mutation calculation.
3) Compared with the other developing calculations, it shows bigger optimization ability.
4) PSO can be combining with other conventional methods easily to solve special problems.

5.1.2 Disadvantages:

1) The method cannot works for the problems of non-coordinate system. [4]
2) Original PSO can easily trap into local optima.

5.2 Advantages and disadvantage of CI

5.2.1 Advantages:

1) CI has great potential to solve variety of optimization problems.
2) It uses self supervising mechanism which new in optimization field.

5.2.2 Disadvantages:

1) It may converge very slowly [3].

VI. CONCLUSION

The PSO and CI both have common similarity as they are inspired from natural behavior of the living organism. PSO applies swarm intelligence to achieve solution to complex scientific problems where as CI considers human tendency to solve optimization problem. Both have its pros and cons but they can combine with other heuristic methods to improve their performance. In case of cluster optimization CI can show best results as compared to PSO because of its self supervised nature whereas PSO requires less parameter tuning as compared to CI.

REFERENCES

[1] Anil K. Jain, Richard C. Dubes "Algorithm for Clustering Data"**.** Prentice Hall Advanced Reference Series. 1998.
[2] Anand J. Kulkarni Ishan P. Durugkar, Mrinal Kumar. "Cohort Intelligence: A Self Supervised Learning Behavior", IEEE International Conference on Systems, Man, and Cybernetics. 2013.
[3] Ganesh Krishnasamy, Anand J. Kulkarni, Raveendran Paramesran "A hybrid approach for data clustering based on modified cohort intelligence and K-means" Science Direct. Expert Systems with Applications 41 (2014) 6009–6016.
[4] V.Selvi, Dr.R.Umarani Comparative Analysis of Ant Colony and Particle Swarm Optimization Techniques International Journal of Computer Applications (0975 – 8887)Volume 5– No.4, August 2010.
[5] Venayagamoorthy, Ganesh K. and Harley, Ronald G., "Swarm intelligence for transmission system control" (2007). Faculty Research & Creative Works. Paper 1009.
[6] Riccardo Poli, James Kennedy, Tim Blackwell "Particle swarm optimization: An overview**"** Springer Science and Business Media, LLC (2007).
[7] W. Xu, ZGeng X. Gu. "A Piecewise Linear Chaotic Map and Sequential Quadratic Programming based Robust Hybrid PSO" Information Science. 218.pp. 85-102. 2013.
[8] L. Lu W. M. Zhong, F. Quan, "An Improved Chao-particle swarm optimization algorithm" Journal of East China University of Science and Technology (Natural Science Edition) 36, 267-272, 2010.
[9] J. Kennedy, and R. Eberhart, "Particle swarm optimization," in Proc. Of the IEEE Int. Conf. Neural Networks (ICNN), Vol. 4, Nov. 1995, pp. 1942-1948.
[10] R. Eberhart, and J. Kennedy, "A new optimizer using particle swarm theory," in Proc. 6th Int. Symp. Micro Machine and Human Science (MHS), Oct. 1995, pp. 39-43.
[11] R. Eberhart, Y.Shi and J. Kennedy, Swarm intelligence. San Francisco, CA: Morgan Kaufmann, 2001.
[12] Y. Shi, R. Eberhart, "A Modified Particle Swarm Optimizer," Proc. IEEE Int. Conf. on Evolutionary Computation, pp. 69-73, 1998.
[13] Sonia Sharma, ShikhaRai "Genetic K-Means Algorithm – Implementation and Analysis**"** International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-1, Issue-2, June 2012.
[14] Selim, Sz; Alsultan, K "A Simulated Annealing Algorithm for the Clustering Problem" Pergamon-Elsevier Science Ltd, Pattern Recognition; pp: 1003-1008; Vol: 24.
[15] Fathian, M., Amiri, B., & Maroosi, A. (2007). Application of honey-bee mating optimization algorithm on clustering. Applied Mathematics and Computation, 190, 1502–1513.
[16] Taher Niknam, Bahman Bahmani Firouzi and Majid Nayeripour "An Efficient Hybrid Evolutionary Algorithm for Cluster Analysis" World Applied Sciences Journal 4 (2): 300-307, 2008, ISSN 1818-4952.
[17] Shelokar, P. S., Jayaraman, V. K., & Kulkarni, B. D. (2004). An ant colony approach for clustering. Analytica Chimica Acta, 509, 187–195.
[18] Pavel Berkhin "Survey of clustering data mining technique" Accrue Software Inc.
[19] Ujjwal Maulik, Sanghamitra Bandyopadhyay "Genetic algorithm-based clustering technique" Pergamon-Elsevier Science Ltd Pattern Recognition 33 (2000) 1455-1465.
[20] Arthur, D., & Vassilvitskii, S. "K-means++: The advantages of careful seeding". In Proceedings of the eighteenth annual ACM–SIAM symposium on discrete algorithms SODA '07, Philadelphia, PA (pp. 1027– 1035) 2007.
[21] Abazar Shamekhi, " AN IMPROVED DIFFERENTIAL EVOLUTION OPTIMIZATION ALGORITHM**"** IJRRAS 15 (2) May 2013.
[22] Mohammad Fathian, Babak Amiri, Ali Maroosi "Application of honey-bee mating optimization algorithm on clustering" Elsevier, Applied Mathematics and Computation 190 (2007) 1502–1513

