



Designing a New Hybrid K-Means Optimization Algorithm

Navreet Kaur

Research Scholar: Department of Computer Science and
Engineering
Sri Guru Granth Sahib World University
Fatehgarh Sahib, India

Shruti Aggarwal

Assistant Professor: Department of Computer Science and
Engineering
Sri Guru Granth Sahib World University,
Fatehgarh Sahib, India

Abstract- Clustering is an unsupervised learning process with the main objective of organizing data into certain clusters and groups such that the data objects in the same cluster have higher similarity but dissimilar to the objects in other clusters. Many algorithms have been presented till now. These algorithms are k-means algorithm, hybrid k-means algorithms, and variants of k-means. The hybrid k-means optimization algorithms are those that combines k-means algorithm with various optimization algorithms like KBAT, CGA, KFA etc. The k-means algorithms have many problems like k-means algorithm converges to local minima rather than giving a global optimum results, hybrid k-mean algorithms have accuracy and efficiency problems. A new hybrid k-means optimization clustering algorithm is presented in the proposed study to solve the problems of existing k-means algorithm and hybrid k-mean algorithms. The output parameters like intra cluster distance, purity, recall etc are use to compare the performance of existing hybrid k-means algorithm with a new hybrid k-means optimization clustering algorithm. The results of these output parameters show that the proposed ABCGA technique is more efficient and it requires less computational time than existing hybrid k-means algorithm. Also the recall and purity of proposed hybrid algorithm is better than existing hybrid k-means algorithms and k-means algorithm but still there is scope for improvement in the proposed technique to extend this approach to handle variety of situations and information.

Keywords- Data mining, Clustering, K-means algorithm, Clustering based genetic algorithm, ABPPO algorithm, ABCGA technique

I. INTRODUCTION

Data mining [1] is a powerful concept for data analysis and process of discovering interesting pattern from the large amount of data. It is also defined as mining of knowledge from huge amount of data. The goals of data mining are fast retrieval of data or information and knowledge discovery. Clustering [2] is an unsupervised learning process with the main objective of organizing data into certain clusters and groups such that the data objects in the same cluster have higher similarity but dissimilar to the objects in other clusters. Clustering is known as unsupervised learning because the class label information is not present. Clustering techniques [3] are widely used in many application areas such as Artificial intelligence, image processing, biology, marketing, pattern reorganization, machine learning, data mining and Information retrieval. To perform clustering various clustering techniques are used and these techniques are partitioning based, hierarchical based, density based and grid based techniques. The partitioning based clustering technique is that which organizes the object of data set into groups or clusters and in this partitioning technique the clusters formed are of spherical shape. For example k-means, k-medoids, k-mode etc. The hierarchical based techniques are also used to find the clusters of spherical shaped clusters but the clusters are formed in the form of hierarchy or tree. To find the clusters of arbitrary shape density based clustering technique is used. For example DBSCAN (Density-based clustering based on connected regions with high Density). Grid based techniques used space-driven approach by partitioning the embedding space into cells independent of distribution of input objects. For Example: STING (Statistical Information Grid).

A. K-means Algorithm

The k-means algorithm [4] is popular clustering method which falls in the category of partitioning based clustering method. It was firstly proposed by MacQueen in 1967. K-means is an unsupervised, numerical, iterative method of clustering. To measure similarity, mean value of the objects within a cluster is used.

The algorithm consists of two separate phases:

1st Phase: The k centroids are selected randomly where the value of k is fixed.

2nd Phase: In data set each object belongs to the nearest the centroid. The distance between each object of data set and centroid of cluster is measured by using the euclidean distance.

Steps for k-means algorithm

The steps for the k-means algorithm for clustering are given below:

Start

Input: Number of desired clusters, k and a database $X = \{x_1, x_2, x_3 \dots x_n\}$ containing n data objects.

Output: A set of k clusters

Method:

Let $X = \{x_1, x_2, x_3 \dots x_n\}$ be the set of data points

1. Choose the number of clusters k to be formed.
2. The cluster centroids are selected randomly.
3. The distance between each data point and cluster centers is calculated.
4. The data point is assigned to the cluster center where distance between cluster center and data point is minimum than other cluster centers.
5. The new cluster center is recalculated.
6. Also the distance between each data point and new obtained cluster centers is recalculated.

7. Repeat from third step if data point was reassigned otherwise stop.

The algorithm of k-means shows the steps to be carried out for clustering. In this approach, First the number clusters to be formed are selected and then the cluster centers are chosen randomly from data points that are being clustered, after that choose centroids initially and then compute distance between objects and centroid. Clusters are formed based on minimum distance. The data point is assigned to the cluster center where distance between cluster center and data point is minimum than other cluster centers. The new cluster center is recalculated. Also the distance between each data point and new obtained cluster centers is recalculated. Repeat from third step if data point was reassigned otherwise stop.

The K-means algorithm is simple and flexible and also easily implemented. But it has various limitations like in k-means algorithm user need to specify the number of cluster in advanced, k-means algorithm converges to local minima rather than giving a global optimum result, the definition of means limits the application only to numerical variables and also k-means is sensitive to outliers and noise.

II. REVIEW OF LITERATURE

In review of literature the various hybrid k-mean algorithms are discussed like KFA, KABC, K-Krill herd, KACO, PSO-ACO-K etc and also the Clustering based genetic algorithm used for clustering is explained here. The optimization technique adaptive biogeography based predator-prey optimization is also described in the literature review.

A. Hybrid k-mean Algorithms

Hybrid k-means algorithms are those algorithms that combines k-means algorithm with k-means variants or k-means optimization algorithms. The hybrid k-means algorithms are used to improve performance of k-means algorithm and these algorithms provides better clustering than k-means algorithm. The hybrid k-mean algorithms KFA, KABC, KACO etc. are described below.

In KGA [5] genetic algorithms are commonly used to generate high-quality solutions for optimization and search problems by relying on bio-inspired operators such as mutation, crossover and selection. In KFA algorithm [6] the k-means clustering algorithm is combined with firefly algorithm [7]. The firefly is used to find the centroids for specified number of clusters and then k-means is used to refine the centroids and clusters. In KACO algorithm [8] firstly ACO is applied to starting condition because cluster quality is based on it. PSO-ACO-K algorithm [9] combines the particle swarm optimization, ant colony optimization and k-means. In KCUCKOO algorithm [10] Cuckoo search randomly select initial centroids and leads too many iterations to overcome the problem of k-means. In KFP algorithm [11] the flower pollination algorithm is used to reduce the disadvantages of k-means local optima and its results are used to select the centroids of clusters in k-means. In KBAT algorithm [12] the optimization BAT algorithm helps to reduce the local optimal problem of k-means clustering algorithm. In KABC algorithm [13] the k-means is combined with artificial bee colony algorithm for

optimization and clusters formed are better than in simple k-means algorithm. In K-Krill herd algorithm [14] the krill herd is used to initialize the centers for clusters in k-means. The local search strategy of krill herd is used to provide local optimal results. In BAT k-medoids [15] k-medoids clustering algorithm is combined with BAT algorithm to solve the optimization problems of k-medoids algorithm. IGSA-KHM algorithm [16] not only helps the KHM clustering escape from local optima but also overcomes the slow convergence speed of the IGSA. In KPSSO algorithm [17] the results of PSO algorithm [18] is used as the initial seed for the k-means algorithm. The k-means algorithm will be applied for refining and final results. The Tabu-KHM algorithm [19] combines the k-harmonic algorithm with the tabu search algorithm. It combines the optimization property of tabu search and the local search capability of k-harmonic means algorithm together.

These all hybrid k-mean algorithms have many limitations like efficiency and accuracy problems. Due to these limitations there is needed to design a new hybrid k-means optimization clustering algorithm.

B. Clustering based Genetic Algorithm

CGA [20] is clustering based Genetic Algorithm with polygamy selection and dynamic population control technique. Genetic Algorithm is related to biological background, in sense that it follows Darwin theory of natural evolution “the survival of the fittest”. This theory state that the individual who is fit among the whole population will survive and next generation is reproduces by that individual. The genetic algorithm is first proposed by John Holland in 1975. In genetic algorithms first of all some terminologies are understand to get insight of the process. Main terms are gene, chromosome, individual, population. Gene is smallest unit of information carrying capacity. Individual is a set of genes carrying information and further set of individuals is population. According to CGA the fitness values obtained from chromosomes in each generation were clustered into two non-overlapping clusters. The surviving chromosomes in the selected cluster were subjected to polygamy crossover mating process while the population of the off springs which would form the next generation was subjected to dynamic population control mechanisms.

Steps for CGA

The steps for Clustering based genetic algorithm are described below:

Start

Input: Read unstructured data

Output: Clustered Data

Method:

1. The unstructured data is read.
2. After unstructured data is readed then initial value for centroids are generated.
3. The centroids are chosen randomly.
4. if cluster formed then results are generated
5. Else
6. The population (t) is initialized randomly.
7. Fitness of population (t) is determined.
8. Repeat the steps given below until best individual is good enough.

9. The parents from population (t) are selected and then polygamy crossover is performed on parents and it creates population (t+1)
10. Apply dynamic population control involves birth control and population growth control.
11. The mutation of population (t+1) is performed.
12. Fitness of population (t+1) is determined.
13. Until best individual is good enough
14. Generate Results
15. End

11. The solutions are hunt using equation (5) and equation (6).
12. t = t+1
13. End

The equations used in the above ABPPO algorithm are given:

$$\mu_k = E^{\max} \left(\frac{k}{n} \right) \quad (1)$$

Here, E^{max} is maximum emigration rate and n is maximum number of species in the population.

$$\lambda_k = I^{\max} \left(1 - \left(\frac{k}{n} \right) \right) \quad (2)$$

Here, I^{max} is maximum immigration rate and n is maximum number of species in the population.

$$P_m(t+1) = P_m^0 \cdot \frac{1 + (D(1) - D(T))}{D(1)} \quad (3)$$

Here, D (t) is diversity measure at generation t, P_m is rate of mutation and P_m⁰ is initial mutation rate.

$$D(t) = \frac{1}{nh \cdot X} \cdot \sum_{i=1}^{nh} 1 \left(\sum_{j=1}^{nh} 1 (h_{ij} - h_j)^2 \right)^{1/2} \quad (4)$$

Here, D (t) is diversity measure at generation t, nh is number of habitats, h_{ij} is jth SIV of ith habitat.

Here, D (t) is diversity measure at generation t, nh is number of habitats, h_{ij} is jth SIV of ith habitat.

$$h_{\text{predator}}(t) = h_{\text{worst}}(t) + \rho \cdot \left(1 - \frac{1}{t_{\text{max}}} \right) \quad (5)$$

Here, ρ is rate of hunting, h_{predator} (t) is a possible solution that represents a predator at generation t, h_{worst} (t) the worst solution in the population at generation t and t_{max} is maximum number of generations.

$$h(t+1) = h(t) + \rho \cdot e^{-|d|} \quad \text{if } (d > 0) \\ \text{And } h(t+1) = h(t) - \rho \cdot e^{-|d|} \quad \text{if } (d < 0) \quad (6)$$

Here, ρ is rate of hunting, emigration, d is euclidean distance between predator and prey and n is maximum number of species in the population.

The ABPPO algorithm is more efficient, fast and robust, ABPPO algorithm requires less number of iterations to reach the optimal point, it does not form the grouping of the habitats that have similar characteristics and its solutions survive forever.

III. PROPOSED WORK

In proposed work the new designed hybrid k-mean algorithm is described. This algorithm combines the ABPPO optimization technique with the Clustering based genetic algorithm that is CGA algorithm.

In CGA algorithm mutating the chromosomes does not give additional new chromosomes to population due to polygamy crossover and population control. The CGA algorithm solves the problem with multiple solutions but it has many problems like the CGA algorithm does not give better performance when number of generation is more, some optimization problems cannot be solved by CGA algorithm, no absolute assurance that it will find global optimum results and it requires more computation time when population is large.

C. ABPPO Algorithm

ABPPO Algorithm [21] means adaptive biogeography based predator-prey optimization is also a bio inspired optimization technique that includes adaptive mutation combined with predator-prey pattern for attaining the global optimal point. In adaptive mutation scheme, the diversity measure of distance-to-average point is the predominant feature that dodges the supremacy of extremely feasible solutions throughout enhancing the population diversity. The predators explore around the elite prey in a determined way, whereas the preys search the solution space to evade from the predators. This tool improves the utilization and searching abilities of the BBO exploration procedure.

Steps for ABPPO

The steps for algorithm of adaptive biogeography based predator-prey optimization are described below:

Start

1. Initial population of habitats is generated
2. While t < t max
3. The HSI of individual habitat is calculated
4. The immigration rate and emigration rate for each habitat are calculated using equation (1) and equation (2).
5. Based on the immigration rates probabilistically choose the immigrating habitats.
6. To select the emigrating habitats use roulette wheel selection based on the emigration rates.
7. After this based on the selected habitats in previous step migrate randomly chosen SIVs.
8. For each habitat based on the mutation rate probabilistically perform mutation.
9. The population diversity is evaluated based on equation (4).
10. The mutation rate is computed based on equation (3).

A. Clustering using ABCGA Hybrid Technique

The ABCGA means Adaptive Biogeography Clustering based genetic algorithm. In hybrid technique the CGA which means Clustering based Genetic Algorithm is used along with ABPPO which stands for Adaptive biogeography based predator-prey optimization. In the proposed technique the clustering process is similar to k-means algorithm and then ABPPO technique is applied for selection in genetic algorithm which is much robust and efficient in converging to the global optimal point with minimum computational time. After this the other steps like mutation, polygamy crossover of genetic algorithm are performed. The ABCGA algorithm combines the functionalities of the CGA algorithm with the functionalities of ABPPO optimization algorithm. The ABCGA technique is used to reduce the limitations of Clustering based genetic algorithm and improves the results of existing Clustering based genetic algorithm. The Steps for ABCGA algorithm are given below.

Steps for ABCGA

The steps for algorithm of adaptive biogeography clustering based genetic algorithm are described below:

Start

Input: Read unstructured data

6. if cluster formed then results are generated
7. Else
8. The population (t) is initialized randomly.
9. Fitness of population (t) is determined.
10. Repeat the steps given below until best individual is good enough.
11. The parents from population (t) are selected and then polygamy crossover is performed on parents and it creates population (t+1)
12. Apply the dynamic population control involves the birth rate control and the population growth rate control.
13. The mutation of population (t+1) is performed.
14. Fitness of population (t+1) is determined.
15. Until best individual is good enough
16. Generate Results
17. End

In flowchart for ABCGA algorithm firstly the unstructured data is read and then the initial population is generated. The flowchart for ABCGA is shown in Figure1.

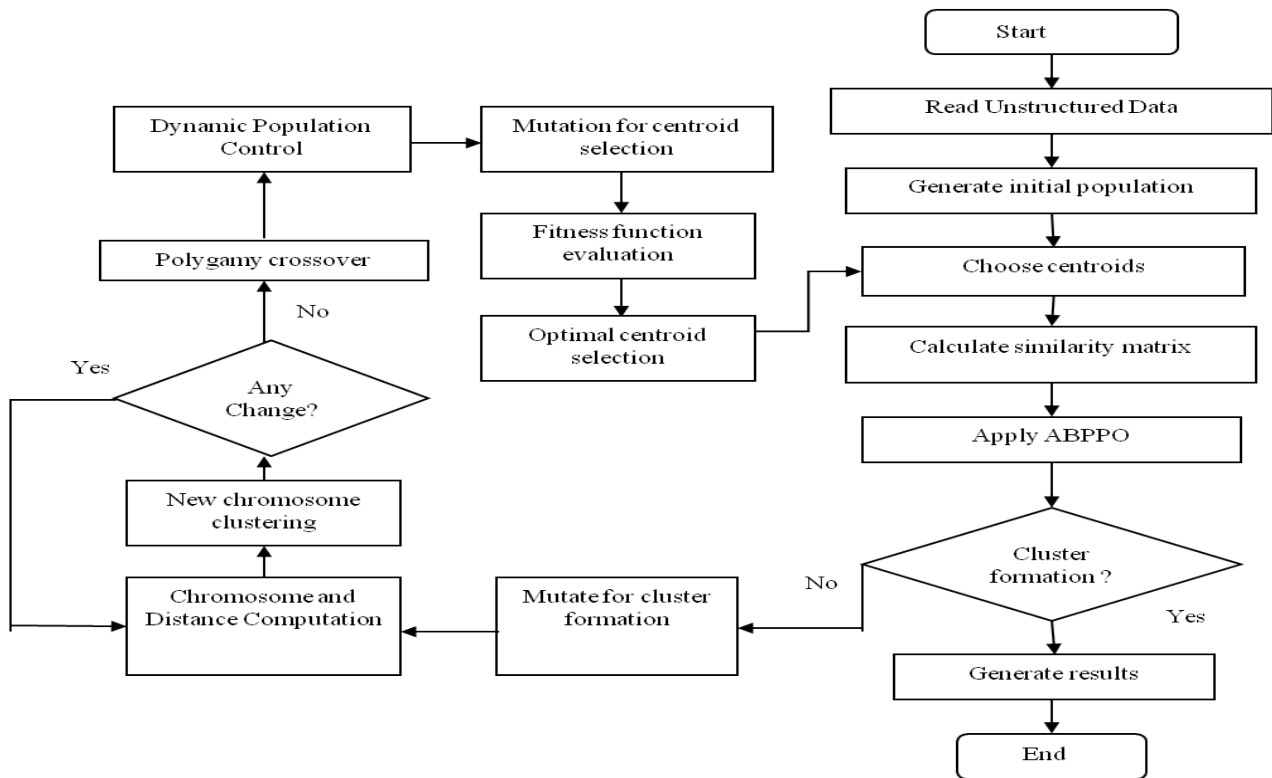


Fig 1: Flowchart for ABCGA algorithm

Output: Clustered Data

Method:

1. The unstructured data is read.
2. After unstructured data is readed then initial population is generated.
3. The centroids are chosen randomly.
4. Choose Probable elements for ABPPO optimization technique.
5. Apply Adaptive biogeography based predator-prey optimization for selection

After this the centroids are chosen randomly. Next step is to calculate the similarity matrix and after calculating the similarity matrix the ABPPO technique is applied. The mutation operation is performed for cluster formation and then chromosome and distance computation is performed. The next step is to perform the new chromosome clustering and if there are changes then again the then chromosome and distance computation is performed. After that the new chromosome clustering is performed and if no change then polygamy crossover is applied and after applying polygamy crossover the dynamic population control is performed and

then mutation is performed for centroid selection and then fitness function is evaluated. After this the optimal centroids are selected and the centroids for clustering are chosen from these optimal selected centroids

ABCGA algorithm requires less computation time than CGA algorithm, Also this algorithm converges to global solution with few generations and efficiency of ABCGA algorithm is more.

IV. RESULTS

In results the various data sets used and also the output parameters are discussed that are used to compare the performance of new designed algorithm with the existing algorithms like k-means and CGA algorithm. The experimental results are shown with the help of bar graphs.

A. Data Set

The data sets used are data sets from the UCI Machine Learning Repository are Wine, Iris, Seed, Breast Cancer and Liver Disorders data set. Also number of attributes and number of instances in these data sets are described. These data sets are shown in Table 1.

Table 1: Data Sets

Sr. No.	Data Set Name	Number of Instances	Number of Attributes
1	Wine	179	14
2	Iris	150	4
3	Seed	210	5
4	Breast Cancer	699	10
5	Liver Disorders	345	7

This table shows that the total number of attributes in Wine data set is 14 and number of instances is 179, total number of attributes in Iris data sets is 4 and total number of instances is 150. The total number of attributes in Seed data set is 5 and total number of instances is 210. The total number of attributes in breast cancer data set is 10 and total number of instances is 699. The total number of attributes in Liver disorders data set is 7 and total number of instances is 345.

B. Output Parameters

The output parameters are those parameters based upon which the performance of existing clustering algorithm is compared with the new hybrid k-means optimization algorithm. Some output parameters are described below:

- **TP : True Positive**
It measures the proportion of positives that are correctly identified. It is also known as Sensitivity. Example: Sick people who are correctly identified as having the condition.
- **TN : True Negative**
It measures the proportion of negatives that are correctly identified. It is also known as Specificity. Example: The percentage of sick people who are correctly identified as not having the condition.
- **FP : False Positive**

They are incorrectly identified. Example: Healthy people incorrectly identified as sick.

- **FN : False Negative**
They are those which are incorrectly rejected. Example: Sick people incorrectly identified as healthy.

Some major parameters based upon which the performance of proposed algorithm is evaluated and compared with existing algorithm are described below:

- **Recall**
It may be defined as the number of most relevant features data set of cluster upon total number of data set features clustered.

$$\text{Recall} = \frac{\text{No of most relevant features data set of cluster}}{\text{total no of data set features Clustered}}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{1}$$

Where, TP= True Positive, FN= False Negative
Larger value of it gives better clustering results.

- **Purity**
Purity is defined as the percent of the total number of objects that were classified correctly. To compute purity each cluster is assigned to the class which is most frequent in a cluster. It describes the cluster quality.

$$\text{Purity} = \frac{1}{N} \sum_k \max_j |w_k \cap c_j| \tag{2}$$

Where N = number of objects, k = number of clusters, cj is set of classes and wk is the set of clusters.

- **Intra Cluster Distance**
Intra Cluster Distance may be simply the distance between a point and its cluster centre and by taking the average of all of these distances the intra cluster distance is given below.

$$\text{Intra Cluster Distance} = \frac{1}{N} \sum_{i=1}^k \sum_{x \in C_i} ||x - z_i||^2 \tag{3}$$

Where N is number of objects, k is the number of clusters and zi is the cluster centre of Ci

- **Computation Time**
Computation time means the time taken required by the computer to perform a given set of computations. If Computation time is less than the clustering algorithm is better than algorithms having more Computation time.
- **Computational Complexity**
Computational Complexity of an algorithm is a measure of how many steps the algorithm will require in the worst case for an instance or input of given size. The number of steps is measured as a function of size.

C. Experimental Results

The k-means, CGA and ABCGA algorithms are executed on JAVA version jdk1.8 with hardware configuration Intel Core i3 and above, RAM: 4GB and software used are Net Beans IDE 8, Microsoft Excel and Operating System: XP, windows 7. Experimental results shows the comparison of k-means algorithm, CGA algorithm and new designed ABCGA algorithm for various output parameters using five data sets that are mentioned above.

1) Comparison using Recall

The comparison of CGA algorithm and k-means algorithm with the newly designed ABCGA algorithm using recall when number of clusters is 8 for five data sets that are Wine, Iris, Seed, Breast cancer and Liver Disorders is done.

The results for recall value of CGA, k-means and proposed ABCGA algorithms is shown using the bar graph that is shown in the Figure 2.

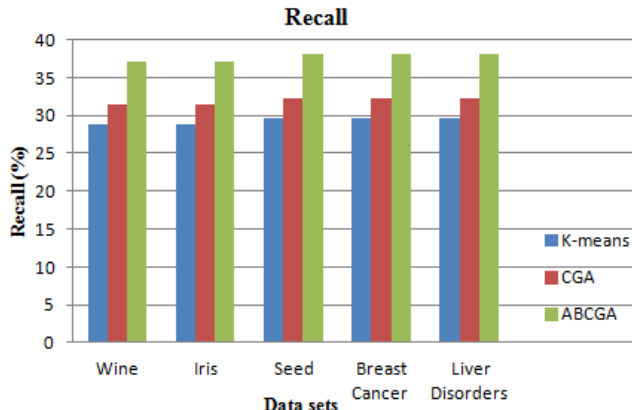


Fig 2: Comparison using Recall

It shows that the recall is improved for proposed ABCGA algorithm than CGA algorithm and k-means algorithm.

2) Comparison using Purity

The comparison using purity when number of clusters is 8 for five data sets that are Wine, Iris, Seed, Breast cancer and Liver Disorders. The values for purity are shown in Table 2 given below.

Table 2: Comparison using Purity

Sr. No.	Dataset	K-means	CGA	ABCGA
1	Wine	30.52	34.04	40.7
2	Iris	30.52	34.04	40.7
3	Seed	31.34	34.92	41.8
4	Breast Cancer	31.35	34.96	41.5
5	Liver Disorders	31.37	34.97	41.9

This table shows that the value of purity is increased for ABCGA algorithm than k-means and CGA algorithm for all these five Data sets.

3) Comparison using Intra Cluster Distance

The comparison for three algorithms using intra cluster distance when number of clusters is 8 for five data sets that are Wine, Iris, Seed, Breast cancer and Liver Disorders. These results are shown in Figure 3.

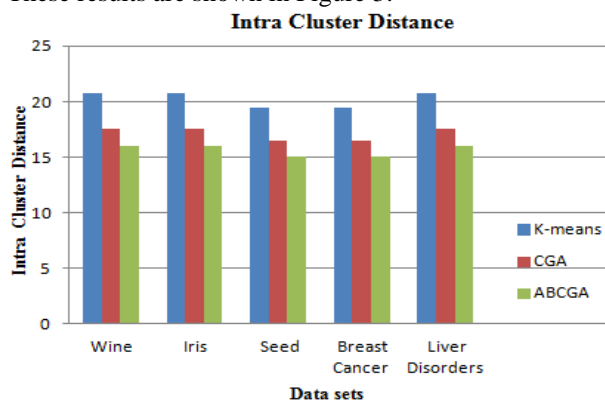


Fig 3: Comparison using Intra Cluster Distance

It shows that the intra cluster distance is improved for proposed ABCGA algorithm than CGA algorithm and k-means algorithm.

4) Comparison using Computation Time

The comparison of CGA algorithm and k-means algorithm with the newly designed ABCGA algorithm using computation time when number of clusters is 8 for five data sets that are Wine, Iris, Seed, Breast cancer and Liver Disorders is done. The values for computation time are shown in Table 3.

Table 3: Comparison using Computation Time (m sec)

Sr. No.	Dataset	K-means	CGA	ABCGA
1	Wine	31	47	16
2	Iris	16	47	16
3	Seed	32	47	15
4	Breast Cancer	31	47	15
5	Liver Disorders	16	46	16

This table shows that the value of Computation time is increased for ABCGA algorithm than k-means and CGA algorithm for all these five Data sets.

5) Comparison using Computational Complexity

The comparison of CGA algorithm and k-means algorithm with the newly designed ABCGA algorithm using computational complexity when number of clusters is 8 using five data sets that are Wine, Iris, Seed, Breast cancer and Liver Disorders. These results are shown with the help of bar graph in Figure 4.

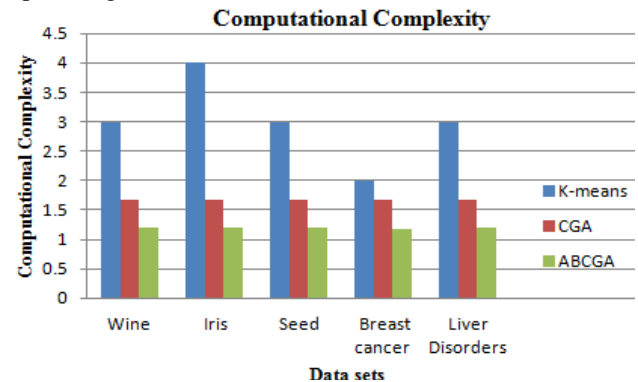


Fig 4: Comparison using Computational Complexity

It shows that the Computation time is improved for proposed ABCGA algorithm than CGA and k-means algorithm.

V. CONCLUSION

In this work, the k-means algorithm for clustering and various hybrid k-mean algorithms are studied. K-means algorithm is the popular method for cluster analysis but it has some drawbacks that it does not guarantee for optimal solution, user need to specify the number of clusters in advance and also k-means is sensitive to noise and outliers. Also hybrid k-mean algorithms have many drawbacks like have efficiency, accuracy problems and some of them do not provide optimum results. To overcome the drawbacks of k-means clustering algorithm and its hybrids a new hybrid k-means optimization clustering algorithm is designed. The performance of new ABCGA algorithm is compared with the k-means algorithm and other exiting k-mean hybrid algorithms based upon various performance parameters like

recall, purity, computation time etc. The ABCGA algorithm requires less computation time than existing CGA hybrid k-means algorithm. The recall is also improved for clustering in ABCGA algorithm than k-means and its hybrid CGA algorithm.

There is scope for improvement in the proposed technique to handle variety of information. Also more than one nature inspired algorithm for optimization is used to hybrid with k-means and variant of k-means. This approach can be used for largest data sets. Other optimization techniques can be used to improve the performance of clustering and also other clustering algorithm can be used for best results.

VI. REFERENCES

- [1] Ming-Syan Chen, Jiawei Han, Ps Yu, "Data Mining: An overview from database perspective" IEEE Transaction on knowledge and data engineering, Vol. 8, Issue 6, pp. 886-883, 1996.
- [2] Anil .Jan, "Data Clustering: 50 years beyond k-means" Pattern Recognition Letters, Elsevier, Vol. 31, pp. 651-666, 2010.
- [3] Lior Rokach and Oded Maimon, "Clustering methods" Data mining and Knowledge handbook, pp. 321-352, 2005.
- [4] J.B. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations" Proceedings of Fifth Berkeley Symposium on Mathematics Statistics and Probability, University of California Press, Vol. 1, pp. 281-297, 1967.
- [5] M.Mary and Shanthi Rani, "A Genetic Algorithm Based K-Means Algorithm For Compressing Images" International Journal Of Engineering And Computer Science ISSN, Vol. 4, Issue 9, pp. 14359-14362, 2015.
- [6] S.J Nanda, G. Panda, "A Survey on nature inspired metaheuristic algorithm for partition clustering" Swarm and Evolutionary Computation, Elsevier, Vol. 16, pp. 1-18, 2014.
- [7] Xin-She Yang, "Firefly algorithms for multimodal optimization" Stochastic Algorithms: Foundations and Applications, SAGA 2009. Lecture Notes in Computer Sciences, Vol.5792, pp.169-178, 2009.
- [8] K.Aparana and Mydhili K.Nair, "Enhancement of k-means algorithm using ACO as optimization technique on high dimensional data" 2014 international conference on Electronics and Communication Systems (ICECS) IEEE, pp. 1-5, 2014.
- [9] Taher Niknam and Babak Amiri, "An efficient hybrid approach based on PSO, ACO and k-means for cluster analysis" Applied Soft Computing, Elsevier, vol. 10, pp. 183-197, 2010.
- [10] Saida Ishak Boushaki, Kamel Nadjet and Omar Bendjeghaba, " A New Hybrid Algorithm for Document Clustering based on Cuckoo Search and K-means", Recent advances on Soft Computing and Data Mining SCDM Springer, pp. 59-68, Vol. 287, 2014.
- [11] Tang Rui, Fong Simon, Yang Xin-She and Deb Sujay, "Integrating nature-inspired optimization algorithms to k-means clustering" 2012 seventh International Conference on Digital Information Management ICDIM, IEEE, pp. 116-123, 2012.
- [12] E. Obsab, X. S. yang, F. Diaz, P. Lopez-Garcia, R. Carballedo, "An Improved Discrete Bat Algorithm for Symmetric and Asymmetric Travelling Salesman Problems", Engineering Applications of Artificial Intelligence, Vol. 48, pp. 59-71, 2016.
- [13] Karaboga K, Dervis D, Ozturk, "A novel clustering approach: Artificial Bee Colony (ABC) algorithm" Applied Soft Computing, Springer, Vol. 11, pp. 7-652, 2011.
- [14] Hamed Nikbakht, Hamid Mirvaziri, "A new clustering approach based on K-means and Krill Herd algorithm" 23rd Iranian Conference on Electrical Engineering, IEEE, 2015.
- [15] Monica Sood and Shilpi Bansal, "K-Medoids Clustering Technique using Bat Algorithm", International Journal of Applied Information Systems, pp. 20-22, 2013.
- [16] Minghao Yin, Yanmei Hu, Fengqin Yang, Xiangtao Li, Wenxiang Gu, "A novel hybrid K- harmonic means and gravitational search algorithm approach For clustering" Expert Systems with Applications, Elsevier, Vol. 38, pp. 9319-9324, 2011.
- [17] Yucheng Kao, Szu-Yuan Lee, "Combining K-means and particle swarm optimization for dynamic data clustering problems", IEEE International Conference on Intelligent Computing and Intelligent Systems, 2009.
- [18] Kennedy.J and Eberhart.R, "Particle Swarm Optimization" Proceedings of IEEE International Conference on Neural Networks. pp. 1942-1948, 1995.
- [19] Gungor. Z and Unler. A, "k-Harmonic Means Data Clustering with Tabu Search Method" Applied Mathematical Modeling, Vol. 32, pp. 1115-1125, 2008.
- [20] A.M. Aibinu, H.Bello Salau, Najeeb Arthur Rahman, M.N. Nwohu, C.M. Akachukwu, "A novel Clustering based Genetic Algorithm for route optimization", Engineering Science and Technology an International Journal, Vol. 19, pp. 2022-2034, 2016.
- [21] Ananthy Christy, Perianayagam Ajay-D-Vimal Raj, Sanjeevikumar Padmanaban, Rajasekar Selvamuthukumar, Ahmet H.Ertas, "A bio-inspired novel optimization technique for reactive power flow", Engineering Science and Technology an International Journal, Vol. 19, pp. 1682-1692, 2016.