An Empirical Evaluation of Density-Based Clustering Techniques

Glory H. Shah, C. K. Bhensdadia, Amit P. Ganatra

Abstract*— Emergence of modern techniques for scientific data collection has resulted in large scale accumulation of data pertaining to diverse fields. Conventional database querying methods are inadequate to extract useful information from huge data banks. Cluster analysis is one of the major data analysis methods. It is the art of detecting groups of similar objects in large data sets without having specified groups by means of explicit features. The problem of detecting clusters of points is challenging when the clusters are of different size, density and shape. The development of clustering algorithms has received a lot of attention in the last few years and many new clustering algorithms have been proposed. This paper gives a survey of density based clustering algorithms. DBSCAN [15] is a base algorithm for density based clustering techniques. One of the advantages of using these techniques is that method does not require the number of clusters to be given a prior nor do they make any kind of assumption concerning the density or the variance within the clusters that may exist in the data set. It can detect the clusters of different shapes and sizes from large amount of data which contains noise and outliers. OPTICS [14] on the other hand does not produce a clustering of a data set explicitly, but instead creates an augmented ordering of the database representing its density based clustering structure. This paper shows the comparison of two density based clustering methods i.e. DBSCAN [15] & OPTICS [14] based on essential parameters such as distance type, noise ratio as well as run time of simulations performed as well as number of clusters formed needed for a good clustering algorithm. We analyze the algorithms in terms of the parameters essential for creating meaningful clusters. Both the algorithms are tested using synthetic data sets for low as well as high dimensional data sets.

***Index Terms*— DBSCAN, OPTICS, DENCLUE, Spatial Data, Intra Cluster, Inter Cluster.**

I. INTRODUCTION

Clustering is an initial and fundamental step in data analysis. It is an unsupervised classification of patterns into groups or we can say clusters. Intuitively, patterns within a valid cluster are more similar to each other and dissimilar when compared to a pattern belonging to other cluster. Clustering is useful in several fields such as pattern analysis, machine learning situation also pattern classification and many other fields.

Clustering can be classified into five major types – Partitioned, Hierarchical, Density-Based, Grid-Based and Model-Based methods. Fig. 1 shows the detailed clustering methods along with its subtypes.

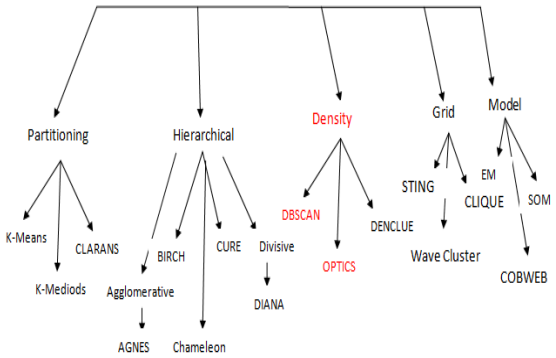

Fig 1 Types of Clustering Techniques

In *Partitioning Method*, given a database of n objects it constructs k partitions of the data, where each partition represents a cluster and k<=n. That is, it classifies the data into k groups, which together satisfy the following requirements: [1] each group must contain at least one object and [2] each object must belong to exactly one group.

In *Hierarchical Method*, it creates a hierarchical decomposition of the given set of data objects. It can be either agglomerative or divisive, based on how hierarchical decomposition is formed. The *agglomerative approach*, also called the *bottom-up* approach, starts with each object forming a separate group. It successively merges the objects or groups that are close to one another, until all of the groups are merged into one (the topmost level of the hierarchy), or until a termination condition holds. The *divisive approach*, also called the *top-down* approach, starts with all of the objects in the same cluster. In each successive iteration, a cluster is split up into smaller clusters, until eventually each object is in one cluster, or until a termination condition holds.

In *Density-Based Method*, most partitioning methods cluster objects based on the distance between objects. Such methods can find arbitrary shaped clusters. The general idea here is to continue growing the given cluster as long as the density or say number of objects or data points in the

Glory H. Shah, Computer Engineering at Dhramsinh Desai University, Nadiad and Assistant Professor at Charotar University of Science Technology (CHARUSAT), Education Campus, Changa, Gujarat, India. E-mail: glory.ce2006@gmail.com.

C. K. Bhensdadia, Professor & Head at Department of Computer Engineering, Faculty of Technology, Dharmsinh Desai University, Nadiad, Gujarat, India. E-mail: ckbhensdadia@yahoo.co.in.

Amit P. Ganatra, Associate Professor at Charotar University of Science Technology (CHARUSAT), Education Campus, Changa, Gujarat, India E-mail: amitganu@yahoo.com.

neighborhood exceeds some threshold. Such methods can be used to filter our noise or outliers.

In *Grid-Based Method*, it quantizes the object space into a finite number of cells that form a grid structure. The main advantage of this approach is its fast processing time, which is independent of number of data objects and dependent on the number of cells in each dimension in the quantized space.

In *Model-Based Method*, it hypothesizes a model for each of the clusters and finds the best fit of the data to given model.

II. LITERATURE SURVEY

Spatial Clustering is an active research area in spatial data mining with various methods reported.

In [22], Xin et al. and Howard et al. made a comparative analysis of two density-based Clustering algorithms i.e. DBSCAN and DBRS which is a density-based clustering algorithm. They concluded that DBSCAN gives extremely good results and is r is efficient in many datasets. However, if a dataset has clusters of widely varying densities, than DBSCAN is not able to perform well. Also DBRS aims to reduce the running time for datasets with varying densities. It also works well on high-density clusters.

In [20] Mariam et al. and Syed et al. made a comparision for two density-based clustering algorithm i.e. DBSCAN and RDBC i.e. Recursive density based clustering. RDBC is an improvement of DBSCAN. In this algorithm it calls DBSCAN with different density distance thresholds ε and density threshold MinPts. It concludes that the number of clusters formed by RDBC is more as compared to DBSCAN also we see that the runtime of RDBC is less as compared to DBSCAN.

In [1] K.Santhisree et al. described a similarity measure for density-based clustering of web usage data. They developed a new similarity measure
named sequence similarity measure and enhanced DBSCAN [14] and OPTICS [15] for web personalization. As an experimental result it was found that the average intra cluster distance in DBSCAN is more as compared to OPTICS and the average intra cluster distance is minimum in OPTICS.

In [17] K.Mumtaz et al. and Dr. K.Duraiswamy described an analysis on Density-Based Clustering of Multi-Dimensional Spatial Data. They showed the results of analyzing the properties of density-based clustering characteristics of three clustering algorithms namely DBSCAN, k-means and SOM using synthetic two dimensional spatial data sets. It was seen that DBSCAN performs better for spatial data sets and produces the correct set of clusters compared to SOM and k-means algorithm.

In [10] A.Moreia, M.Santos and S.Corneiro et al. described the implementation of two density based clustering algorithms: DBSCAN [15] and SNN [12]. The no of input required by SNN is more as compared to DBSCAN. The results showed that SNN performs better than DBSCAN since it can detect clusters with different densities while the former cannot.

III INTRODUCTION TO DENSITY-BASED CLUSTERING TEQNIQUES

A. Background Study

Density based clustering is to discover clusters of arbitrary shape in spatial databases with noise. It forms clusters based on maximal set of density connected points. The core part in Density-Based clustering is density-reach ability and density connectivity. Also it requires two input parameters i.e. Eps which is known as radius and the MinPts i.e. the minimum number of points required to form a cluster. It starts with an arbitrary starting point that has not visited once. Then the ε - neighborhood is retrieved, and if it contains sufficiently many points than a cluster is started. Otherwise, the point is labeled as noise. This section describes two density based clustering algorithms briefly i.e. DBSCAN (Density Based Spatial Clustering of Application with Noise) and OPTICS (Ordering Points to Identify the Clustering Structure).

Here,

Density=number of points within a specified radius.

Density-Based clustering Algorithms mainly include three techniques:

- DBSCAN [15] which grows clusters according to a density-based connectivity analysis.
- OPTICS [14] extends DBSCAN to produce a cluster ordering obtained from a wide range of parameter settings.
- DENCLUE [24] clusters objects based on a set of density distribution functions.

Table 1 Comparison of Density-Based Clustering Methods

Name	Noise	Varied Density	Primary Input Required	Complexity	Data Type	Cluster Type	Data Set
DBSCAN	Yes	No	Cluster radius, Minimum no. of Objects	O(nlogn)	Numerical	arbitrary	High-Dim ensional
OPTICS	Yes	Yes	Density Threshold	O(nlogn)	Numerical	arbitrary	High-Dim ensional
DENCLUE	Yes	Yes	Radius	$O(n^2)$	Numerical	arbitrary	High-Dim ensional

B. DBSCAN Algorithm

This algorithm grows regions with sufficiently high density into clusters and discovers clusters of arbitrary shape in spatial databases with noise. Some definitions to be known before understanding the algorithm are as follows:

Definition 1: A *noise point* is any point that is not a core point or a border point. Noise points are discarded.

Definition 2: A ε - *neighbourhood* is objects within a radius of ε from an object.

Definition 3: *Core Objects*, if the ε - neighbourhood of an object contains at least a minimum number MinPts of objects then the object is called a core object.

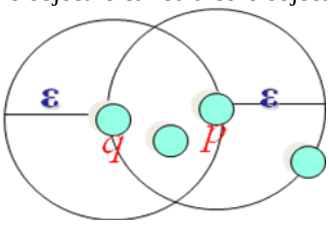

Fig 2 ε - neighbourhood of p, q. p is core object with MinPts=4

Definition 4: *Directly density-reachable*: An object q as shown in fig. 3 is directly density-reachable from object q if q is within the ε - neighbourhood of p and p is a core object.

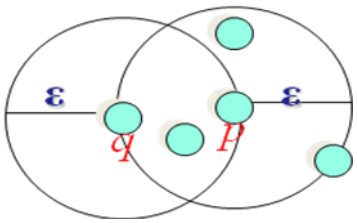

Fig 3 q is directly density reachable from p

Definition 5: *Density-Reachable*: An object pas shown in fig. 4 is density-reachable from q w.r.t ε and MinPts if there is a chain of objects $p_1,\ldots,p_n$, with $p_1=q$, $p_n=p$ such that $p_{i+1}$is directly density-reachable from p_i w.r.t ε and *MinPts* for all $1 <= i <= n$

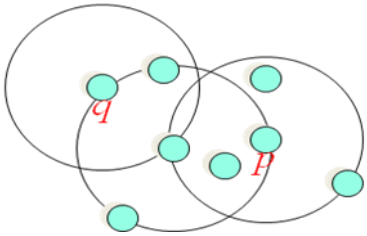

Fig 4 q is density-reachable from p

Definition 6: *Density-Connectivity*: Object p as shown in fig. 5 is density-connected to object q w.r.t ε and MinPts if there is an object o such that both p and q are density reachable from o w.r.t ε and MinPts

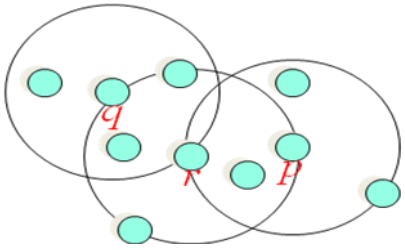

Fig 5 p and q are density-connected to each other by r

DBSCAN [15] is an important and widely used technique for class identification in spatial databases. The explanation of algorithm can be summarized as below:

- Select a point ***p.***
- Retrieve all points density-reachable from ***p*** w.r.t ε and MinPts.
- If ***p*** is a core point, a cluster is formed.
- If ***p*** is a border point, no points are density-reachable from ***p*** and it visits the next point of the database.
- Continue the process until all the points have been processed.

Advantages of DBSCAN are as follows:

Most of the clustering methods use distance as a measure between two clusters, which fails in detecting arbitrary shaped clusters. DBSCAN [15] can detect arbitrary shaped clusters, which is the main feature of this technique in identifying clusters.

1. DBSCAN does not require you to know the number of clusters in the data a priori, as opposed to k-means.
2. DBSCAN can find arbitrarily shaped clusters. It can even find clusters completely surrounded by (but not connected to) a different cluster. Due to the MinPts parameter, the so-called single-link effect (different clusters being connected by a thin line of points) is reduced.
3. DBSCAN has a notion of noise.
4. DBSCAN requires just two parameters and is mostly insensitive to the ordering of the points in the database.

Disadvantages include:

As the first density-based clustering algorithm that discovers clusters with arbitrary shape and outliers, DBSCAN has

certain limitations, which are listed below:

- It is not easy to determine proper initial values of Eps and MinPts. Even in the same database, when the number of samples is changed, the two parameters have to be adjusted accordingly.
- The computational complexity of DBSCAN without any special structure is $O(n^2)$, where n is the number of data objects. If a spatial index is used, the complexity can be reduced to O (nlogn). However, the task of building a spatial index is time-consuming and less applicable to high dimensional data sets.

C. OPTICS

While the partitioning density-based clustering algorithm DBSCAN [15] can only identify a "flat" clustering, the newer algorithm OPTICS [14] computes an ordering of the points augmented by additional information, i.e. the reachabilitydistance, representing the intrinsic hierarchical (nested) cluster structure. The result of OPTICS [14], i.e. the cluster ordering, is displayed by the so-called reachability plots which are 2D-plots generated as follows: the clustered objects are ordered along the *x*-axis according to the cluster ordering computed by OPTICS [14] and the reachabilities assigned to each object are plotted along the abscissa. An example reachability plot is depicted in Fig. 6. Valleys in this plot indicate clusters: objects having a small reachability value are closer and thus more similar to their predecessor objects than objects having a higher reachability value.

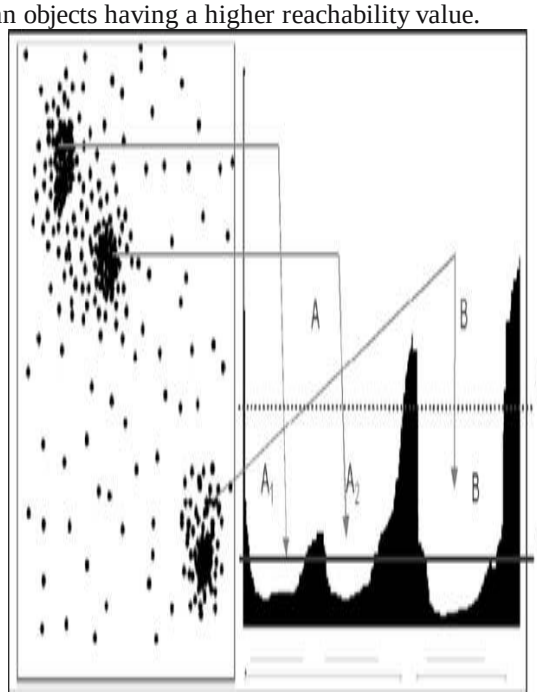

Fig. 6 Reachability plot (*right*) computed by OPTICS for a 2D data set (*left*)

Thus it is possible to explore interactively the clustering structure, offering additional insights into the distribution and correlation of the data. This section shortly introduces the definitions underlying the OPTICS algorithm, the core-distance of an object *p* and the reachability-distance of an object *p* w.r.t. a predecessor object *o*.

Definition 1: *Core-distance***:** Let *p* be an object from a database DB, let *Nε(p)* be the *ε*-neighborhood of *p*, let MinPts be a natural number and let MinPts-dist*(p)* be the distance of *p* to its MinPts-the neighbor. Then, the *core-distance* of *p*, denoted as core-dist *ε* , MinPts*(p)* is defined as MinPts-dist*(p)* if *|Nε(p)* |≥ MinPts and INFINITY otherwise. This is illustrated in fig. 7.

Definition 2: *Reachability-distance***:** Let *p* and *o* be objects from a database DB, let *Nε(o)* be the *ε*-neighborhood of *o*, let dist*(o, p)* be the distance between *o* and *p*, and let MinPts be a natural number. Then the *reachabilitydistance* of *p* w.r.t. *o* as shown in fig. 7, denoted as reachability-dist *ε,* MinPts*(p, o)*, is defined as max*(*core-dist *ε,* MinPts*(o)*, dist*(o, p))*.

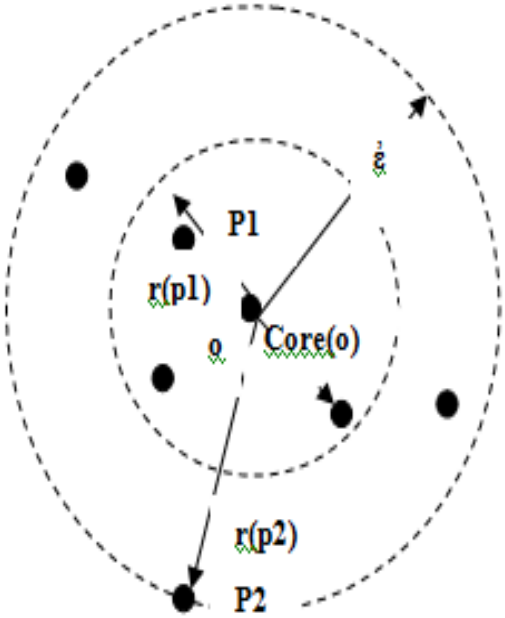

Fig. 7 Core-distance(*o*), reachability-distances r(p_1,*o*), r(p_2,*o*) for *MinPts*=4

The OPTICS [14] algorithm creates an ordering of a database, along with a reachability-value for each object. Its main data structure is a *seedlist*, containing tuples of points and reachability-distances. The seedlist is organized w.r.t. ascending reachability-distances. Initially the seedlist is empty and all points are marked as *not-done*.

IV EXPERIMENTS AND RESULTS

This section illustrates the comparative study of two density based clustering algorithms i.e. DBSCAN[15] and OPTICS[14] based on essential parameters which includes type of database used, type of distance used, no. of clusters formed, and time taken to form a cluster, unclustered instances as well as the content of noise found. This evaluation is done on Low to high Dimensional Data Sets. The datasets used for experimental evaluation are in ARFF (Attribute-Relation File Format) format which is an ASCII text file that describes a list of instances sharing a set of attributes. Table 2 describes the information about datasets used for experimental purpose.

The experiment is carried out in data mining tool WEKA 3.6.Table 3 describes the working of DBSCAN [15] algorithm, with ε= 1.2 and MinPts=2 and distance type=EUCLIDIAN distance. Table 4 describes working of DBSCAN [15] but with using ε= 1.2 and MinPts=2 and distance type=MANHATTAN. Table 5 describes OPTICS [14] with using ε= 1.2 and MinPts=2 distance type=EUCLIDIAN and table 6 also describes OPTICS [14] with using ε= 1.2 and MinPts=2 distance type=MANHATTAN. Study is based on the following features of the algorithm.

- No of Samples for each data set used for experiment.
- No of Clustered as well as unclustered instances.
- Noise Ratio which can have either of the values i.e. High, Very High, Less, Very Less, No noise and almost negligible means there is noise but only some percent.
- Time taken when distance is changed

Table 2 Data Set Information

Data Set	# Attributes	# Instances	# Nominal Attr.	# Numerical Attr.	# Classes
Heart-Stat log	14	270	1	13	1
Arrhythmia	280	452	70	210	1
Kr-vs-kp	37	3196	37	0	1
Waveform	41	5000	1	40	1
Ipums_la_98-small	61	7485	61	7485	0

Example 1: DBSCAN clustering technique:

Table 3 DBSCAN using ε= 1.2 and MinPts=2 distance type=EUCLIDIAN

DBSCAN	CLUSTERING RESULTS USING EUCLIDIAN				
No. Of Samples	270	452	3196	5000	7485
No of Clusters formed	1	16	158	1	16
No of Unclustered Instance	2	316	159	0	7453
Noise Level	Almost Negligible	Very Less	Less	No	High
Time Taken(Min.Sec)	0.18	15.36	18.46	64.37	1852.18
DataBase Type	Sequential	Sequential	Sequential	Sequential	Sequential

Table 4 DBSCAN using ε= 1.2 and MinPts=2 distance type=MANHATTAN

DBSCAN	CLUSTERING RESULTS USING MANHATTAN				
No. Of Samples	270	452	3196	5000	7485
No of Clusters formed	27	0	158	0	16
No of Unclustered Instance	80	452	159	5000	7453
Noise Level	Very High	High Compared To Euclidian	High Compared To Euclidian	All Noise	High
Time Taken(Min.Sec)	0.07	9.42	12.79	34.23	1813.65
DataBase Type	Sequential	Sequential	Sequential	Sequential	Sequential

Example 2: OPTICS clustering technique:

Table 5 OPTICS using ε= 1.2 and MinPts=2 and distance type=EUCLIDIAN

OPTICS	CLUSTERING RESULTS USING EUCLIDIAN				
No. Of Samples	270	452	3196	5000	7485
No of Clusters formed	0	0	0	0	0
No of Unclustered Instance	270	452	3196	5000	7485
Noise Level	Very Low	Very Low	Very High	No	Very High
Time Taken(Min.Sec)	0.06	3.31	18.64	301.16	7365.66

DataBase Type	Sequential	Sequential	Sequential	Sequential	Sequential

Table 6 OPTICS using ε= 1.2 and MinPts=2 distance type=MANHATTAN

OPTICS	CLUSTERING RESULTS USING MANHATTAN				
No. Of Samples	270	452	3196	5000	7485
No of Clusters formed	0	0	0	0	0
No of Unclustered Instance	270	452	3196	5000	7485
Noise Level	High	All	Less As compared to Euclidian	No	Very High
Time Taken(Min.Sec)	0.06	1.90	47.17	161.67	7044.50
DataBase Type	Sequential	Sequential	Sequential	Sequential	Sequential

V PERFORMANCE EVALUATION

DBSCAN [15] constructs clusters using distance transitivity based on a density measure defined by the user. In case of OPTICS [14], using the same approach as DBSCAN [15], the number of clusters formed always remains zero. The reason behind this might be that the data types supported by DBSCAN [15] may not be supported by OPTICS [14].

Also the runtime comparison of DBSCAN [15] and OPTICS [14] on all types of data sets shows that the runtime of algorithm when using the Manhattan distance is always less in both the cases. Also it was found that the number of clusters formed by using Euclidian distance is always better as compared to Manhattan distance.

More clusters are good because they are able to separate noise while generation of clusters. OPTICS [14] has more ability than DBSCAN [15] to handle noise as DBSCAN [15] is generating more clusters.

VI CONCLUSION AND FUTURE WORK

Clustering algorithms are attractive for the task of class identification in spatial databases. This work focus on making an comparative analysis of two density-based clustering algorithms i.e. DBSCAN[15] and OPTICS [14] based on essential parameters needed for good clustering algorithm. Performance evaluation was performed on low to high dimensional data sets based on essential parameters. These parameters include time taken to execute, Distance used, No of Clusters formed, unclustered Instances and noise ratio. Based on the experimental evaluation carried out for low as well as high dimensional data set, it was found that DBSCAN [15] forms more clusters as compared to OPTICS [14] as we don't have more unclustered instances. Furthermore, it was observed that the time taken by Euclidian is always more as compared to Manhattan distance but the number of clusters formed by using Euclidian as a distance measure is more as compared to using Manhattan. Also the noise ratio when using the Euclidian distance is less as compared to Manhattan. So we can conclude that Euclidian distance is always better than Manhattan. As a future work we can carry out the same result using other types of database types as in this paper only sequential database is used.

ACKNOWLEDGMENT

I am thankful to Mr. C. K. Bhensdadia, Professor and Head, D. D. I. T, Nadiad and Mr. Amit Ganatra, Associate Professor and Head, C. S. P. I. T, Changa for their valuable advices and providing an environment for research throughout the work.

REFERENCES

1. Ms K. Santhisree, Dr. A. Damodaram, SSM-DBSCAN and SSM-OPTICS : Incorporating new similarity measure for Density based clustering of Web usage data, in International Journal on Computer Sciences and Engineering, August 2011
2. S. Chakraborty, Prof. N. K. Nagwani, Analysis and Study of Incremental DBSCAN Clustering Algorithm, International Journal of Enterprise Computing And Business Systems, Vol. 1, July 2011
3. M. Parimala, D. Lopez, N. C. Senthilkumar, A Survey on Density Based Clustering Algorithms for Mining Large Spatial Databases, International Journal of Advanced Science and Technology, Vol. 31, June 2011.
4. Dr. Chandra. E, Anuradha. V. P, A Survey on Clustering Algorithms for Data in Spatial Database Management System, International Journal of Computer Applications, Col. 24, June 2011
5. J. H. Peter, A. Antonysamy, An optimized Density based Clustering Algorithm, International Journal of Computer Applications, Vol. 6, September 2010
6. A. Ram, S. Jalal, A. S. Jalal, M. Kumar, A Density based Algorithm for Discovering Density varied clusters in Large Spatial Databases, International Journal of Computer Applications, Vol. 3, June 2010
7. Tao Pei, Ajay Jasra, David J. Hand, A. X. Zhu, C. Zhou, DECODE: a new method for discovering clusters of different densities in spatial data, Data Min Knowl Disc, 2009
8. Zhi-Wei SUN, A Cluster Algorithm Identifying the clustering Structure, International Conference on Computer Science and Software Engineering, 2008
9. Marella Aditya,"DBSCAN And its Improvement",june 2007
10. Stefan Brecheisen, Hans-Peter Kriegel, and Martin Pfeifle, Multi-step Density Based Clustering, Knowledge and Information Systems, Vol. 9, 2006
11. A. Moreira, M. Y. Santos and S. Carneiro, Density-based clustering algorithms-DBSCAN and SNN, July 2005

12. M. Rehman and S. A. Mehdi, Comparision of Density-Based Clustering Algorithms, 2005
13. Levent Ertoz, Michael Steinback, Vipin Kumar, Finding Clusters of Different Sizes, Shapes, and Density in Noisy, High Dimensional Data, Second SIAM International Conference on Data Mining, San Francisco, CA, USA, 2003
14. Yong-Feng Zhou, Qing-Bao Liu, S. Deng, Q. Yang, An Incremental Outlier Factor Based Clustering Algorithm, Proceedings of First International Conference on Machine Learning and Cybernetics, Beijing, 4-5 Nov 2002
15. M. Ankerst, M. M. Breunig, H. P. Kriegel and J. Sander, OPTICS: Ordering Points To Identify Clustering Structure, at International Conference on Management of Data, Philadelphia, ACM 1999
16. Martin Ester, Hans-Peter Kriegel, Jörg Sander and Xiaowei Xu, A Density- Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, The Second International Conference on Knowledge Discovery and Data Mining (KDD-96), Portland, Oregon, USA, 1996
17. X. Wang, H. J. Hamilton. A Comparative Study of Two Density-Based Spatial Clustering Algorithms for Very Large Datasets.
18. K. Mumtaz, Dr. K. Duraiswamy, An Analysis on Density Based Clustering of Multidimensional Spatial Data in Indian Journal of Computer Science and Engineering Vol 1 No 1 8-12
19. D.T.Pham and A.A. Afify, Clustering techniques and their applications in engineering
20. Pavel Berkhin, Survey of Clustering Data Mining Techniques
21. Mariam Rehman, Syed Atif Mehdi, Comparision of Density-Based Clustering Algorithms
22. S. Maji, R. S. Mondal, S. Banerjee, DBSCAN Algorithm with automated parameter selection
23. X. Wang, H. J. Hamilton. A Comparative Study of Two Density-Based Spatial Clustering Algorithms for Very Large Datasets.
24. Alexander Hinneburg, Hans-Henning Gabriel, DENCLUE 2.0: Fast Clustering based on kernel Density Estimation", Martin-Luther-University, Germany
25. Tutorial for WEKA https://blog.itu.dk/SPVC-E2010/files/2010/11/**wekatutorial**.pdf
26. Weka manual for version3.6.3 by Eibe Frank and Mark Hall
27. Data mining Concepts and Techniques by Jiawei Han and Kamber
28. Data Mining: Practical Machine Learning Tools and Techniques, 2nd Edition, Morgan Kaufmann Series in Data Management Systems.

AUTHOR PROFILE

Glory H. Shah is a student of Master of Technology in computer Engineering at Dharmsinh Desai University, Nadiad, Gujarat, India. She is also an Assistant Professor at U & P U. Patel Department of Computer Engineering at Charotar University of Science & Technology, Changa, Dist. Anand, Gujarat, India. She has received her B.E. Computer Engineering degree from Vyavasai Vidhya Pratishtan Engineering College, Rajkot, Gujarat, India in 2010. She has joined M. Tech. at Dharmsinh Desai University, Nadiad, Gujarat, India in 2010. Her current research interest includes Data Mining Clustering (Density-Based Clustering), Distributed Database, Distributed Computing, Compiler Design.

C.K. Bhensdadia is Professor and Head of Department of Computer Engineering at the Dharmsinh Desai University, Nadiad- Gujarat, India. He received a B.E. degree from Dharmsinh Desai University, Nadiad -Gujarat, India in 1990, and M.Tech. degree from IIT Mumbai in 1996. He is currently pursuing Ph.D. His main areas of research are the Genetic Algorithms and Data Mining. He has presented numerous papers on these topics in International Conferences.

Amit P. Ganatra (B.E.-'00-M.E. '04-Ph.D.* '11) has received his B.Tech. and M.Tech. degrees in 2000 and 2004 respectively from Dept. of Computer Engineering, DDIT-Nadiad from Gujarat University and Dharmsinh Desai University, Gujarat and he is pursuing Ph.D. in Information Fusion Techniques in Data Mining from KSV University, Gandhinagar, Gujarat, India and working closely with Dr.Y.P.Kosta (Guide). He is a member of IEEE and CSI. His areas of interest include Database and Data Mining, Artificial Intelligence, System software, soft computing and software engineering. He has 11 years of teaching experience at UG level and concurrently 7 years of teaching and research experience at PG level, having good teaching and research interests. In addition he has been involved in various consultancy projects for various industries. His general research includes Data Warehousing, Data Mining and Business Intelligence, Artificial Intelligence and Soft Computing. In these areas, he is having good research record and published and contributed over 70 papers (Author and Co-author) published in referred journals and presented in various international conferences.