ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

IJESRT
INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

HYBRID GENETIC K-MEANS ASSISTED DENSITY BASED CLUSTERING ALGORITHM

Suresh Kurumalla[1], P.Srinivasa Rao[2]
[1]Research Scholar in CSE Department, JNTUK Kakinada
[2]Professor, CS&SE Department, Andhra University, Visakhapatnam, AP, India

DOI: 10.5281/zenodo.1419501

ABSTRACT
The data mining algorithms performance is key issue, when the data becomes more and more. Clustering analysis is a dynamic and challenge research direction in the area of data mining for compound data samples. DBSCAN is a density based clustering algorithm with numerous advantages in numerous applications. However, DBSCAN has quadratic time complexity i.e. $O(n^2)$ making it difficult for practical applications especially with large complex data samples. Therefore, this paper suggested a hybrid approach to minimize the time complexity by exploring the core properties of the DBSCAN in the initial stage using genetic based K-means partition algorithm. The scientific experiments showed that the proposed hybrid approach obtains competitive results when compared with the traditional approach and significantly improves the computational time.

KEYWORDS: Density based Clustering Algorithm, DBSCAN, Genetic Algorithm, K-Means algorithm, Image database.

. 1. INTRODUCTION
Clustering analysis comes under unsupervised learning and it plays a key role in research field of machine learning and data mining [1]. Clustering is one of the efficient techniques that issued in finding important knowledge for patterns. The fundamental objective for clustering is to dividing the specified dataset into clusters so that objects which are present in a cluster have high equivalence in compare with one another and contrast to objects in remaining clusters. Clustering is an essential data mining issue that is usually begins in various areas as well as biology, social science and marketing. Fast and efficient clustering algorithms perform a key role in given inherent exploration and browsing mechanisms by arranging high volumes of data into less number of valid clusters.

Researchers have conferred different clustering algorithms [2, 3], the fundamental division on cluster algorithms is partitioning clustering, and hierarchical clustering, grid and density based clustering. In some cases, these algorithms can stimulate clustering speed and upgrade effectiveness by taking of feature selection [4], reducing the count of dimensions [5], or applying reference points [6], etc. Even though there is a specific inconsistency and complications in cluster analysis by certain algorithms, so that the inconsistency among cluster definiteness and effectiveness, average of primary value setting and so on. Cluster analysis has applications in various sectors of business and science, Data reduction, Hypothesis generation, Hypothesis testing, Prediction based on groups,, Biology, Spatial data analysis, Web mining.

In density-based clustering, the clusters are begin with the objects associating slowly and detached by inadequate regions. It has the specified benefits over remaining clustering algorithms [6, 7]. First, it recovers the clusters of random shapes apart from curved shapes. Second, it is effectively removes the noise. Third, it does not need a pre-described number of clusters unlike k-means [6]. Fourth, the clustering output is not damaged by the input order of objects. However, there also certain 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 better performance [3,12].

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

DBSCAN algorithm is generally 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 partially 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 explore the core properties from the initial clustering approach. The K- Means algorithm is used to obtain initial clusters.

1.1 Organization of the Paper

A brief 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 different types clustering techniques and survey done on these approaches. The proposed hybrid Genetic K-Means based DBSCAN algorithm is concisely 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.

2. LITERATURE SURVEY

Bohm et al[10] suggested CUDA-DClust algorithm. It enhance the consecutive DBSCAN up to15 times using the GPU. They also suggested CUDA-DClust that enhanced CUDA-DClust up to 11.9 times using a simple catalog structure. In [14], a methodology is presented 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 select points for DBSCAN clustering at second step. Experiments show that our method obtains competitive results with the original DBSCAN, while significantly improving the computational time.

The suggested 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 a future based service. This is closest to the outlier cluster data. The amount of definiteness of algorithm raised on a actual threshold value P. Viswanath et al[9] suggested a technique that is obtained a prototype identified as leader from a dataset that has the data of the prototypes along with density. These are used to form density based clusters.

The fundamental disadvantage of k-mean is efficiency, as user must specify the no. of clusters during the initial process. This constraint of predefined number of clusters tends some points of the dataset to remain un-clustered. So by enlarging the cluster methods, the predictions can be enhanced. Therefore, in [11], the normalization is used to obtain accurate results by calculating distance to have definite centroid and to eliminate noise data. Backtracking fashion is adopted to find the definite figure of clusters that is defined to analyze the data in better way. The results showed that there is a development in clustering when correlated to the actual procedures.

A prototype-based hybrid approach [12] is proposed to accelerate the k-means method. The data-set is initially divided into insignificant clusters of differing sizes. Then, the fixed prototype is divided into k clusters applying the modified k-means method. This clustering method is same to the conventional k-means method by removing vacant clusters in the repeated technique. In every cluster of model, every model is regained by its parallel set of patterns to determine the separation of data-set. So, this separation 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 suggested method is correlated with the standard method and the other current methods that are suggested to accelerate the k-means method.

A two-stage hybrid clustering algorithm is proposed in [13] where DBSCAN is enhanced to progress the data with definite aspects. By merging DBSCAN clustering algorithm and one-pass clustering algorithm then two-stage hybrid clustering algorithm is obtained. In the fundamental step, to group the data one pass clustering algorithm is used. In the second step, improved DBSCAN clustering algorithm is merged with the partition then only ultimate clusters are gathered. The given clustering algorithm is of closely linear time complexity, it is used to progress the extensive datasets. The experimental result on actual datasets and synthetic datasets displays that the two-stage hybrid clustering algorithm can assist to analyze the data with random shape similar to DBSCAN, the operating effectiveness of which is not only better than the DBSCAN, but also efficient and capable.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

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

In this section, a hybridized clustering algorithm is introduced that is amalgamation of density based clustering algorithm and partition based clustering algorithm. The partition based clustering algorithm is employed in this paper to obtain the initial clusters from the complex data samples. From the obtained initial clusters, the density based clustering algorithm is functioned to attain the final clusters. The K-Means algorithm explores the complete 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 get 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**:

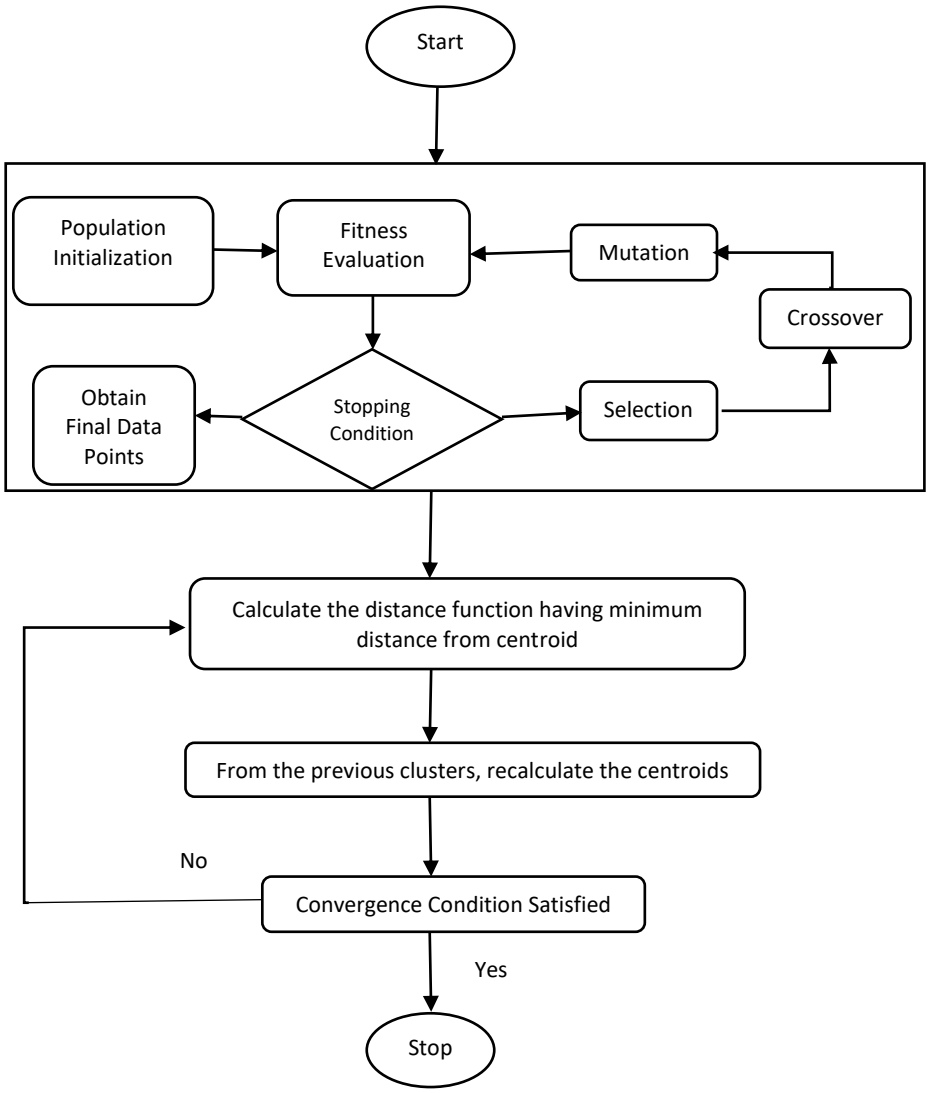

Fig 1: Flow Chart of Genetic K-Means Clustering Algorithm

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

The issue of DBSCAN algorithm specifically lies in the construction of initial clusters where the problem lies in the exploration of core properties of objects are slightly 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 obtain 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 values for clustering are selected by means of genetic algorithm.
   a. The population for genetic algorithm is initialized by randomly selecting n data points from the data sets.
   b. These data points are evaluated using the fitness function. Maximization of Euclidean distance between the data points is the fitness evaluation function given as the Euclidean distance $D(x_i, x_j)$ between the data points x_i and x_j as: $D(x_i, x_j) = \sqrt{\sum_{i=1}^{d}(x_i - x_j)^2}$
   c. The individual data points are selected using roulette wheel selection operation for further computation.
   d. The crossover and mutation operation is implemented on the selected individual data points.
   e. The above steps are repeated till the termination criteria are reached. The number of generations is considered as the termination criteria.
   f. The obtained first n data points are considered as centroids for further 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 particular cluster.

C. Centroid Recalculation: Clusters produced already, the centroid is again and again 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 swapping of data points between the clusters, then it is halted.
   - When a threshold value is obtained, then the algorithm is halted.

E. If the above specifications are not fulfilled, then move to step 2 and the entire process same thing over, until the given specifications are not satisfied.

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 $s \in O$ combined through accepted density-connected neighbor form the initial cluster, referred as $iclus(s)$, where s denotes cluster agent. If the initial cluster contains of only s and its $\in$ −neighborhood $N_{\in}(s)$, referred as initial circle i.e. $icir(s)$.

ii. **Determining States of Nodes in the Clusters**:

The relationship amongst different initial clusters is captured in this stage. The initially clusters are directly connected at a particular time T as the chain of objects are connected 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$\exists M = \{m_1, m_2, m_3, \ldots . m_n\} \in iclus(s) \cup iclus(a)$such that$s \lhd m_1 \ldots \rhd m_n \rhd a$.If given two initial circles $iclus(s)$ and $iclus(a)$then there would be two cases:

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

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

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, denoted as$state(x, y)$, that indicates the connectivity status of the two initial clusters$iclus(x)$and$iclus(y)$.

- $if\ \forall T: \neg iclus(x) \bowtie iclus(y), (x, y) \nexists E(state\ (x, y) = No)$
- $else\ if\ iclus(x) \bowtie iclus(y), state(x, y) = Yes$
- $else\ if\ iclus(x) \cap iclus(y)\ \neq 0, state(x, y) = Weak$
- $else\ if\ state(x, y) = unknown$

Usually, The graph G of each edge is attaches two initial clusters that may associate to the equivalent cluster. And its state returns how capable the connection is. For example, at the particular time T, if$iclus(x)$ and $iclus(y)$ divide an object s and they are not directly density-connected, $state\ (x, y)$ is thus weak. When compare with unknown case they have further 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 specified time $T\ + 1$ meanwhile$iclus(x)$and $iclus(y)$are now directly density-connected resulting 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 belong to the edge set $E$of$G$. To build the graph G, Proposed DBSCAN locates 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 mustly 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 subsequent moves, initial circles will be combined to form the more general initial and stable clusters.

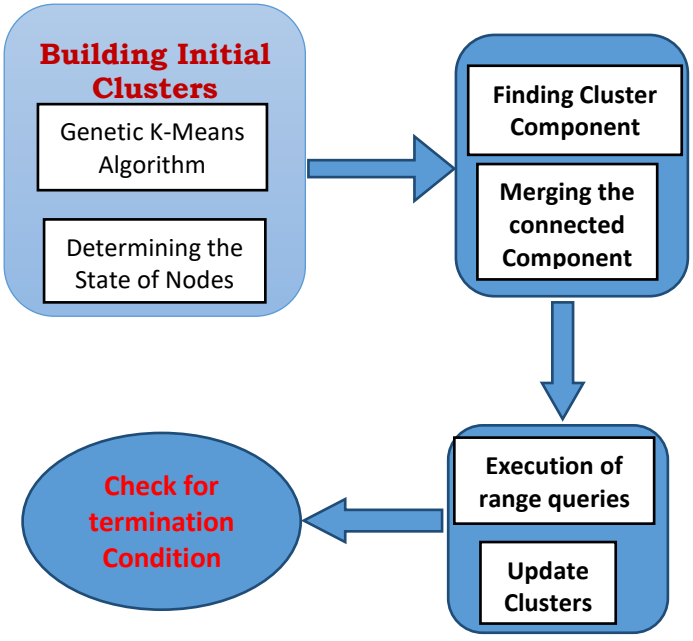

Fig 2: Block Diagram of Hybrid Genetic K-Means Based DBSCAN Algorithm

iii. **Finding and Merging the Connected Components**:
Usually DBSCAN describes sequence of exactly density-connected unsophisticated clusters through identifying joined modules of the graphG_0.

Definition5. (Cluster connection graph) A cluster connection graph$G_0 = (V, E_0)$is the subgraph of cluster graph$G$, where$E_0 = \{(x,y)|(x,y) \in E \wedge state(x,y) = yes\}$.Given two nodes x and y in a connected modules C ofG_0, $iclus(x)$and$iclus(y)$belongs to the same cluster at a specified time T.

At a specified time T, a midway clustering outcome of proposed DBSCAN can be created by identifying all nodes of G0 rendering to their associated factors. Then objects are identified corresponding to the labels of their illustrative 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 associated component are combined 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$B = \{b_1, \ldots . b_2\}andD = \{d_1, \ldots . d_2\}ofG_0$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: $otherwise\ state(C, D) = unknown$

iv. **Checking for the Termination Criterion**:

Understanding when to terminate the algorithm is necessary 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 ultimate clustering outcome following Corollary 3. Asa result, its functioning is automatically improved.

v. **Execution of Range Queries**:

A naïve approach of executing range queries in DBSCAN is randomly selecting an unprocessed object. However, it is not essential since there are no unprocessed objects and need to further exploit the cluster. The general approach in this stage is letting the algorithm can run repeatedly and passionately determine the existing cluster structure from the graph G and select the objects that itreflectsvaluable for modifying the cluster structure. As reported in Theorem 3, we delete all inadequate and unspecified 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 particular 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 neighboring nodes $N(x)$ in the graph G. The degree of x, denotedas $\deg(x)$, at a specified time T is given as follows:

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

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

Where$\psi(x)$=0if x does not have border objects elseit gives the count of inadequate and unspecified 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 on$x$, 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 directly density connected. , Although if$(state(x,y) = unkown)$, it is so difficult to find the true connection status of x and y.

So, a sophisticated weight $w = |V|$ for edges withdelicate positions than the edges with remaining positions. Even though, if node x having border objects, it will be fixed advanced due to the result of Lemma 4 and the combining scheme of each latest query for establishing an conclusion of the algorithm. The aim is enclosing a capable 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 an 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 neighbors of s at a specified time T.

The count of an object p is measured depends on the total degrees of all nodes $y \in V$ that hold$s$ and its presentintegers of neighbors. The node degrees described over, we choose objects with less count of neighbors in view of achieving range queries on them would indicate to more core objects to be disclosed with each query. Since$a$ has less neighbors, it is a highly unreliable area. So, inspecting it earlier can help to remove this faster.

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

> *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.*

vi. **Updating the Cluster**:

In this method, the graph G is revised to react modifications in the present cluster arrangement after forming the novel 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 completely processed, all of its bordering nodes y, where$state(x,y) = weak\ or\ unkown$, closed up having if their associations 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$\exists k \in icir(x) \cap icur(y): state(k) = processed - core \vee state(k) = noise\ or\ outlier \Rightarrow icir(x) \bowtie icir(y)$.

5. 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 compared with existing approaches such as K-Means Clustering, Genetic based Clusterig and DB scan Clustering Algoithms. Fig 3 represents 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.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

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 segmented by means of Proposed Clustering Algorithm. The approach is experimented with two different clustering radius i.e. with radius 1 and 1.5 and with the Matching 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 3. The image with clustering radius 1 seen clearly with segmented position compared 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.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

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

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

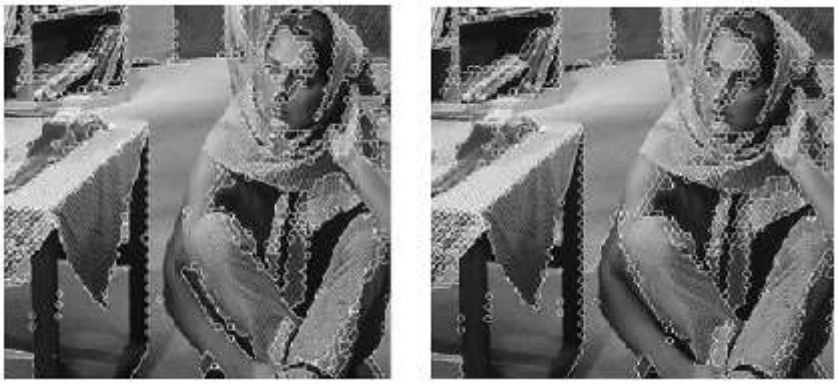

Fig 7: Clustered Lady Image using Proposed Hybrid Clustering Algorithm with Clustering Radius and Threshold value 5

Fig 6 represents 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 segmented by means of Proposed Clustering Algorithm. The approach is experimented with two different clustering radius i.e. with radius 1 and 1.5 and with the Matching 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 segmented position compared with the radius 1.5. Fig 8. 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 8: Clustered Lady Image using Proposed Hybrid Clustering Algorithm with Clustering Radius 1 and different threshold values

6. CONCLUSIONS

Clustering is one of the most efficient techniques that isused in discovering 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 obtain initial clusters.The partition based method is employed in this paper to obtain the initial clusters from the complex data samples. From the obtained initial clusters, the density based method is functioned to attain the final clusters. The K-Means algorithm explores the complete 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 get the similarities and dissimilarities amongst the data points. The performance of the proposed methodology is compared using two different image samples and related 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, SanyogitaSoni, 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] T. HitendraSarma, P. Viswanath, B. Eswara Reddy, "A hybrid approach to speed-up the k-means clustering method", International Journal of Machine Learning and Cybernetics, Vol. 4, No. 2, pp. 107–117, April 2013.
[13] Tianjin, China, Sheng-Yi Jiang, Xia Li, "A Hybrid Clustering Algorithm", Fourth International Conference on Fuzzy Systems and Knowledge Discovery, 2009.
[14] 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
[15]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.
[16]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.

CITE AN ARTICLE

Kurumalla, S., & Rao, P. S. (2018). HYBRID GENETIC K-MEANS ASSISTED DENSITY BASED CLUSTERING ALGORITHM. *INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY, 7*(9), 209-219.