Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 167 (2020) 531–540

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Computational Intelligence and Data Science (ICCIDS 2019)

Hybrid Artificial Chemical Reaction Optimization Algorithm for Cluster Analysis

Hakam Singh[a] and Yugal Kumar[a,*]

[a] Department of Computer Science and Engineering, Jaypee University of Information Technology, Solan (H.P.) 173234, India.

Abstract

Heuristic algorithms have significant contribution in the clustering field. In present work, a hybrid version of the artificial chemical reaction optimization algorithm (HACRO) is proposed to optimize clustering problems. As exploration and exploitation are two major aspects that require balanced coordination among algorithmic steps. The artificial chemical reaction suffers from slower convergence speed due to its poor exploitation mechanism. Moreover, it requires more execution time. Henceforth, to enhance the convergence speed and to make balance among algorithmic space a hybrid version of ACRO is developed. In present work, the artificial chemical reaction optimization algorithm is incorporated with crossover and mutation operator of genetic algorithm. Further, the efficiency of the HACRO algorithm is examined on seven benchmark datasets and collated with ACO, PSO, K-means, GA, ICSO and ACRO clustering algorithms. The present investigation indicated that the proposed algorithm works efficiently in clustering field.

Keywords: Heuristic; hybridized; clustering; crossover; mutation.

1. Introduction

Data Mining is the process of discovering unrevealed patterns from the database. Mainly four types of techniques, supervised, unsupervised, semi-supervised and active user are used to extract patterns. The supervised technique inspects the patterns with consulting the class labels and unsupervised technique inspects the patterns without consulting the class labels. The semi-supervised technique uses both supervised and unsupervised techniques, while

* Corresponding author.
E-mail address: yugalkumar.14@gmail.com

10.1016/j.procs.2020.03.312

in active user approach user plays an active role in the pattern extraction process. The clustering process is an unsupervised technique that groups data objects into different clusters without consulting the class labels. From the last few decades, clustering has gained wide attention of researchers and also accepted in diverse research domains [1-6]. To obtain the optimal solution for clustering problems number of meta-heuristic algorithms have been reported in literature. These algorithms are inspired from natural phenomenon. These natural phenomena can include swarm intelligence, insect's behavior, well-defined law of physics and some other natural process of living beings [7-11]. Meta-heuristic algorithms possess unique characteristics that attain optimal results. The major advantage of these algorithms is flexibility i.e. the algorithm can adapt itself according to the nature of the problem. Moreover, these algorithms are significant due to a variety of solution strategies.

In literature, numbers of meta-heuristic algorithm have been developed based on natural phenomenon's, swarm behaviors and other natural inspirations. It is observed that algorithms get circumscribed from few issues like local optima, diversity and slower convergence rate. To solve these issues certain enhancements in procedural space and improvements are done. In this work, a hybrid version of the ACRO algorithm is proposed to unravel the clustering problems. The exploration and exploitation processes are major aspects that directly affect the convergence speed. The artificial chemical reaction optimization algorithm suffers from slower convergence rate, due to unbalanced local and global search mechanism. While on initialization aspect, it is better than random initialization method. To augment the convergence rate and to make the balance among exploration and exploitation process the crossover and mutation operator of the GA are introduced in the proposed algorithm. The hybridization results have improved coordination among exploration and exploitation processes. The efficacy of the HACRO algorithm is examined over seven datasets. Further, the accomplishment of HACRO is compared with ACO, PSO, K-means, GA, ICSO and ACRO clustering algorithms. The previous studies related to this work summarized in Section 2 and the Hybrid ACRO algorithm is presented in Section 3. The simulation results and conclusion of the developed algorithm are discussed in section 4 and 5, respectively.

2. Literature

In the past few decades, the various algorithms are reported in the clustering field, some of them are summarized in this section. Prakash and Singh have reported a hybrid Gbest-guided artificial bee colony algorithm for clustering problems [12]. In this work, a crossover operator is introduced for the improvement of social learning among bees and search space. Moreover, it is incorporated with Gbest-guided search procedure to enhance the convergence speed. The efficiency of the addressed algorithm is examined on twelve datasets and compared with ABC and several other clustering algorithms. Further, Das et al. have reported a modified bee colony optimization algorithm in clustering field [13]. In the modified bee colony optimization algorithm, the fairness and cloning concepts are added. In the forward pass, the fairness concept is applied on all bees to investigate the search space. While in the backward pass, bee's comparison concept is introduced to obtain the best quality of the solution. Moreover, to enhance convergence speed a cloning concept is also incorporated in addressed work. Further, it is combined with K-means and named as MKCLUST and KMCLUST. The proficiency of the addressed method is examined on seven benchmark datasets and compared with ACO, PSO and K-means algorithms. The MBCO algorithm delivers the state of art results in clustering field.

To solve clustering problems K-means algorithm is combined with mussels wandering optimization algorithm [14]. In this work, a new clustering ensembles (CE) framework based on weights is introduced to perform clustering. The efficiency of the addressed algorithm is examined on nine datasets and compared with K-PSO and K-means algorithms. From experimental results, it is noticed that addressed algorithm performs well in the clustering field. Cura has introduced a metaheuristic algorithm named as particle swarm optimization algorithm in the clustering field [11]. In this work, the ''gbest neighbourhood topology'' is followed, according to which each particle move towards the best position. The best position at the end of iteration is preserved. Kumar and Sahoo have reported an opposition learning based improved cat swarm algorithm for cluster analysis [15]. In this work, the opposition learning based mechanism is prescribed to improve the diversity mechanism. Moreover, Cauchy and mutation operator are also incorporated to prevent algorithm from local optima situation especially in tracing mode. The accomplishments of addressed algorithm are evaluated on ART1, ART2, iris, wine, CMC, cancer datasets and compared with, PSO, CSO and K-means clustering algorithms. From simulation results, it is seen that addressed algorithm provides better result in clustering field. Senthilnath et al. have reported the robust optimization technique inspired from echolocation behavior of bats. The swarm intelligence of bats is used to solve crop classification problem in an efficient manner [16].

Zabihi and Nasiri have presented a new Hd-ABC (History Driven Artificial Bee Colony) algorithm in the clustering field [17]. In this work, binary space partitioning tree is used to store appreciated information. Moreover, the guided anisotropic search strategy is also adopted to enhance the exploitation ability of the addressed algorithm. The proficiency of Hd-ABC algorithm is examined on both artificial and real-life datasets and compared with K-means, ACO, BB-CS, CPSO, IKH, ABC, CABC, HABC, Two-step ABC, BSF-ABC and ABCL algorithms. From simulation results it is cleared that reported algorithm provides good quality clusters. Kumar et al. have combined the ant colony optimization and ant lion optimization algorithms to solve clustering problems efficient manner [18]. To avoid local optima situation and maintain balance between exploration and exploitation process a local search mechanism is introduced in this work. The efficiency of the mentioned algorithm is evaluated on five benchmark datasets and compared with K-means, ACO and Hybrid ALO algorithms. Authors have claimed that aforementioned algorithm is superior than other compared algorithms.

To optimize clustering problems a meta-heuristic algorithm inspired from the cuttlefish behavior is addressed by Kowalski et al. [19]. In this work, reflection and visibility processes are utilized to generate new solutions. Further, cuttlefish cells are grouped into four types to identify new patterns. The ability of the addressed algorithm is examined on twelve datasets and compared with K-means algorithm. The addressed algorithm provides better quality results. Bijari et al. have introduced a MEBB-BC (Memory-Enriched Big Bang Big Crunch) algorithm in clustering field [20]. In order to make a balance among the exploration and exploitation processes, a memory concept is introduced in this work. Further, the fitness function is evaluated to update the memory space. The performance is examined on several test functions and six real-life datasets. Moreover, to prove its importance it is compared with GA, PSO, GWO, BB-BC and K-means clustering algorithms. Authors have claimed its superiority over other compared algorithms. To achieve automatic data clustering a symbiotic organism search algorithm is reported in [21]. In this work, symbiotic interaction strategies are utilized. The new candidate solution is generated by emulating the biological cooperation between organisms. The working procedure of the proposed algorithm is divided into three phases: mutualism phase, commensalism phase and parasitism phase. Each organism arbitrarily interacts with other organisms. The performance is examined on both artificial and real-life datasets and compared with well-known clustering algorithms. From experimental results, it is clarified that addressed algorithm performs well in the clustering field.

3. Proposed work: Hybrid Artificial Chemical Reaction Optimization Algorithm for Cluster Analysis.

In this section, the hybrid version of the ACRO algorithm incorporated with crossover and mutation operators to solve clustering problems is discussed. The clustering process decomposes the dataset into several disjoint clusters. The data objects associated with the clusters are more similar in nature in comparison to data objects associated with other clusters. In clustering process, distance among data objects and centroids is calculated and data objects are assigned to respective clusters using these distance values. In this work, the Euclidean distance is taken as a distance measure Equation (1).

$$\mathrm{dis}(X_i, C_j) = \sqrt{\sum_{k=1}^{d} (X_{ik}, C_{jk})^2} \tag{1}$$

Where X_i and C_j denote data points and cluster centres. The steps of Hybrid ACRO algorithm for clustering are detailed below. Further, graphical view of the executional steps is also provided in Fig. 2.

3.1 Steps of Hybrid Artificial Chemical Reaction Optimization Clustering Algorithm

Step 1: Initialization

In this step, basic algorithmic parameters like number of reactants, number of clusters, termination condition etc are set. Here, the number of reactants is directly proportional to the number of clusters (K).

Step 2: Selection and evaluation

The selection process starts with random assemblage of reactants (R_1, R_2) from the datasets. These two reactants are further evaluated to generate rest of reactants using Equations (2-9). For example, if K = 2 means two new reactants i.e. R_3, R_4 are generated from R_1 and R_2 using Equations (2,3).

$$R_3 = \begin{Bmatrix} \alpha * x_{i,1}, \alpha * x_{i,2}, \ldots, \alpha * x_{i,\frac{d}{2}}; \\ \alpha * x_{j,\frac{d}{2}+1}, \alpha * x_{j,d-\frac{1}{2}+2}, \ldots, \alpha * x_{j,1} \end{Bmatrix} \tag{2}$$

$$R_4 = \begin{Bmatrix} \alpha * x_{j,1}, \alpha * x_{j,2}, \ldots, \alpha * x_{j,\frac{d}{2}}; \\ \alpha * x_{i,\frac{d}{2}+1}, \alpha * x_{i,d-\frac{1}{2}+2}, \ldots, \alpha * x_{i,1} \end{Bmatrix} \tag{3}$$

Where, α is a random variable between 0 and 1. Further, for more reactants following procedure is followed.

$$R_5 = \begin{Bmatrix} \alpha * x_{i,1}, \alpha * x_{i,2}, \ldots, \alpha * x_{i,\frac{2d}{3}}; \\ \alpha * x_{j,2d/3+1}, \alpha * x_{j,2(d-1)/3+2}, \ldots, \alpha * x_{j,1} \end{Bmatrix} \tag{4}$$

$$R_6 = \begin{Bmatrix} \alpha * x_{i,1}, \alpha * x_{i,2}, \ldots, \alpha * x_{i,\frac{d}{3}}; \ \alpha * x_{j,d/3+1}, \alpha * x_{j,2(d-1)/3+2}, \\ \alpha * x_{i,\frac{2d}{3}}, \ldots, \alpha * x_{j,1} \end{Bmatrix} \tag{5}$$

$$R_7 = \left\{ \alpha * x_{i,1}, \alpha * x_{i,2}, \ldots, \alpha * x_{i,\frac{d}{3}}; \ \alpha * x_{j,d/3+1},, \ldots, \alpha * x_{j,1} \right\} \tag{6}$$

$$R_8 = \left\{ \alpha * x_{j,1}, \alpha * x_{j,2}, \ldots, \alpha * x_{j,\frac{d}{3}}; \ \alpha * x_{i,\frac{d}{3}}, \quad \ldots, \alpha * x_{i,1} \right\} \tag{7}$$

$$R_9 = \left\{ \alpha * x_{j,1}, \alpha * x_{j,2}, \ldots, \alpha * x_{j,\frac{d}{3}}; \ \alpha * x_{i,\frac{d}{3}+1}, \alpha * x_{i,\frac{2(d-1)}{3}}, \alpha * x_{j,2d/3+1} \ldots, \alpha * x_{i,1} \right\} \tag{8}$$

$$R_{10} = \left\{ \alpha * x_{j,1}, \alpha * x_{j,2}, \ldots, \alpha * x_{j,\frac{2d}{3}}; \ \alpha * x_{i,\frac{2(d-1)}{3}+1}, \quad \alpha * x_{i,\frac{2d}{3}+2}, \ldots, \alpha * x_{i,1} \right\} \tag{9}$$

Step 3: Objective function calculation

After initial cluster center selection, next step is to compute the objective function. The values of objective function are computed using Equation (1) and data objects are assigned to the respective clusters using these values.

Step 4: Fitness function evaluation

In this step, the fitness of newly generated reactants is calculated using Equation (10). If the fitness value of newly created reactant is better, it is included in reactant pool otherwise it is discarded.

$$\text{Fitness}\,(R_p) = \sum_{j \in 1}^{K} \frac{SSE(R_p)}{\sum_{j=1}^{K} SSE(R_p)} \tag{10}$$

The SSE is sum of squared error and R_p denotes the reactants in pool.

Step 5: Crossover operator

In-order to generate new reactant the single point crossover operator is introduced in this work. The basic step to perform crossover operation are given below.

Step 5.1: Select a crossover point (C_p) in range [$1 < C_p < d$]. Here, *d* denotes the length of the reactant and C_p is crossover point.

Step 5.2: Copy the left portion of P_1 reactant into C_1 and left portion of P_2 into C_2. The P_1 and P_2 are parent reactants and C_1 and C_2 are child or newly generated reactants Fig. 1.

Step 5.3: Put elements into C_1 from P_2 in the same order as they appear but not already present in left portion.

Step 5.4: Repeat the procedure for C_2.

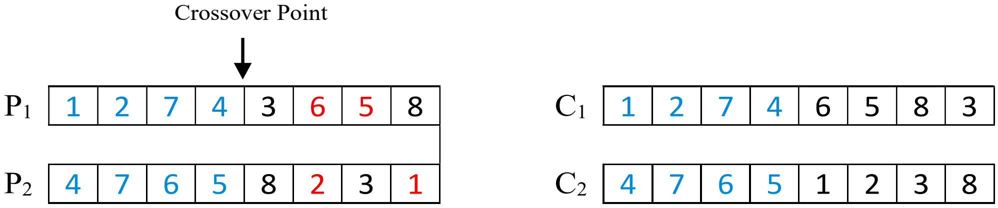

Fig. 1. Single point crossover

Step 6: Mutation operator

After crossover operation, each reactant undergoes mutation operation. In mutation operation, the positions of elements get changed either sipping its values or exchanging the columns.

In this work, a polynomial mutation operation is performed to mutate the reactants. The mutations steps are listed below.

Step 6.1: Calculate the perturbation factor using Equation (11).

$$\delta = \begin{cases} (2r)^{\frac{1}{q+1}} - 1 & if\ r \leq 0.5 \\ 1 - [(2(1-r)]^{\frac{1}{q+1}} - 1 & if\ r \geq 0.5 \end{cases} \quad (11)$$

Where r is random variable between 0 and 1, q is an exponent (user defined parameter).

Step 6.2: Obtain the mutated using Equation (12).

$$R_{mutated} = R_{original} + \delta \times \Delta \quad (12)$$

Where, Δ is maximum perturbation allowed between original and mutated values.

Step 7: Check local optima

If local optima situation arises apply “neighborhood strategy”, Else “next step”.

Step 8: Checking termination criteria

If termination criteria achieved “stop”, Else “repeat” steps 3 to 8.

Table 1. Algorithmic steps of the Hybrid ACRO for clustering process.

Hybrid artificial chemical reaction optimization algorithm (HACRO) for cluster analysis	
Step 1:	Initialize user-defined parameters, number of chemical reactants i.e. population and number of clusters (K) etc.
Step 2:	Generate the subpopulation i.e. reactant from the initial population using Equations (2-9).
Step 3:	Calculate the objective function via Equation (1).
Step 4:	Investigate the fitness of newly generated reactants with Equation (10).
Step 5:	Implement crossover and mutation operator to generate new reactant using Equations (11-12).
Step 6:	If, fitness value of the newly created reactants is better than previous reactants, then replace ‘Else’ consider current reactant and perform update.
Step 7:	Check local optima, if occurs apply “neighborhood strategy”, Else “next step”.
Step 8:	If termination criteria are not met, repeat the steps 3-8, otherwise find optimal cluster centers.
Step 9:	Optimal result

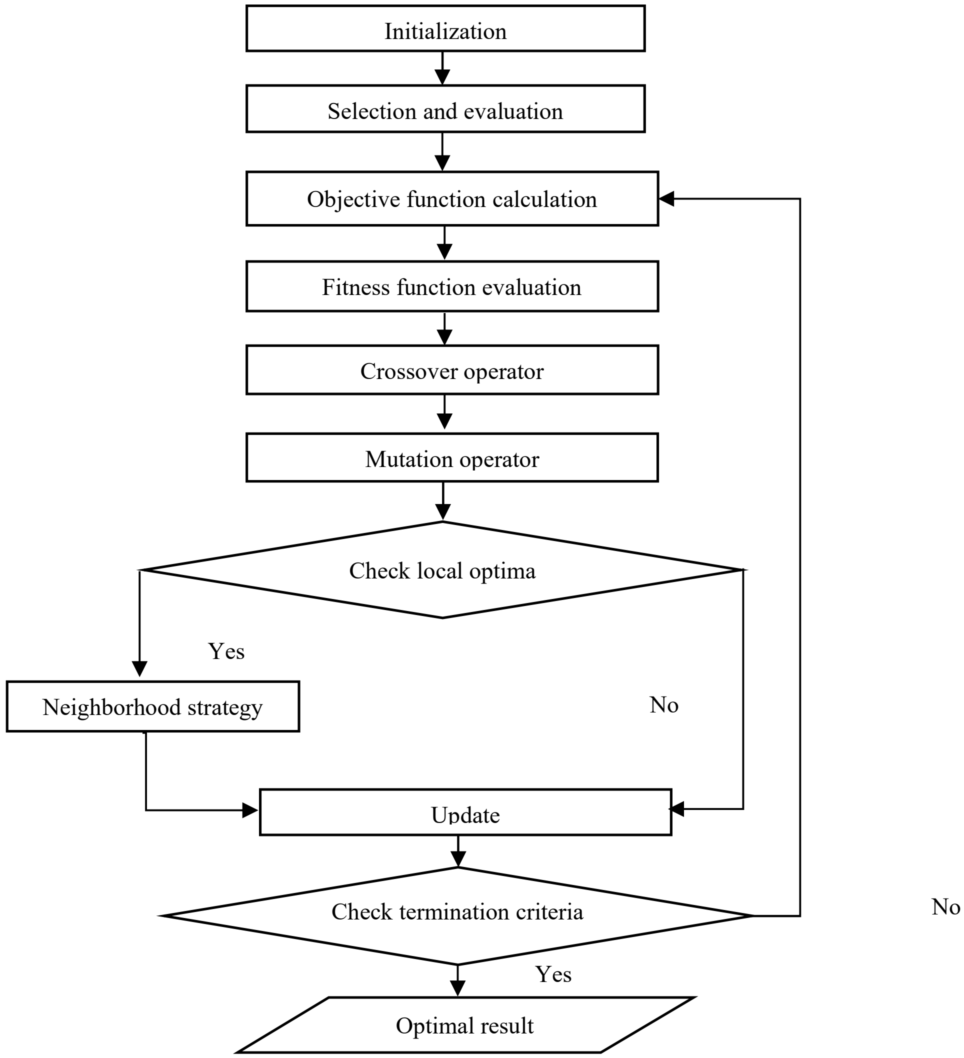

Fig. 2 Hybrid ACRO algorithm

4. Simulation Results

The simulation results of Hybrid ACRO algorithm in clustering space are demonstrated in this section. The proposed algorithm is implemented in MATLAB simulator installed on windows operating system. The efficacy of the developed algorithm is examined on seven datasets including two artificial and five real-life datasets. The detailed description of datasets is provided in Table 2. Further, to examine the performance of proposed algorithm, three performance measures are taken into consideration in the present work.

4.1 *Performance measures*

4.1.1 *Intra-cluster distance*

Intra cluster distance is the sum of distance between the data instances present in one cluster to its corresponding cluster centre. The results are restrained in terms of average values of solutions.

4.1.2 *Standard deviation*

Standard deviation provides the information about the dispersion of data instances present in cluster from its centroid. The minimum value of standard deviation indicates that the data instances are close to its centroid, while large value indicates that the data are far from its centre points.

4.1.3 *F-measure*

The F-measure is a measure of a test's accuracy and is defined as the weighted harmonic mean of the precision and recall of the test.

Table 2. Description of the datasets.

Datasets	K	D	N	Description
Art1	3	2	300 (100, 100, 100)	Artificial data 1
Art2	3	3	300(100, 100, 100)	Artificial data 2
Iris	3	4	150 (50, 50, 50)	Fisher's iris data
CMC	3	9	1,473 (629, 334, 510)	Contraceptive method choice
Wine	3	13	178 (59, 71, 48)	Wine data
Glass	6	9	214 (70, 76, 17, 13, 9, 29)	Glass identification data
Cancer	2	9	638 (444, 239)	Cancer data

Table 3 illustrates the simulated results of Hybrid ACRO, K-means, ACRO, GA, PSO, ICSO and ACO clustering algorithms tested on seven datasets. From experimental outcomes, it is observed that proposed algorithm performs well in clustering field. However, the performance of ACRO and HACRO is nearly similar in case of wine data set. Further, to validate the significance of HACRO algorithm the Friedman statistical test is also conducted in this work.

Table 3. Performance comparison of HACRO with other well-known clustering algorithms.

Datasets	Parameters	Clustering Algorithms						
		K-means	GA	ACO	PSO	ICSO	ACRO	HACRO
Art1	Intra Cluster Distance	161.12	157.01	154.29	153.45	158.17	143.36	143.21
	SD	7.625	0.192	4.712	6.437	0.145	5.243	5.070
	F-Measure	0.94	0.95	0.99	0.96	1	1	0.99
Art2	Intra Cluster Distance	768.38	743.10	766.15	759.82	745.74	752.26	745.01
	SD	6.837	6.621	6.845	7.614	0.247	4.214	4.227
	F-Measure	0.89	0.87	0.93	0.91	0.99	0.97	0.94
Iris	Intra Cluster Distance	114.00	125.00	98.40	98.70	97.08	96.70	95.50
	SD	15.30	14.60	0.43	0.47	0.156	0.20	0.20
	F-Measure	0.78	0.78	0.78	0.78	0.783	0.79	0.78
CMC	Intra Cluster Distance	5910.00	5760.00	5830.00	5850.00	5756.31	5746.00	5730.00
	SD	54.20	50.40	44.30	48.90	35.04	36.41	33.40
	F-Measure	0.34	0.34	0.33	0.33	0.341	0.340	0.340
Wine	Intra Cluster Distance	18100.00	16500.00	16500.00	16400.00	16357.89	16334.00	16300.00
	SD	785.00	78.40	53.70	91.30	40.73	34.60	33.50

	F-Measure	0.52	0.51	0.52	0.52	0.524	0.53	0.51
Glass	Intra Cluster Distance	247.00	239.00	281.00	279.00	256.01	266.00	231.00
	SD	18.30	74.50	6.58	8.59	8.74	8.11	7.190
	F-Measure	0.43	0.42	0.40	0.41	0.46	0.43	0.43
Cancer	Intra Cluster Distance	3250.00	3250.00	3180.00	3120.00	3036.49	3063.34	3048.00
	SD	257.00	230.00	93.50	107.00	43.56	71.22	70.60
	F-Measure	0.83	0.83	0.83	0.83	0.834	0.84	0.83

Fig. 3 and 5 demonstrate the dispersion of the data objects in wine and glass datasets respectively. These datasets are downloaded from UCI repository. The Fig. 4 show the clustering of the wine dataset objects into different clusters using HACRO algorithm. The data objects of wine dataset are divided into three clusters i.e. wine type 1, type 2 and type 3. From graphical representation it is seen that objects of wine type 1 cluster are linearly separable form wine type 2 and type 3. The Fig. 6 show the clustering of the glass dataset objects into different clusters using HACRO algorithm. The glass dataset constitutes six number of clusters building float, building non float, vehicle float, vehicle non float, containers, tableware and headlamps.

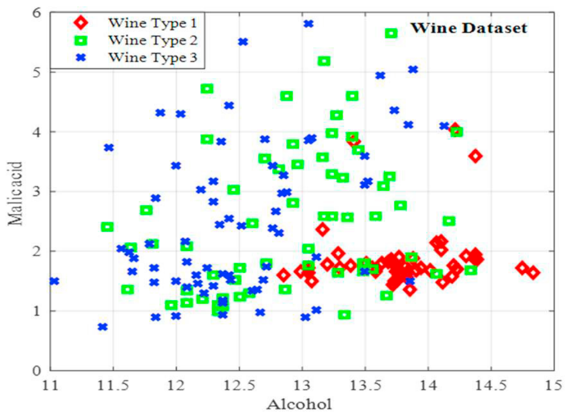

Fig. 3. Dispersion of data objects in Wine Dataset

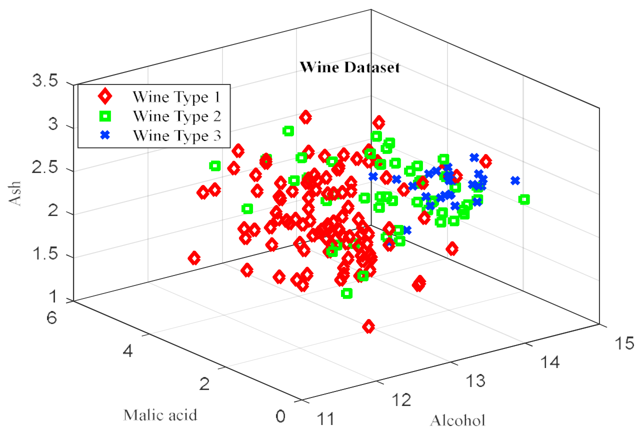

Fig. 4. Clustering results of HACRO on Wine Dataset

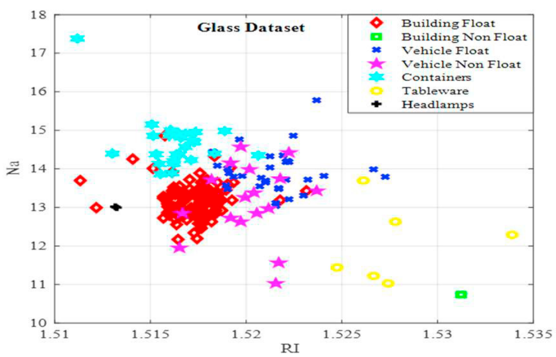

Fig. 5. Dispersion of data objects in Glass Dataset

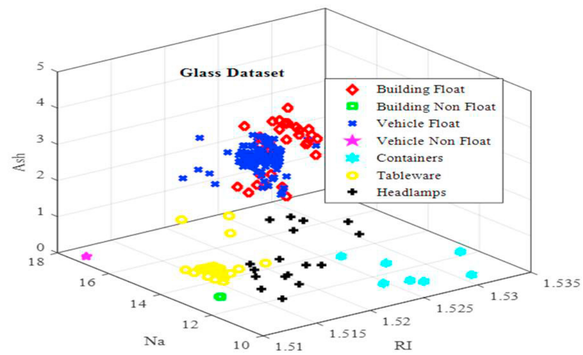

Fig. 6. Clustering results of HACRO on Glass Dataset

Table 4. Friedman statistical test on avg. intra-cluster distance constraint of compared algorithms.

Datasets	Clustering Algorithms						
	K-means	GA	ACO	PSO	ICSO	ACRO	HACRO
Art1	7	5	4	3	6	2	1
Art2	7	1	6	5	3	4	2
Iris	6	7	4	5	3	2	1
CMC	7	4	5	6	3	2	1
Wine	7	5.5	5.5	4	3	2	1
Glass	3	2	7	6	4	5	1
Cancer	6.5	6.5	5	4	1	3	2
Sum	43.5	31	36.5	33	23	20	9
Rank	6.21	4.43	5.21	4.71	3.29	2.86	1.29

Number of observations: 49	Number of problems: 07	Number of algorithms: 7
Sum of squares of rank sums: 6284.5	Correction factor: 784	Friedman test statistic: 24.5076
Degree of freedom: 6	p-value: 0.000421	Critical value: 12.5915

Table 4 illustrates the results of Friedman statistical test conducted on average intra-cluster distance constraint. To perform statistical test two hypotheses (H_0 and H_1) are assumed in this work. According to the first hypothesis, there is a significant difference among performances of compared algorithms. While, according to second hypothesis there is insignificant difference among performances of compared algorithms. The statistical test is conducted at confidence level 0.05. The value of Friedman statistical test is 24.5076. Whereas, the critical value is 12.5915 and p-value is 0.000421. From results, the second hypothesis is rejected that validated the significance of proposed algorithm.

5. Conclusion and Future Scope

In present work, a hybrid version of the artificial chemical reaction optimization algorithm is proposed to optimize clustering problems. The traditional ACRO algorithm suffers from slower convergence speed due to its large comparative calculation. Hence, to improve the convergence speed and make a balance among exploration and exploitation processes the crossover and mutation operator of genetic algorithm are inherited in the proposed algorithm. The hybridization results enhanced coordination among exploration and exploitation processes. The efficiency of developed algorithm is examined on seven datasets and compared with K-means, GA, ACO, PSO, ICSO and ACRO clustering algorithms. The HACRO algorithm achieves significant results in comparison to other algorithms. The significance of proposed algorithm is also confirmed via Friedman statistical test. In future, the proposed algorithm will be implemented in e-healthcare application for data analysis.

References

[1] Lo, Kenneth, Ryan Remy Brinkman, and Raphael Gottardo. "Automated gating of flow cytometry data via robust model-based clustering." *Cytometry Part A: the journal of the International Society for Analytical Cytology* 73.4 (2008): 321-332.

[2] Kavakiotis, Ioannis, Olga Tsave, Athanasios Salifoglou, Nicos Maglaveras, Ioannis Vlahavas, and Ioanna Chouvarda. "Machine learning and data mining methods in diabetes research." *Computational and structural biotechnology journal* 15 (2017): 104-116.

[3] Gilbert, Edwin Prem Kumar, Baskaran Kaliaperumal, Elijah Blessing Rajsingh, and M. Lydia. "Trust based data prediction, aggregation and reconstruction using compressed sensing for clustered wireless sensor networks." *Computers & Electrical Engineering* 72 (2018): 894-909.

[4] Ma, Zhanyu, Jiyang Xie, Hailong Li, Qie Sun, Zhongwei Si, Jianhua Zhang, and Jun Guo. "The role of data analysis in the development of intelligent energy networks." *IEEE Network*31.5 (2017): 88-95.

[5] Miah, Shah Jahan, Huy Quan Vu, John Gammack, and Michael McGrath. "A big data analytics method for tourist behaviour analysis." *Information & Management* 54.6 (2017): 771-785.

[6] Kamilaris, Andreas, Andreas Kartakoullis, and Francesc X. Prenafeta-Boldú. "A review on the practice of big data analysis in agriculture." *Computers and Electronics in Agriculture* 143 (2017): 23-37.

[7] Hatamlou, Abdolreza. "Black hole: A new heuristic optimization approach for data clustering." *Information sciences* 222 (2013): 175-184.
[8] Alatas, Bilal. "ACROA: artificial chemical reaction optimization algorithm for global optimization." *Expert Systems with Applications* 38.10 (2011): 13170-13180.
[9] Alatas, Bilal. "A novel chemistry-based metaheuristic optimization method for mining of classification rules." *Expert Systems with Applications* 39.12 (2012): 11080-11088.
[10] Zhan, Zhi-Hui, Jun Zhang, Yun Li, and Henry Shu-Hung Chung. "Adaptive particle swarm optimization. "*IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 39.6 (2009): 1362-1381.
[11] Cura, Tunchan. "A particle swarm optimization approach to clustering." *Expert Systems with Applications* 39.1 (2012): 1582-1588.
[12] Prakash, J., and P. K. Singh. "Hybrid Gbest-guided Artificial Bee Colony for hard partitional clustering." *International Journal of System Assurance Engineering and Management*9.4 (2018): 911-928.
[13] Das, Pranesh, Dushmanta Kumar Das, and Shouvik Dey. "A modified Bee Colony Optimization (MBCO) and its hybridization with k-means for an application to data clustering." *Applied Soft Computing* 70 (2018): 590-603.
[14] Kang, Qi, ShiYao Liu, MengChu Zhou, and SiSi Li. "A weight-incorporated similarity-based clustering ensemble method based on swarm intelligence." *Knowledge-Based Systems* 104 (2016): 156-164.
[15] Kumar, Yugal, and Gadadhar Sahoo. "An Improved Cat Swarm Optimization Algorithm Based on Opposition-Based Learning and Cauchy Operator for Clustering." *JIPS* 13.4 (2017): 1000-1013.
[16] Senthilnath, J., Sushant Kulkarni, Jón Atli Benediktsson, and Xin-She Yang. "A novel approach for multispectral satellite image classification based on the bat algorithm." *IEEE Geoscience and Remote Sensing Letters* 13.4 (2016): 599-603.
[17] Zabihi, Farzaneh, and Babak Nasiri. "A Novel History-driven Artificial Bee Colony Algorithm for Data Clustering." *Applied Soft Computing* 71 (2018): 226-241.
[18] Mageshkumar, C., S. Karthik, and V. P. Arunachalam. "Hybrid metaheuristic algorithm for improving the efficiency of data clustering." *Cluster Computing*: 1-8.
[19] Kowalski, Piotr A., Szymon Łukasik, Małgorzata Charytanowicz, and Piotr Kulczycki. "Optimizing clustering with cuttlefish algorithm." *Conference on Information Technology, Systems Research and Computational Physics*. Springer, Cham, 2018.
[20] Bijari, Kayvan, Hadi Zare, Hadi Veisi, and Hossein Bobarshad. "Memory-enriched big bang–big crunch optimization algorithm for data clustering." *Neural Computing and Applications* 29.6 (2018): 111-121.
[21] Zhou, Yongquan, Haizhou Wu, Qifang Luo, and Mohamed Abdel-Baset. "Automatic data clustering using nature-inspired symbiotic organism search algorithm." *Knowledge-Based Systems* 163 (2019): 546-557.