

Dbscan Assisted by Hybrid Genetic K Means Algorithm



Suresh Kurumalla, Chandusha Kanda

Abstract: The data mining algorithms functioning is main concern, when the data becomes to a greater extent. Clustering analysis is a active and dispute research direction in the region of data mining for complex data samples. DBSCAN is a density-based clustering algorithm with several advantages in numerous applications. However, DBSCAN has quadratic time complexity i.e. $O(n^2)$ making it complicated for realistic applications particularly with huge complex data samples. Therefore, this paper recommended a hybrid approach to reduce the time complexity by exploring the core properties of the DBSCAN in the initial stage using genetic based K-means partition algorithm. The technological experiments showed that the proposed hybrid approach obtains competitive results when compared with the usual approach and drastically improves the computational time.

Keywords: DBSCAN, Genetic Algorithm, K-Means algorithm, Image database, clustering.

I. INTRODUCTION

Unsupervised learning plays a vital role in research field of machine learning and data mining [1]. Clustering comes under unsupervised learning. It is one of the efficient procedures that issued in finding significant knowledge for patterns. The primary objective for clustering is to isolating the precise dataset into clusters so that objects which are present in a cluster have high equivalence in compare with one another and distinguish to objects in left over clusters. Clustering is a fundamental data mining issue that is regularly begin in various areas as well as biology, social science and marketing. Fast and efficient clustering algorithms make a key role in given inherent exploration and browsing mechanisms by arranging high volumes of data into a smaller number of valid clusters.

Researchers have conferred various clustering algorithms [2,3], the essential division on cluster algorithms are partitioning clustering, hierarchical clustering, grid and density-based clustering.

Manuscript received on February 10, 2020.
Revised Manuscript received on February 20, 2020.
Manuscript published on March 30, 2020.

* Correspondence Author

Dr. K. Suresh, Assistant professor in Anil Neerukonda Institute of Technology & Sciences, Visakhapatnam

Mrs. K. Chandusha, Assistant professor in Anil Neerukonda institute of Technology & Sciences, Visakhapatnam.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

In some cases, these algorithms can inspire clustering speed and improve effectiveness by taking of feature selection [4], reducing the count of dimensions [5], or applying reference points [6], etc. Even though there is a detailed variations and complications in cluster analysis by certain algorithms, so that the inconsistency among cluster definiteness and effectiveness, average of key value setting and so on.

Cluster analysis has applications in various sectors of business and science, Data reduction, Generating Hypothesis, testing the Hypothesis, Prediction based on groups, Biology, Spatial data analysis and Web mining.

In density-based clustering, the clusters are commencing with the objects associating slowly and isolated by insufficient regions. It has the specified benefits over outstanding clustering algorithms [6,7]. First, it recovers the clusters of arbitrary shapes away from each other from curved shapes. Second, it is efficiently removing the noise. Third, it does not need a pre-described number of clusters unlike k-means [6]. Fourth, the clustering output is not offended by the input order of objects. However, there is also exact drawbacks in the traditional DBSCAN algorithm that exists till today where one of the issues is addressed in this paper. The efficient density-based algorithms are DBSCAN [5], OPTICS [2], and DENCLUE [8], and there have been many methods for enhanced performance [3,12].

DBSCAN algorithm is usually employed density-based clustering algorithms mostly for large complex datasets. However, the DBSCAN algorithm due to its high complexity suffers from some of the limitations. The time complexity of DBSCAN algorithm is $O(n^2)$ which is very high and has more computational time. The major drawback that was addressed in this paper was that the core properties of the objects in DBSCAN algorithm are only to some extent resolved since only certain range queries are executed. Therefore, so as to over this issue, this density-based algorithm is hybridized with the partition-based algorithm as to discover the core properties from the initial clustering approach. The K- Means algorithm is used to get hold of initial clusters.

1.1 Organization of the Paper

A concise discussion on the introduction of the clustering algorithm and the motivation for the suggested approach is given in this section. Section 2 briefly discusses the various types clustering techniques and survey done on these approaches. The proposed hybrid Genetic K-Means based DBSCAN algorithm is quickly interpreted in section 3. The experimental outputs and its analysis is given in section 4 followed by the conclusion and references given in the section 5 and section 6 respectively.

II. LITERATURE SURVEY

Bohm et al [10] recommended CUDA-DClust algorithm. It enhances the successive DBSCAN up to 15 times using the GPU. They also recommended CUDA-DClust that enhanced CUDA-DClust up to 11.9 times using a simple catalog structure. In [14],

a methodology is accessible to decrease the time complexity based on K-Means algorithm. This approach divides the data in k partitions at first step and then uses a Min-Max method to choose points for DBSCAN clustering at second step. Experiments show that our method obtains competitive results with the novel DBSCAN, while extensively improving the computational time

The recommended algorithm in [8] find the noise cluster data. It reduces the outlier problems. DBSCALE algorithm (Density Based Spatial Clustering of Applications for Large Databases) is performed with Naïve Bayes' theorem. It is an upcoming service. This is nearest to the outlier cluster data. The amount of definiteness of algorithm recommended a technique that is obtained a prototype recognized as leader from a dataset that has the data of the prototypes along with density. These are used to form density-based clusters.

The basic disadvantage of k-mean is efficiency, as user necessarily identifies the no. of clusters throughout the initial process. This limitation of predefined number of clusters tends some points of the dataset to keep on un-clustered. So, by enlarging the cluster methods, the predictions can be improved. Therefore, in [11], the normalization is used to get hold of exact outputs by calculating distance to have specific centroid and to get rid of noise data. Backtracking fashion is adopted to find the definite figure of clusters that is defined to analyze the data in better way. The outcomes showed that there is an expansion in clustering when correlated to the concrete procedures

A prototype-based hybrid approach [12] is proposed to accelerate the k-means method. The data-set is primarily separated into irrelevant clusters of conflicting sizes. Then, the fixed prototype is separated into k clusters applying the customized k-means method. This clustering method is identical to the conventional k-means method by removing vacant clusters in the repeated technique. In every cluster of models, every model is regained by its parallel set of patterns to gain the partition of data-set. So, this disjointing of the data-set can be gained by using the conventional k-means method over the entire data-set, a reviewed step is suggested. Experimentally, the recommended method is related with the traditional method and the other current methods that are recommended to quicken the k-means method.

A two-stage hybrid clustering algorithm is proposed in [13] where DBSCAN is improved to proceed the data with accurate aspects. By merging DBSCAN clustering algorithm and one-pass clustering algorithm then two-stage hybrid clustering algorithm is obtained. In the elementary step, to cluster the data one pass clustering algorithm is used. In the second step, improved DBSCAN clustering algorithm is combined with the partition then only ultimate clusters are gathered. The given clustering algorithm is of directly linear time complexity, it is used to evaluate the wide-ranging datasets. The experimental result on actual datasets and synthetic datasets displays that the two-stage hybrid clustering algorithm can assist to evaluate the data with random shape similar to DBSCAN, the operating effectiveness of which is

not only better than the DBSCAN, but also well-organized and accomplished.

III. HYBRID GENETIC K-MEANS BASED DBSCAN APPROACH (GK-DBSCAN ALGORITHM)

In this section, a hybridized clustering algorithm is introduced that is combination of density-based clustering algorithm and partition-based clustering algorithm. The partition-based clustering algorithm is worked in this paper to get hold of the initial clusters from the complex data samples. From the obtained initial clusters, the density-based clustering algorithm is functioned to achieve the ending clusters. The K-Means algorithm look at the absolute the data points and initializes the clusters which in turn explores all the core objects of the DBSCAN algorithm. The Euclidean Distance function is used in this paper as to obtain the similarities and dissimilarities amongst the data points. The proposed Hybrid Genetic K-Means DBSCAN algorithm is described in six major stages given as:

i. Building Initial Clusters:

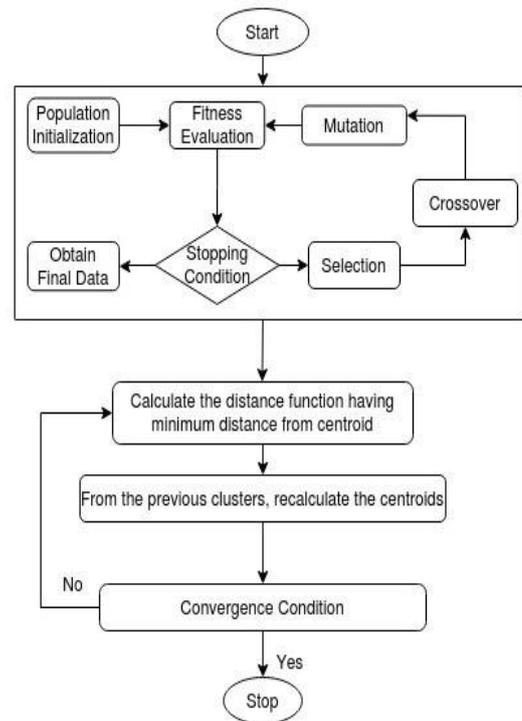


Fig 1: Flow Chart of Genetic K-Means Clustering Algorithm

The issue of DBSCAN algorithm exclusively lies in the assembly of initial clusters where the problem lies in the searching of core properties of objects are to some extent decided. Therefore, an Enhanced K-Means clustering algorithm is introduced in this section. The Enhanced K-Means algorithm is implemented by means of employing Genetic Algorithm as to get the initial centroid values. The Flow for the approach is given in Fig 1.

Procedure for Generating Initial Clusters using Genetic K-Means:



- A. **Initialization:** The data standards for clustering are chosen by means of genetic algorithm.
- The population for genetic algorithm is initialized by arbitrarily ranging n data points from the data sets.
 - These data points are estimated using the fitness function. Maximization of Euclidean distance between the data points is the fitness evaluation role given as the Euclidean distance $D(x_i, x_j)$ between the data points x_i, x_j as:

$$D(x_i, x_j) = \sqrt{\sum_{i=1}^d (x_i - x_j)^2}$$

- The individual data points are preferred using roulette wheel selection operation for additional computation.
- The crossover and mutation operation are executed on the preferred individual data points.
- The above steps are continued till the termination criteria are attained. The number of generations is considered as the termination criteria.
- The obtained first n data points are measured as centroids for advance clustering approach

B. **Clustering:** The distance is determined for every data point having minimum distance from the centroid of a cluster and individual data point from the centroid is consigned to that exact cluster.

C. **Centroid Recalculation:** Clusters produced already, the centroid is over and over computed by means of recalculation of the centroid.

D. **Convergence Condition:**

- When it reaches a given number of repetitions, then it is halted.
 - When there is no trade off data points between the clusters, then it is halted.
 - When a threshold value is acquired, then the algorithm is halted.
- E. If the above provisioned are not satisfied, then move to step 2 and the entire process same thing over, until the given provisions are not fulfilled.

Definition 1. (Object State) The position of an object s, indicated as $state(s)$, indicates the facts about s at a specified time T. Based on its core property, s can be defined as core, border or noise.

Note: Since K-means clustering algorithm is employed initially, there would be no untouched states in this step.

Definition 2. (Initial Clusters) At a particular time T, a core object $s \in O$ shared through accepted density-connected neighbour form the initial cluster, referred as $iclus(s)$, where s denotes cluster agent. If the initial cluster contains of only s and its ϵ -neighbourhood $N_\epsilon(s)$, referred as initial circle i.e. $icir(s)$.

II Formative States of Nodes in the Clusters:

The relationship amongst different initial clusters is confined in this stage. The primarily clusters are precisely connected at a specific time T as the chain of objects are associated to each other.

Definition 3. (Direct cluster connectivity) Two initiative clusters $iclus(s)$ and $iclus(a)$ are directly density-connected at a particular time T, identified as $iclus(s) \bowtie iclus(a)$, if $\exists M = \{m_1, m_2, m_3, \dots, m_n\} \in iclus(s) \cup iclus(a)$ such that $s \triangleleft m_1 \dots \triangleright m_n \triangleright a$. If given two initial circles $icir(s)$ and $icir(a)$ then there would be two cases:

- CaseA:
 $d(s, a) > 3\epsilon \Rightarrow \forall T: \neg icir(s) \bowtie icir(a)$
- CaseB:
 $d(s, a) > 3\epsilon \wedge |icir(s) \cap icir(a)| \geq \mu \Rightarrow icir(s) \bowtie icir(a)$

Definition 4. (Cluster graph) It is a graph $G = (V, E)$, here each vertex $x \in V$ corresponds to an initial cluster $iclus(x)$, and each edge (x, y) is assigned a state, signified as $state(x, y)$, that point toward the connectivity position of the two initial clusters $iclus(x)$ and $iclus(y)$.

- if $\forall T: \neg iclus(x) \bowtie iclus(y), (x, y) \in E, state(x, y) = No$
- else if $iclus(x) \bowtie iclus(y), state(x, y) = Yes$
else if $iclus(x) \cap iclus(y) \neq \emptyset, state(x, y) = Weak$
- else if $state(x, y) = unknown$

Generally, the graph G of each edge is put together two initial clusters that may connect to the correspondent cluster. And its state proceeds how accomplished the connection is. For example, at the particular time T, if $iclus(x)$ and $iclus(y)$ partition an object s and they are not precisely density-connected, $state(x, y)$ is thus weak. When contrast with unknown case they have additional opportunities to be present in the corresponding cluster. The edge state also modifies overtime. Suppose that at time $T + 1$, few extra queries notify that s is a core object, then $state(x, y)$ turn into yes at précised time $T + 1$ meanwhile $iclus(x)$ and $iclus(y)$ are now directly density-connected ensuing Definition 3. If $state(x, y)$ is No, it means that $iclus(x)$ and $iclus(y)$ will not be directly connected while all range queries have been accomplished. Thus, (x, y) does not fit in to the edge set E of G. To build the graph G, Proposed DBSCAN situated all the initials clusters obtained from stage 1 into V.

Lemma 1. Given two nodes $icir(x)$ and $icir(y)$ of G, if x and y are mostly density-connected, there should exist a path that associates $icir(x)$ and $icir(y)$ in G.

At the end of the Stage 1 and 2, initial circles are obtained by means of Enhanced K-Means Clustering. In the consequent moves, initial circles will be pooled to form the more general initial and stable clusters.

Dbscan Assisted by Hybrid Genetic K Means Algorithm

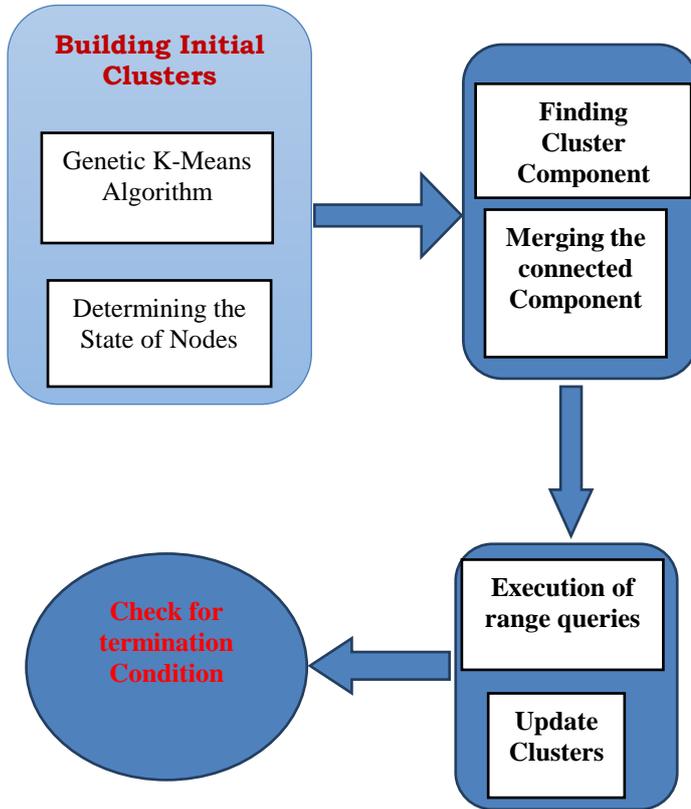


Fig 2: Block Diagram of Hybrid Genetic K-Means Based DBSCAN Algorithm

IV. FINDING AND MERGING THE CONNECTED COMPONENTS

Generally, DBSCAN depict sequence of exactly density-connected unrefined clusters through recognizing joined modules of the graph G_0 .

Definition 5. (Cluster connection graph) A cluster connection graph $G_0 = (V, E_0)$ is the subgraph of cluster graph G , where $E_0 = \{(x, y) | (x, y) \in E / state(x, y) = yes\}$. Given two nodes x and y in a associated modules C of G_0 , $iclus(x)$ and $iclus(y)$ belongs to the same cluster at a precise time T .

At a particular time, T , a midway clustering resulting of proposed DBSCAN can be created by recognizing all nodes of G_0 rendering to their related factors. Then objects are acknowledged equivalent to the labels of their advisory nodes. When graph G size is trivial, then label producing time is automatically decreased. So, a high quality, maximum is obtained.

All primitive clusters on a connected component are mutated in sync to decrease the number of nodes in G , so elevating the achievement, e.g., decreasing the time for identifying related modules in consecutives phases. For every individual module, inconstantly pick the model of a node inside it as a model for the entire cluster. For instance, then the total number of graph nodes are decreased as speed and efficient at the first few repetitions due to this convergence system.

Theorem 1. Each connected component $B \in G_0$ represented by an object sis an initial cluster attained after merging.

Lemma 2. Given two connected components $B = \{b_1, \dots, b_2\}$ and $D = \{d_1, \dots, d_2\}$ of G_0 at a particular time T . Then:

- Case A: $\forall b_i \in B, \forall d_i \in D: state(c_i, c_j) = No \Rightarrow state(B, D) = No$
- Case B: $\exists b_i \in B, \forall d_i \in D: state(c_i, c_j) = Weak \Rightarrow state(B, D) = Weak$
- Case C: otherwise $state(C, D) = unknown$

V. CHECKING FOR THE TERMINATION CRITERION:

Appreciating when to conclude the algorithm is essential for increasing the enforcement of proposed DBSCAN.

Theorem 2. At a specific time, T , if $\forall (x, y) \in E: state(x, y) = yes$, then $\forall T_0 > T \forall (x, y) \in E: state_T(x, y) = state_{T_0}(x, y)$, where $state_T(x, y)$ and $state_{T_0}(x, y)$ are the states of the edge (x, y) at time T and T_0 .

Ensuring Theorem 3, if all edges of graph G are states yes, the proposed approach can halt without inspecting all range queries since Graph g has no change the eventual clustering outcome following Corollary 3. As a result, its functioning is involuntarily improved.

VI. EXECUTION OF RANGE QUERIES

A naïve approach of executing range queries in DBSCAN is arbitrarily selecting an unrefined object. However, it is not necessary since there are no unrefined objects and need to additional develop the cluster. The general approach in this stage is letting the algorithm can run repetitively and keenly determine the existing cluster structure from the graph G and choose the objects that it reflects valuable for adjust the cluster structure. As reported in Theorem 3, we delete all inadequate and indeterminate edges from G as fast as possible, the early Proposed DBSCAN reaches the final state, so that, the lesser queries are obtained. To perform this, Proposed DBSCAN initially calculates the influence of each and every node of G . In this way, it ranks all noise or outlier objects confer to its current neighbors and its arrangements inside G . Those with maximum scores are selected as destination for achieving range queries.

Definition 6. (Node statistic) At a specific time T , the numerical knowledge of a node $\in V$, noted as $state(x)$, is denoted as follows:

$$state(x) = \frac{xsize(x)}{|iclus(x)|} + \frac{|iclus(u)|}{n}$$

where $xsize(x)$ is the number of noisy or outlier objects inside $iclus(x)$ and n is the number of objects.

Definition 7. (Node degree) Given a node x and its neighbouring nodes $N(x)$ in the graph G . The degree of x , denoted as $deg(x)$, at a specified time T is given as follows:

$$deg(x) = w \left(\sum_{y \in N(x) / state(x,y)=weak} state(y) \right) + \sum_{x \in N(x) / state(x,y)=unknown} state(y) - \psi(x)$$

Where $\psi(x)=0$ if x does not have border objects else it gives the count of not enough and indefinite edges of x . The degree of a node x calculates how the node x is with respect to its neighboring nodes. Naturally, high $deg(x)$ means that x remains inside a highly undefined area with number of uncertain connections. So, if a range query is achieved don't x , it has higher

transformations to either associate x and its neighboring nodes y ($state(x,y) = yes$) or to divide x from its adjacent nodes ($state(x,y) = No$). Moreover, if two nodes x and y have some common objects, they are exactly density connected. , Although if ($state(x,y) = unknown$), it is so difficult to find the true connection status of x and y .

So, a complicated weight $w = |V|$ for edges with slight positions than the edges with remaining positions. Even though, if node x having border objects, it will be fixed complex due to the result of Lemma 4 and the combining scheme of each most recent query for create a conclusion of the algorithm. The aim is enclosing an accomplished node with managed objects consequently any further queries can take it near to the vacant state as determined in Lemma 5.

Definition 8. (Object score) The score of a noisy and outlier object p , indicated as $score(s)$, at a particular time T is described as given:

$$score(s) = \sum_{y \in V \wedge s \in iclus(x)} deg(x) + \frac{1}{nei(s)}$$

where $nei(s)$ is the count of neighbours of s at a specified time T .

The count of an object p is précised depends on the entire degrees of all nodes $y \in V$ that whole and its present integers of neighbours. The node degrees described over, we prefer objects with less count of neighbours view of accomplishing range queries on them would indicate to more core objects to be reveal with each query. Since a has less neighbours, it is a highly unpredictable area. So, inspecting it earlier can suggest eliminating this faster.

If object s is choose for resulting range queries and if S is not a core object, it is identified as managed-border if it is nearby inside a cluster. Else, it is acknowledged as executed-core. And for all objects $a \in N_{\epsilon}(s)$, the status of a modifies subsequent the evaluation schema in Figure 3. Additionally, $N_{\epsilon}(s)$ is combined into all nodes x that holds.

Theorem 3: At a particular time, T , if $core(s) \wedge s \in iclus(x), N_{\epsilon}(s) \cup iclus(x)$ is a primitive cluster.

VII. UPDATING THE CLUSTER

In this method, the graph G is revised to retort amendment in the present cluster arrangement after forming the noval queries q following Definition 10, and Lemma 4 and 5 labeled below.

Lemma 3. Given two nodes x and y at a particular time T , if $xsize(x) = 0 \wedge xsize(x) = 0$ and $\neg iclus(x) \bowtie iclus(y)$, then $\forall T_0 > T: state(x,y) = no$, where $xsize(x)$ and $ysize(y)$ are the numbers of noise and outlier of $iclus(x)$ and $iclus(y)$.

According to Lemma 4, if the node x is entirely processed, all of its bordering nodes y , where $state(x,y) = weak$ or $unkown$, closed up having if their association is broken, i.e., deleted from E ($state(x,y) = No$). This drives the algorithm moving quicker to the halt situation described in Theorem 3.

Lemma 4. Given two primitive clusters $iclus(x)$ and $iclus(y)$, if

$$\exists k \in icur(x) \cap icur(y): state(k) = processed - core \vee state(k) = noise \text{ or outlier} \Rightarrow icur(x) \bowtie icur(y)$$

VIII. EXPERIMENTAL RESULTS AND ITS ANALYSIS

The Experimental Results for the proposed approach is carried out using two different images such as a cameraman and a lady. The proposed Hybrid Clustering Algorithm is evaluate with accessible approaches such as K-Means Clustering, Genetic based Clusterig and DB scan Clustering Algorithms. Fig 3 represents the measured Original image beside with the customized image by means of K-Means Clustering, DB Scan Clustering and Genetic Algorithm based Clustering Algorithm on image data sample.



Fig 3: Original Cameraman image with K-means, DB-Scan and Genetic algorithm based Clustering Algorithm



Fig 4: Clustered Cameraman image using Proposed Hybrid Clustering Algorithm with Clustering Radius and Threshold value 5

In Fig 4, the clustered cameraman image is shown which is fragmented by means of Proposed Clustering Algorithm. The approach is trialed with two distinctive clustering radii i.e. with radius 1 and 1.5 and with the Matching tolerance value or expanse threshold value of 1 as per the proposed methodology. From the figure, it can be contingency that the image is minutely sub divisional when contrast with other existing approaches as given in fig 3. The image with clustering radius 1 seen clearly with segmented position evaluate with the radius 1.5. Fig 5. represents the clustered image of cameraman by means of the proposed hybrid clustering algorithm with clustering radius 1 and using different Matching Tolerance Value or Threshold Value such values 6, 7, 8 and 9 respectively.



Fig 5: Clustered Cameraman image using Proposed Hybrid Clustering Algorithm with Clustering Radius 1 and different threshold values.

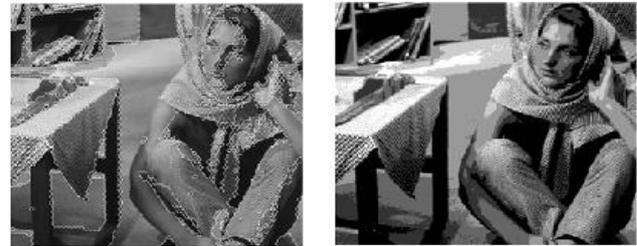


Fig 6: Original Lady image with K-means, DB-Scan and Genetic Algorithm based Clustering Algorithm

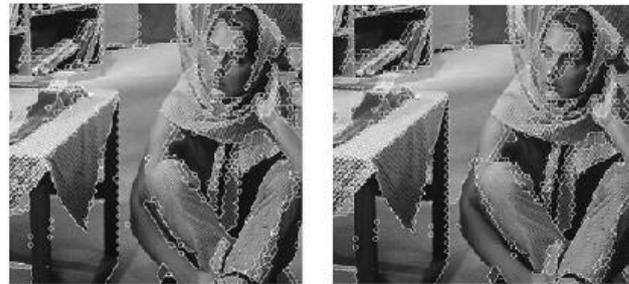


Fig 7: Clustered Lady image using Proposed Hybrid Clustering Algorithm with Clustering Radius and Threshold value 5

Fig 6 signifies the considered Original image along with the modified image by means of K-Means Clustering, DB Scan Clustering and Genetic Algorithm based Clustering Algorithm on image data sample. In Fig 7, the clustered cameraman image is shown which is divisional by means of Proposed Clustering Algorithm. The approach is investigation with two special clustering radii i.e. with radius 1 and 1.5 and with the Corresponding tolerance value or distance threshold value of 1 as per the proposed methodology. From the figure, it can be inferred that the image is minutely segmented when compared with other existing approaches as given in fig 6. The image with clustering radius 1 seen clearly with divisional position evaluate with the radius 1.5. Fig 8. signify the clustered image of cameraman by means of the proposed hybrid clustering algorithm with clustering radius 1 and using different Matching Tolerance Value or Threshold Value such values 6, 7, 8 and 9 respectively.



Fig 8: Clustered Lady Image using Proposed Hybrid Clustering Algorithm with Clustering Radius 1 and different threshold values.

IX. CONCLUSIONS

Clustering is one of the most well-organized techniques that is used in determine important knowledge for patterns. The density-based algorithm is hybridized with the partition-based algorithm as to explore the core properties from the initial clustering approach. The K- Means method is used to get hold of initial clusters. The partition-based method is engaged in this paper to get hold of the initial clusters from the complex data samples. From the attain initial clusters, the density-based method is functioned to attain the final clusters. The K-Means algorithm discover the complete the data points and initializes the clusters which in turn search all the core objects of the DBSCAN algorithm. The Euclidean Distance function is used in this paper as to get the similarities and dissimilarities along with the data points. The performance of the proposed methodology is evaluated using two various image samples and associated with the obtained clustering algorithms. From the outputs it is clearly shown, the proposed approach has higher performance relating to the existing ones.

REFERENCES

1. Wang J, Zhou Z. H., "Machine learning and its application", Beijing, Tsinghua university publisher, 2006.
2. Jain A., Murty M., Flynn P., "Data Clustering: A Review", ACM Computing Surveys, Vol. 31, No. 3, pp. 264-323, 1999.
3. Xu R., Wunsch D., "Survey of clustering algorithms", Transactions on Neural Networks, IEEE, 2005, Vol. 16, No. 3, pp. 645-678.
4. Hammouche K., Diaf M., Postaire J. G., "A clustering method based on multidimensional texture analysis", Pattern Recognition, 2006, Vol. 39, No. 1265-1277.
5. Vesanto J., "SOM-based data visualization methods", Intelligent Data Analysis, Vol. 3, pp. 111-126, 1996.
6. Ma S., Wang T. J., Tang S. W., "A fast clustering algorithm based on reference and density", Journal of Software, 2003, Vol. 14, pp. 1089-1095.
7. Ester, Martin, et al. "A density-based algorithm for discovering clusters in large spatial databases with noise." Knowledge data discovery, Vol. 96. 1996.
8. Agrawal, Jitendra, Sanyogita Soni, Sanjeev Sharma, and Shikha Agrawal, "Modification of Density Based Spatial Clustering Algorithm for Large Database Using Naive Bayes' Theorem", In Communication Systems and Network Technologies (CSNT), 2014 Fourth International Conference on, pp. 419-423. IEEE, 2014.

9. Viswanath, P., V. Suresh Babu. "Rough-DBSCAN: A fast hybrid density based clustering method for large data sets", Pattern Recognition Letters, Vol. 30, No.16, pp. 1477-1488, 2009.
10. C. Bohm, R. Noll, C. Plant, and B. Wackersreuther, "Density-based Clustering using Graphics Processors," In Proceedings Conf. on Information and knowledge management, Hong Kong, China, pp. 661-670, Nov. 2009.
11. VikasVerma, Shaweta Bhardwaj, Harjit Singh, "A Hybrid K-Mean Clustering Algorithm for Prediction Analysis", Indian Journal of Science and Technology, Vol. 9, No. 28, July 2016.
12. Tianjin, China, Sheng-Yi Jiang, Xia Li, "A Hybrid Clustering Algorithm", Fourth International Conference on Fuzzy Systems and Knowledge Discovery, 2009.
13. Vu Viet Thang, D. V. Pantiukhin, A. I. Galushkin, "A Hybrid Clustering Algorithm: The FastDBSCAN", International Conference on Engineering and Telecommunication (EnT), 18-19 Nov. 2015
14. Suresh kurumalla, P. Srinivasarao "k-nearest neighbor based dbscan clustering algorithm for image segmentation" Journal of Theoretical and Applied Information Technology 31st October 2016. Vol.92. No.2.
15. Suresh Kurumalla, P. Srinivasa Rao "An Improved k-means Clustering Algorithm for Image Segmentation" International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 13 (2015) pp 33143-33147.

AUTHORS PROFILE



Dr. K. Suresh, awarded Ph.D from JNTUK, Kakinada in 2019. He has received his Master's degree from Andhra University in 2008. He is currently working as assistant professor in Anil Neerukonda Institute of Technology & Sciences, Visakhapatnam since 2013. He is Member of CSI, ACM, MIE, IAENG.



Mrs. K. Chandusha, has received master's degree from JNTUK, Kakinada in 2017. She is currently working as assistant professor in Anil Neerukonda institute of Technology & Sciences, Visakhapatnam. She is Member of IAENG.