International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 1 | ISSN : 2456-3307

An Efficient Clustering Scheme for Cloud Computing Problems Using Parameter Improved Particle Swarm Optimization (PIPSO) Technique

Baalamurugan K. M.*[1], Dr. S. Vijay Bhanu[2]

*[1]Research Scholar, Department of Computer Science and Engineering, Annamalai University, Chidambaram, Tamilnadu, India

[2]Assistant Professor, Department of Computer Science and Engineering, Annamalai University, Chidambaram, Tamilnadu, India

ABSTRACT

Data clustering partitions the information into helpful classes or groups with no earlier learning. This is a fundamental method in the field of computer data mining and it has turned into an essential element in many other engineering areas including cloud computing. This paper purports a novel clustering technique based on the application of Parameter Improved Particle Swarm Optimization (PIPSO) algorithm. It is an optimization approach for data clustering problem, in which a swarm of particles (candidate solutions) moves to converge a specific positions as final cluster centres by minimizing the fitness function. The proposed method is compared with the common clustering methods such as k-means clustering algorithm, data clustering using particle swarm optimization algorithm, ant colony optimization based algorithm, and proposed clustering method using parameter improved particle swarm optimization algorithm, MATLAB simulation results evidence that the proposed technique gives better results when compared to other existing techniques. The proposed data clustering method can be employed to manipulate vast data sets with different cluster sizes, multi dimensional and densities.

Keywords: Cloud Computing, Data clustering, PIPSO, ACO, PSO, Optimization based clustering.

I. INTRODUCTION

Clustering is characterized as unsupervised categorization of data patterns into purposeful cluster classes [1]. In past two decenniums, numerous contemplations have concentrated around clustering problem on viewpoint of hypothetical and pragmatically. The clustering problems have been tending in distinct fields, for example, analysis of data, patterns recognition, digital signal processing, biological science and social economics particularly on statistical studies. Hence the analyses regarding novel clustering methods are an essential component in the area of researches admitting machine learning, clouding computing, natural science, data mining and market economy.

In past two decenniums, many distinct clustering techniques have been suggested, for example, data partitioning hierarchical, density-based, grid-based and model-based [2]. Data partitioning technique develops distinctive allotments of data set in view of some measurements. In hard data partitional clustering, every pattern has a place in single cluster. Fuzzy clustering, stretch out this concept with every pattern may have place in all clusters with a fuzzy degree of membership. Aside from the above procedures, a precise fundamental relationship between spectral clustering and kernel k-means has been discussed, and by stretching the formula of generalized kernel k-means objective function, the spectral clustering has been arrived [3].

K-means clustering [4] is the most famous method on account of its effortlessness, productivity and ease of calculation. Nevertheless, the conventional solution methodologies for data clustering particularly standard k-means technique is more susceptibility to initial conditions and it is easily to be tuned to the local optimum solutions, due to the non-linear and non-convex nature of certain objective functions of data clustering. Computing the optimal solutions to an objective function of data clustering has become hard non-linear programming (NP) problem owing to the large number of data sets and its dimensions. A few variations to standard k-mean technique gives a quick and local search sub-problems to handle such NP hard problems [5] [6]. Because of the significance of data clustering techniques in many engineering areas, the intelligent optimization algorithms such as Particle Swarm Optimization (PSO) [7] [8], Bacterial Foraging (BF) [9] [10] [11], Ant Colony Optimization (ACO) [12] [13] have been employed to identify the global optimal solution to the data clustering problems [14] [15]. While finding the solutions to the data clustering problems using these intelligent optimization techniques, the implementations start with an initialization of random populations from search space and investigate towards the global optimal solution until stopping convergence criteria achieved. As view point of data clustering, it is an optimization problem; hence a good novel artificial intelligent algorithm can be employed to find the global optimal solution for the data clustering problem.

In this paper, a novel naturally enlivened technique called Parameter Improved Particle Swarm Optimization (PIPSO) algorithm is presented for solving the data clustering problems. The PSO technique is basically exhorted from the biological movement of individuals in a bird flocking or fish schooling behaviors. In 1995, the concept of PSO algorithm to solve optimization problems was discovered by Eberhart and Kennedy [16]. By extending the conception of PSO, PIPSO algorithm has been developed by updating PSO parameters such as inertia weight, social and cognitive agents at each iteration. It is proven that the PIPSO technique has an ability to proficiently tackle an extensive variety of problems and outflanks other existing naturally inspired artificial techniques. This improved version of PSO algorithm can meliorate both local and global solutions and can take minimum number of iterations to convergence for global optimal solution of a problem. Rather than the rapid local search, PIPSO technique is a global optimization method, which gives another perspective to tackle the hard NP clustering problems. In the interim, it is a fresh and novel manipulation of PIPSO.

This paper is further sorted as follows: in section 2, the theoretical concept of clustering problem is portrayed. A PIPSO intelligent method is exhibited thoroughly in Section 3. Section 4 elucidates the philosophy of the proposed PIPSO technique for clustering problem. Section 5 demonstrates the method for validation of clusters. Matlab simulation results are delineated in section 6 and finally conclusion is devoted in Sections 7.

II. PROBLEM FORMULATION

Data clustering partitions the information into helpful classes or groups with no earlier learning. The basic data clustering problem can be generally begun as follows.

For a given sample of data set P = {p1, p2, ...,p_m}, find a segmentation of data into N clusters G1,G2,...,G_N which fulfils the below conditions:

$$\begin{cases} \cup_{i=1}^{N} = P; \\ G_i \cap G_j = 0; \quad i,j = 1,2,\ldots,N; \quad i \neq j \\ G_i = 0; \quad i,j = 1,2,\ldots,N; \end{cases} \tag{1}$$

In perspective of mathematical science, cluster G_i can be dictated by:

$$\begin{cases} G_i = \{x_j \| x_j - z_i \| \le 0 \| x_j - z_k \|, \quad x_j \in P\}, \\ \qquad\qquad k \ne i, k = 1,2,\ldots,N; \\ z_i = \frac{1}{|G_i|} \sum_{x_j \in G_i} x_j, \quad i = 1,2,\ldots,k; \end{cases} \tag{2}$$

where || . || refers the distance between any two data points in a sample data set. Zi refers the center of the cluster Gi, the cluster center Zi is computed by taking the mean of all data points in the cluster Gi. Clustering constraints should be must be followed. Regularly utilized constraint in data clustering problem is Sum of Squared Error (SE)

$$SE = \sum_{i=1}^{N} \sum \|x_j - z_i\|^2 \tag{3}$$

For every data in a sample, the error is the distance between the data and its closest cluster centers. The fundamental intention of the data clustering problem is to acquire the predefined number of cluster groupings or partitions with very least possible sum of squared errors. Moreover, the clustering problem can be treated as a procedure to find optimal N cluster centers $Z_1,Z_2,\ldots,Z_N$; and it should minimizes the objective function as sum of squared distance between every data P = $\{p_1,p_2,\ldots,p_m\}$ and its closest cluster center. Therefore, the clustering problem is conceived as a mathematical optimization problem with objective function as SE.

III. PARAMETER IMPROVED PARTICLE SWARM OPTIMIZATION

The optimal cloud storage problem is one of the constraint marriage problems or the assignment problems. To better tackle with the optimal cloud storage problem of the distributed data centres, the PIPSO algorithm is mainly inspired by the PSO algorithm, which has a number of the successfully existing methods and has been used in the fields of different areas. To utilize the existing method of the continuous PIPSO algorithm, the mapping function from the continuous space to the discrete space can be simply achieved by the sorting operator of continuous values. In the cloud storage optimization problem, the particle in the PIPSO algorithm consists of the mapping order between the node set S and another node set S'. The parameters in each particle can calculate the objective fitness and should subject to the constraints. Main steps of the PIPSO algorithm can be roughly concluded as the following steps.

PSO algorithm explores for global optimum solution in a given N-dimensional problem by collaborating with particles in a swarm. Each particle or individual in a swarm has the characteristics of position and velocity. Mathematically, position 'x' and velocity 'v' of a particle can be represented as feasible solution of the problem and its step length for next iteration respectively. For a given N-dimensional mathematical problem and 'm' number of particles, the position and velocity of the ith particle is denoted as $x_i = [x_{i1}, x_{i2}, \ldots, x_{iN}]$ and $v_i = [v_{i1}, v_{i2}, \ldots, v_{iN}]$ respectively. At the end of each iteration, the best position of ith particle compared with previous iterations is registered as local best solution and denoted by $p_i = [p_{i1}, p_{i2}, \ldots, p_{iN}]$. The overall best position among all particles in a population is considered as global best position $p_g = [p_{g1}, p_{g2},\ldots, p_{gN}]$. The velocity and position update of ith particle for next iteration can be computed by using particle's current velocity and its distance between local best position and global best position, the mathematical expression for ith particle's velocity and position updates are as follows,

$$v_{id}^{t+1} = \omega . v_{id}^{t} + \varphi_1 (p_{gd} - x_{id}^{t}) + \varphi_2 (p_{id} - x_{id}^{t}) \tag{4}$$

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \tag{5}$$

where t is the iteration index; ω is the inertia weight; d is the dimension index; $\varphi_1=c_1r_1$ and $\varphi_2=c_2r_2$; c_1 and c_2 are two positive acceleration coefficients called social and cognitive agents respectively; both r_1 and r_2 are random numbers uniformly distributed in between 0 and 1.

By extending the conception of PSO, Parameter Improved Particle Swarm Optimization (PIPSO) algorithm has been developed by updating PSO parameters such as inertia weight, social and cognitive agents at each iteration. This improved version of PSO algorithm can meliorate both local and global solutions and can take minimum number of iterations to convergence for global optimal solution of a problem. The parameters ω, c_1 and c_2 in (4) can be calculated from the following equations,

$$\omega_1 = \omega_{\max} - \frac{\omega_{\max} - \omega_{\min}}{t_{\max}} \times t \tag{6}$$

$$\omega \ = \omega_{\min} + \omega_1 \times r_3 \tag{7}$$

$$c_1 = c_{1\max} - \frac{c_{1\max} - c_{1\min}}{t_{\max}} \times t, \tag{8}$$

$$c_2 = c_{2\max} - \frac{c_{2\max} - c_{2\min}}{t_{\max}} \times t, \tag{9}$$

where $\omega_{\min}$ and $\omega_{\max}$ are minimum and maximum weights; $c_{1\min}$ and $c_{1\max}$ are minimum and maximum cognitive factors; $c_{2\min}$ and $c_{2\max}$ are minimum and final social factors; $t_{\max}$ is maximum iteration; r_3 are random numbers uniformly distributed in between 0 and 1.

IV. IMPLEMENTATION

As we have just mentioned in Section 2, clustering tasks can be considered as optimization problems. Firstly, the fitness function should be specified. Here we choose SE in (3) to be the required function J in PIPSO-C. The step by step procedures for the implementation of proposed clustering algorithm are as follows,

Let m be the number of populations and N be the dimensions of the problem.

Step 1: Read system data.

Step 2: Choose appropriate values for all PIPSO parameters and set iteration counter as t = 1.

Step 3: Randomly initialize feasible solution for m particles in matrix form as $x_i = [x_{i1}\ x_{i2}]$, i = 1 to m.

Step 4: For each particle i, consider data centre.

Step 5: Calculate fitness value (objective function SE) to ith particle. Repeat step 4 to 5 for all remaining particles.

Step 6: Determine the local best position p_i of ith particle and global best position p_g among all particles based on minimum fitness value.

Step 7: Update PIPSO parameters using (6)-(9).

Step 8: Do velocity update for all particles using (4) and modify the particles position using (5).

Step 9: Check whether the updated particles position satisfies the constraints; if any particle violates the constraints, randomly assign the feasible solution to the violated particle as like step 3; otherwise go to next step.

Step 10: If the convergence criteria are satisfied (t equals $t_{\max}$), output the optimal data centre; otherwise set t = t + 1, and repeat the step 4 to 10.

To validate the efficiency and potential of the proposed technique, two different types of databases are considered for analysis and the simulation results are presented in next section. Figure 1 show the flowchart for PIPSO algorithm application to data clustering problem.

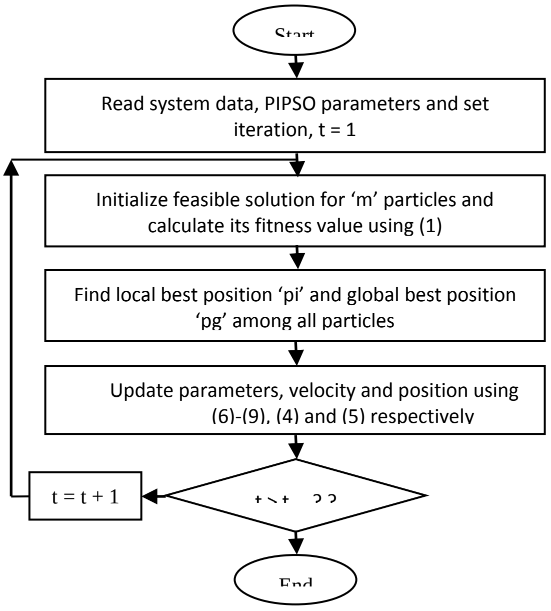
Start
Read system data, PIPSO parameters and set iteration, t = 1
Initialize feasible solution for 'm' particles and calculate its fitness value using (1)
Find local best position 'pi' and global best position 'pg' among all particles
Update parameters, velocity and position using (6)-(9), (4) and (5) respectively
t = t + 1
End

Figure1. Flow Chart for PIPSO

A. Cluster validity

Two sorts of cluster validity methods are picked in this article. The first depends on external criteria, which are utilized to assess the consequences of the purposed PIPSO calculation in view of the correlation with the pre-indicated class name data of the data collection. The second one depends on internal criteria, which we assess the clustering outcomes of the PIPSO calculation execution with no earlier information of data collections. Two external validity measures Rand and Jaccard and also two internal validity measures, Beta (Pal et al. 2000) and Distance index are used for execution assessment of the PIPSO-C calculation and its correlation techniques.

B. Methods for Analysis

For showing the prevalence of the purposed PIPSO calculation, we choose some past clustering procedures for calculation examinations. Firstly we pick the k-means calculation [17] as a strategy to be looked at in light of the fact that it is the most celebrated customary clustering procedure. The k-means calculation is a partition based grouping approach and has been generally employed for quite a long time of years. Also, as a worldwide optimization based method, the PIPSO calculation will be compared with the ACO based clustering method [17], clustering method based on PSO [18] [19] and clustering method based on BF [20]. Ant colony optimization (ACO) [21] [22] was intended to imitate ants' conduct of placing pheromone on the earth while travelling to take care of optimization issues [17] exhibited a case of ACO for clustering which restore an unequivocal dividing of information by a programmed procedure. The ACO calculation impersonates the systems by picking arrangements in view of pheromones and refreshing pheromones in view of the solution quality. Particle swarm optimization (PSO) [23] is a populace based calculation. It is a worldwide optimization strategy and reenacts fish tutoring conduct to accomplish a self-advancement framework. The clustering approach utilizing PSO can seek naturally the server farms of K bunches informational index by advancing the goal work. The Bacterial Foraging (BF) optimization method [24] is a novel random worldwide explore scheme established on the foraging characteristics of E. Coli bacteria living in the human intestine. The thoughts from bacterial foraging can be used to solve engineering optimization problems by the following stages, such as, chemo taxis, reproduction, and elimination and dispersal. In Section 7, we present a progression of investigations to portray strategy correlations between purposed PIPSO, k-means, ACO based, PSO based and BF based clustering techniques.

In order to demonstrate the efficiency and the performance of handling with the cloud storage problem by the PIPSO algorithms, numerical results mainly concentrate on the cloud storage issue of test data’s. Due to the randomness of the discrete PSO algorithm, the suboptimal solution of the cloud storage problem possibly is not equal to each other, especially when the number of data centres is generally larger than 15. Therefore, the objective of this section is to discuss the optimal or suboptimal strategy of the cloud storage problem under different objectives and different numbers of data centres, minimizing the smallest transmitting distance. Table 1 depicts the datasets used. In the case of the constraints in this problem, the new particle should be satisfied three constraints at each evolutionary process. The m-script has been developed and executed in MATLAB simulation software on Intel Pentium dual core personal computer with 2GB RAM.

This section investigates the simulation analysis of the purposed krill herd based clustering technique on different databases to illustrate its performance, feasibility and superiority over other existing techniques. The experimental studies are carried out on MATLAB simulation environment. Two sets of synthetic database with explicit data attributes [17] and five real databases [39] are analyzed to validate the purposed krill herd based clustering technique.

TABLE 1

SUMMARISATION OF VARIOUS DATASETS [24]

Dataset	Samples	Features/ dimensions	Clusters
2D-4C	1,572	2	4
10D-4C	1,289	10	4
Iris	150	4	3
Wine	178	13	3
Glass	214	9	6
Zoo	101	16	7
Ionosphere	351	34	2

C. Synthetic Database

The two synthetic databases used for analyze of the proposed clustering technique complies x dimensional normal distribution $N(\vec{\mu},\vec{\sigma})$ with mean vector $\vec{\mu}$ & standard deviation vector $\vec{\sigma}$ and their data are placed into y distinctive cluster groups. Each cluster group has a sample size of s_i; i ∈ [1, y] and its cluster data are randomly calculated using normal distribution over the range of s ∈ [50, 550], $\mu_i \in [-10, 10]$ and $\sigma_i \in [0, 5]$. Therefore the clusters in each database are with various size and density. The first set of synthetic database used for analysis is named as 2D-4C, since it has 2 dimensional data space disposed in the range of [−20, 20] & [−12, 8] and it contains 4 cluster groups with sample size of 528, 348, 272 and 424 each. The second database is 10D-4C and it has 10 dimensional data space and 4 cluster groups with overall sample size of 1,289.

D. Real Database

The other sets of real database taken from the UCI are widely well known data sets that can be effortlessly seen in the research area of pattern recognition and data mining. The wine database is the consequences of a chemical investigation of wine developed in same area of Italy but obtained from 3 unique varieties. It comprises of 3 cluster groups with sample size of 59, 71 and 48 respectively. The real iris database comprises of 3 distinct cluster groups with each cluster has sample size of 50. Each and every cluster in the database denotes to a subclass of iris plant and can be handled as a cluster group in the simulation studies. Each data in a cluster has 4 properties including petal length and width, sepal length and width. Glass database contains 214 samples depicting 6 clusters of glass in light of 9 properties. Zoo database has animal sample size of 101 with 16 Boolean assessed properties that are characterized into 7 cluster groups. The samples in these real databases are distributed over a wide dimensional space.

V. RESULTS AND ANALYSIS

For each and every database, the MATLAB simulations are carried out for 30 times and their mean and standard deviation of various performance measures are calculated and depicted in table 2.

The Euclidean distance is selected to determine the distance between the data samples. The ranking position for each technique is assigned based on its performance indices. The table 2 illustrates the data clustering results acquired from the k-means, PSO, ACO, BF and the purposed PIPSO-C techniques for both real and synthetic databases. As from the dataset attributes depicted in table 1 and their data clustering results described in table 2, roughly few inferences are discovered as below,

(1) The proficiency of the proposed PIPSO-C technique is better for all databases except 2D-4C for the indices Rand and Jaccard. The PSO provides better performance for 2D-4C database when compared with PIPSO-C. However, the overall performance shows the proposed technique can achieve better optimal solution when compared with the existing methods by overcoming the drawbacks. As an efficient optimization technique, the PISPO-C quickly manifested food convergence properties over other methods. Figure 2 shows the convergence property of proposed and existing techniques.

(2) The inference shows that the proposed PIPSO-C is a very powerful technique for multi clustering databases such as glass and zoo and also provides

advantages over clustering with different data size and clusters.

(3) From the results the proposed PIPSO-C technique outperforms well for β index for majority of databases. The ACO technique shows poor results for β index. From the comparison it is clearly shown that the proposed PIPSO-C technique gives better results for databases which don't have any earlier knowledge about the attributes.

TABLE 2. THE AVERAGE VALUES OF VARIOUS PERFORMANCE INDICES BY THE K-MEANS, ACO, PSO, BF AND THE PURPOSED PIPSO-C TECHNIQUES

Database	Technique	Rand	Jaccard	β	Dis	Time (s)
2D-4C	k-means	0.8636	0.8021	11.1319	0.01079	0.23058
	ACO	0.9941	0.9778	12.565	0.00998	10.87969
	PSO	0.9916	0.9558	13.874	0.01012	10.19844
	BF	0.992	0.9702	13.249	0.01006	6.7922
	PIPSO-C	**0.997053**	**0.981134**	**13.25266**	**0.010005**	**5.687991**
10D-4C	k-means	0.8946	0.7203	2.264	0.0973	0.069018
	ACO	0.8763	0.6924	2.1989	0.09693	27.71563
	PSO	0.9239	0.76142	1.1102	0.096711	19.2719
	BF	0.9319	0.8187	2.2968	0.09202	18.28282
	PIPSO-C	**0.934551**	**0.820092**	**2.300039**	**0.061028**	**17.43687**
Iris	k-means	0.8737	0.6823	7.8405	0.02422	0.00625
	ACO	0.9195	0.7828	8.3579	0.02243	6.753125
	PSO	0.8254	0.6547	1.6159	0.03104	5.5256
	BF	0.9341	0.818	9.1295	0.02111	2.9344
	PIPSO-C	**0.936167**	**0.821171**	**9.130683**	**0.01544**	**2.730947**
Wine	k-means	0.717	0.4127	7.3745	0.02652	0.008235
	ACO	0.7307	0.4312	7.6108	0.02727	19.92031
	PSO	0.683959	0.424734	1.012184	0.030469	11.0644
	BF	0.7516	0.4494	7.9366	0.0259	7.86721
	PIPSO-C	**0.756522**	**0.454859**	**7.93928**	**0.023311**	**7.330549**
Glass	k-means	0.7047	0.2676	3.1188	0.03647	0.034375
	ACO	0.5409	0.1902	2.4245	0.04097	15.29531
	PSO	0.6353	0.2699	1.009839	0.037285	13.9375
	BF	0.7376	0.2765	3.5644	0.03171	11.62502
	PIPSO-C	**0.741284**	**0.278079**	**3.566945**	**0.02337**	**10.48621**
Zoo	k-means	0.7998	0.3758	4.1048	0.01821	0.01875
	ACO	0.8525	0.4768	4.6966	0.01757	44.23438
	PSO	0.8829	0.6867	1.02699	0.01691	17.6563
	BF	0.921	0.6977	5.9665	0.01538	11.38752
	PIPSO-C	**0.922071**	**0.703484**	**5.968584**	**0.013671**	**11.10658**
Ionosphere	k-means	0.5877	0.4323	1.3405	0.75862	0.046875
	ACO	0.5921	0.4261	1.3516	0.75771	45.1
	PSO	0.5398	0.5384	1.3114	0.76174	53.6571
	BF	0.5989	0.4439	1.3528	0.75413	36.39376
	PIPSO-C	**0.600189**	**0.540945**	**1.354569**	**0.745535**	**34.55852**

In the case of intra clustering distance guarantees constrict cluster groups with very minimal deflection from the cluster centers, while the inter clustering distance assures more gap between the distinct cluster groups. The proportion between the intra and inter clustering Euclidean distances (Dis) must be minimum for effective clustering of data items. The proposed PIPSO-C technique provide very minimum Dis index while computing the cluster groups when compared with the other techniques. But in the case of 2D-4C database PSO performs well when compared with other techniques.

TABLE 3. THE STANDARD DEVIATION OF THE PERFORMANCE INDICES BY THE K-MEANS, ACO, PSO, BF AND THE PURPOSED PIPSO-C TECHNIQUES

Database	Technique	Rand	Jaccard	β	Dis	Time (s)
2D-4C	k-means	0.010365	0.017803	0.684979	0.289137	0.011049
	ACO	0.006186	0.009436	0.669946	0.598628	0.419201
	PSO	0.031725	0.0107	0.051462	0.004398	1.74367
	BF	0.002758	0.010182	0.329451	0.002997	0.27472
	PIPSO-C	**0.033327**	**0.022343**	**0.686839**	**0.603692**	**0.228352**
10D-4C	k-means	0.034012	0.013869	0.052886	0.080985	0.023524
	ACO	0.039315	0.089308	0.06987	0.434701	0.919202
	PSO	0.050278	0.035419	0.034022	0.062165	5.57154
	BF	0.011764	0.002192	0.040214	0.044709	0.902037
	PIPSO-C	**0.054723**	**0.09378**	**0.073127**	**0.043248**	**0.898908**

Iris	k-means	0.13534	0.096661	0.602076	0.02267	0.004941
	ACO	0.04271	0.071771	0.374798	0.005974	0.07683
	PSO	0.008045	0.046406	0.53276	0.02067	0.48719
	BF	0.010324	0.02489	0.369183	0.004837	0.058962
	PIPSO-C	**0.135633**	**0.099879**	**0.60283**	**0.002581**	**0.056794**
Wine	k-means	0.006755	0.003063	0.398942	0.002768	0.077548
	ACO	0.011879	0.010041	0.358704	0.003944	0.180339
	PSO	0.0107	0.007082	0.006183	0.008435	0.3288
	BF	0.002899	0.002828	0.270582	0.001703	0.140304
	PIPSO-C	**0.014925**	**0.013937**	**0.402031**	**0.001505**	**0.12468**
Glass	k-means	0.012287	0.029821	0.211435	0.018903	0.07548
	ACO	0.063637	0.054306	0.127317	0.03715	0.699914
	PSO	0.03952	0.035816	0.004898	0.026172	0.40873
	BF	0.012728	0.02086	0.072933	0.007819	0.203714
	PIPSO-C	**0.066201**	**0.05928**	**0.213845**	**0.004052**	**0.155524**
Zoo	k-means	0.048437	0.116199	0.151947	0.003507	0.001055
	ACO	0.372645	0.071418	0.26326	0.009615	3.770243
	PSO	0.023197	0.054153	0.01685	0.0034	0.75973
	BF	0.031157	0.044865	0.093465	0.003154	0.552599
	PIPSO-C	**0.375973**	**0.121545**	**0.266878**	**0.001472**	**0.537638**
Ionosphere	k-means	0.001288	0.004366	0.075354	0.07385	0.02578
	ACO	0.001372	0.006718	0.04071	0.081232	1.094375
	PSO	0.000967	0.001528	0.003408	0.036454	4.6563
	BF	0.001202	0.003147	0.042282	0.023309	0.476011
	PIPSO-C	**0.004237**	**0.010847**	**0.078659**	**0.003555**	**0.459911**

(4) Table 3 depicts the standard deviation of various performances indices. The proposed PIPSO-C technique provides stability over the databases. Figure 3 compares the fitness value function SE for ACO, PSO, BF and proposed PIPSO-C technique.

(5) Overall, the proposed PIPSO-C provides the best results for majority of databases when compared with the other existing cluster techniques for the performance indices.

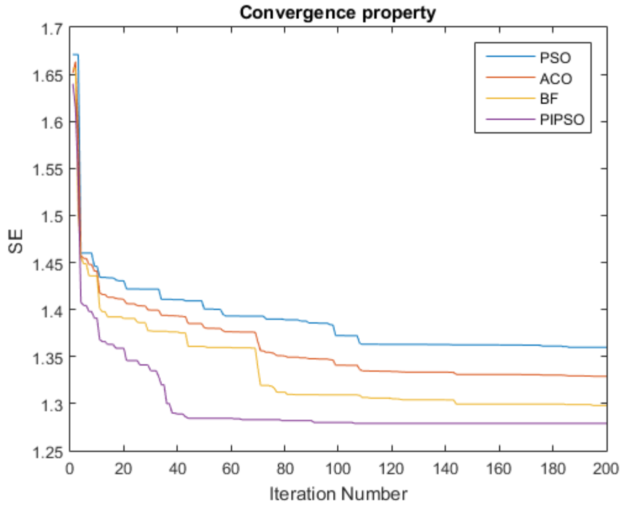

Figure 2. Convergence property of PSO, ACO, BF and purposed PIPSO-C techniques

From the persuasion of above advantages and features of the purposed PIPSO-C technique, it is far a powerful clustering technique and can attain fruitful results for more complex databases with different data cluster sizes, multi dimensional and densities. The characteristic of the presented technique can accomplish itself and offer a best solution for real world cluster problems.

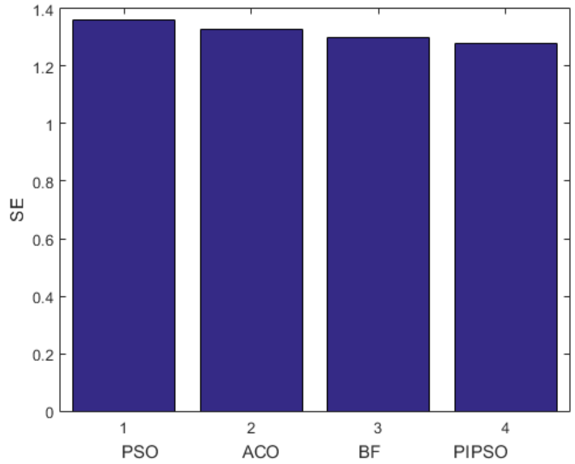

Figure 3. Comparison of fitness value SE for PSO, ACO, BF and purposed PIPSO-C techniques

VI. CONCLUSION

An effectual and superior clustering method primarily based on the application of nature inspired parameter improved particle swarm optimization algorithm has been presented. The data clustering problem is transformed into an optimization search problem to identify the optimum center for each cluster groups by minimizing the objective function. The various real and synthetic databases are analyzed and their comparison studies are described to reveal that the purposed PIPSO-C technique has been utilized to accomplish superior quality on clustering of different dimensional real database as well as synthetic database. The confidence results obtained from the simulation studies apparent that the purposed technique can also classify the optimal cluster groups with various data shapes, sizes, multi dimensional and densities. The presented data clustering method can be utilized in various practical applications to get an optimal solution for real world problems. In coming to the future scope, the purposed PIPSO-C technique can be extended to discover the data cluster groups for web applications. The purposed technique can also broadened to DNA sequence clustering problems that has been turned into an essential research in the field of biomedical.

VII. REFERENCES

[1]. A. K. Jain, M. N. Murty, and P. J. Flynn, Data Clustering: A Review. 1999.

[2]. J. Macqueen, "Some methods for classification and analysis of multivariate observations (1967)," in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, 1967, vol. 1, pp. 281-297.

[3]. Raymond T. Ng and Jiawei Han, "Efficient and Effective Clustering Methods for Spatial Data Mining," in VLDB ’94 Proceedings, 1994, pp. 144-155.

[4]. A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," J. R. Stat. Soc. Ser. B Methodol., vol. 39, no. 1, pp. 1-38, 1977.

[5]. M. Filippone, F. Camastra, F. Masulli, and S. Rovetta, "A survey of kernel and spectral methods for clustering," Pattern Recognit., vol. 41, no. 1, pp. 176-190, Jan. 2008.

[6]. A. G. Delavar and Y. Aryan, "HSGA: a hybrid heuristic algorithm for workflow scheduling in cloud systems," Clust. Comput., vol. 17, no. 1, pp. 129-137, Mar. 2014.

[7]. M. Aruna, D. Bhanu, and S. Karthik, "An improved load balanced metaheuristic scheduling in cloud," Clust. Comput., pp. 1-9, Sep. 2017.

[8]. Baalamurugan K. M and Dr. S. Vijay Bhanu, "Analysis of Cloud Storage Issues in Distributed Cloud Data Centres by Parameter Improved Particle Swarm Optimization (PIPSO) Algorithm," Int. J. Future Revolut. Comput. Sci. Commun. Eng. IJFRSCE, vol. 4, no. 1, pp. 303-307, Jan 18.

[9]. D. Teijeiro, X. C. Pardo, D. R. Penas, P. González, J. R. Banga, and R. Doallo, "A cloud-based enhanced differential evolution algorithm for parameter estimation problems in computational systems biology," Clust. Comput., vol. 20, no. 3, pp. 1937-1950, Sep. 2017.

[10]. Xiuqin Pan, Yong Lu, Na Sun, and Sumin Li, "A hybrid artificial bee colony algorithm with modified search model for numerical optimization," Clust. Comput., no. 2017, pp. 1-8.

[11]. L. W. Koenig and A. M. Law, "A procedure for selecting a subset of size m containing the l best of k independent normal populations, with applications to simulation," Commun. Stat. - Simul. Comput., vol. 14, no. 3, pp. 719-734, Jan. 1985.

[12]. Martin Kotyrba, Eva Volna, and Zuzana Kominkova Oplatkova, COMPARISON OF MODERN CLUSTERING ALGORITHMS FOR TWODIMENSIONAL DATA. Brescia: ECMS, 2014.

[13]. J. Kim, W. Lee, J. J. Song, and S.-B. Lee, "Optimized combinatorial clustering for stochastic processes," Clust. Comput., vol. 20, no. 2, pp. 1135-1148, Jun. 2017.

[14]. L. Li, Y. Zhou, and J. Xie, "A Free Search Krill Herd Algorithm for Functions Optimization," Math. Probl. Eng., vol. 2014, pp. 1-21, 2014.

[15]. B. Mandal, P. K. Roy, and S. Mandal, "Economic load dispatch using krill herd algorithm," Int. J. Electr. Power Energy Syst., vol. 57, pp. 1-10, May 2014.

[16]. T. Zeugmann et al., "Particle Swarm Optimization," in Encyclopedia of Machine Learning, C. Sammut and G. I. Webb, Eds. Boston, MA: Springer US, 2011, pp. 760-766.

[17]. Arthur, David, and Sergei Vassilvitskii, "k-means++: The advantages of careful seeding," Proc. Eighteenth Annu. ACM-SIAM Symp. Discrete Algorithms, pp. 1027-1035, 2007.

[18]. T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "A local search approximation algorithm for k-means clustering," 2002, pp. 10-18.

[19]. F. van den Bergh and A. P. Engelbrecht, "A study of particle swarm optimization particle trajectories," Inf. Sci., vol. 176, no. 8, pp. 937-971, Apr. 2006.

[20]. G. L. Valentini et al., "An overview of energy efficiency techniques in cluster computing systems," Clust. Comput., vol. 16, no. 1, pp. 3-15, Mar. 2013.

[21]. X. Chen and D. Long, "Task scheduling of cloud computing using integrated particle swarm algorithm and ant colony algorithm," Clust. Comput., pp. 1-9, Dec. 2017.

[22]. E. A. Neeba and S. Koteeswaran, "Bacterial foraging information swarm optimizer for detecting affective and informative content in medical blogs," Clust. Comput., pp. 1-14, Sep. 2017.

[23]. S. Mishra and C. N. Bhende, "Bacterial Foraging Technique-Based Optimized Active Power Filter for Load Compensation," IEEE Trans. Power Deliv., vol. 22, no. 1, pp. 457-465, Jan. 2007.

[24]. M. Wan, L. Li, J. Xiao, C. Wang, and Y. Yang, "Data clustering using bacterial foraging optimization," J. Intell. Inf. Syst., vol. 38, no. 2, pp. 321-341, Apr. 2012.