ISSN: 2321-7782 (Online)
Volume 2, Issue 2, February 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Paper / Case Study
Available online at: www.ijarcsms.com

Survey on Different Density Based Algorithms on Spatial Dataset

Noticewala Maitry[1]
Computer Science and Engineering
Parul Institute of Technology
Vadodara – India

Dinesh Vaghela[2]
Asst.Prof.
Computer Science and Engineering
Parul Institute of Technology
Vadodara – India

Abstract: ***DBSCAN is one of the primary clustering algorithms which formed cluster based on the density. This algorithm recognizes arbitrary shape, size and filter out noise. DBSCAN find cluster even surrounded by many other clusters also. This paper gives detail survey on DBSCAN as well as its many other improved methods like FDBSCAN, ODBSCAN,VDBSCAN,ST-DBSCAN and Incremental DBSCAN. Here above algorithms are analyzed based on its essential parameter to determine good clusters.***

Keywords: DBSCAN, FDBSCAN, ODBSCAN, VDBSCAN, ST-DBSCAN, Incremental DBSCAN, Spatial data.

I. INTRODUCTION

Data mining is a process of collecting lots of information from different source then search capabilities and statistical algorithm to discover pattern and correlation in large pre existing database and evaluating it and discover useful knowledge from that data. Data mining is now a day fast growing field and clustering is one of the data mining methods which are performed on unsupervised data. Technique of grouping same character data into one class is called clustering .Different types of clustering methods are Partitioning method, Hierarchical method, Density Based Method, Grid Based Method.

DBSCAN is a one of the density based algorithm to find clusters of arbitrary shapes. DBSCAN is high quality noiseless output cluster algorithm.

Rests of the paper's sections are as followed. Section 2 is on Introduction of DBSCAN algorithm. Section 3 describes different techniques of DBSCAN algorithm. Section 4 presents the comparison table and conclusion.

II. INTRODUCTION OF DBSCAN ALGORITHM

For the DBSCAN algorithm following terms is used in consideration of Database D, Core point (q), Border point (p), Minimum no of points in cluster (Minpts) and Radius (Eps).

Definition 1: Minimum number of points.
Minpts are use to determine whether a neighborhood is denser or not. Minpts specify the density threshold of the denser regions.

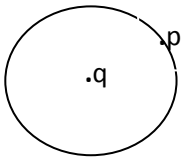
Definition 2: Distance of point p within given Eps
Neighborhood point p within Eps value that is referred NEps (p).
Here **NEps (p)= {q€D | dist(p,q)>=Minpts}**

Definition 3: Core point q condition
Number of NEps (p) is greater than equal to Minpts
i.e.| **NEps (p)|>= Minpts**

Definition 4: Directly density reachable points.
Directly density reachable points are core point q and border point p.

Core point: Minimum numbers of points are needed within Eps-neighborhood.
|NEps (q)| > = Minpts

Border Point: Eps-neighborhood of border point has less point than the Eps of core point.
p€ NEps (q)



q= Core point
p= Border Point

Definition 5: Density reachable points.

Point p is referred as density reachable from another point q in order to Eps and Minpts. If there is a connected chain of point p_1 to p_i, p_1=q, p_i=p such as p_{n+1} is directly density reachable from p_n.

Definition 6: Noise

Any point that is neither core point nor border point and as well as not belongs to any of the cluster is called noise point.

Following are the Algorithm Steps for DBSCAN: [1]

```
DBSCAN (D,Eps,Minpts)
        C=0
        For each unvisited point p in the dataset D
        Mark p as visited
        N=regionQuery(p,Eps)
        If sizeof(N)< Minpts
                Mark p as Noise
        Else
                Put p into new cluster C
        Enlargecluster(Eps,Minps,C,p,N)

Enlargecluster(Eps,Minpts,C,p,N)
        Add p to cluster C
        For each point p' in N
        If p' is unvisited
                Mark p' as visited
                N'=regionQuery(p',Eps)
        If sizeof(N')>= Minpts
                N=N' combine to N
                If p' is not in any cluster than add p' to cluster C
```

Advantages of DBSCAN:

1. DBSCAN can find arbitrary shape cluster.
2. 0DBSCAN can remove noise from the dataset.
3. It is requires only two parameter which are mostly insensitive ordering of the point in the database.

Disadvantages of DBSCAN:

1. Multi density dataset are not complete by DBSCAN.
2. Run time complexity is high.
3. DBSCAN cannot cluster data sets well with large differences in densities.

III. VARIOUS METHODS OF DBSCAN ALGORITHM

FDBSCAN (Fast DBSACN algorithm)

A **FDBSCAN[2]** Algorithm has been invented to improve the speed of the original DBSCAN algorithm and the performance improvement has been achieved through only few selected representative objects belongs inside a core object's neighbor region as seed objects for the further expansion. This algorithm is faster than the basic version of DBSCAN algorithm and suffers with the loss of result accuracy.

The Fast DBSCAN Algorithm's selected seed objects in Region Query has been improved to give the better output, at the same time within less time using Memory effect in DBSCAN algorithm

The remaining objects present in the border area have been examined separately during the cluster expansion which is not done in the Fast DBSCAN Algorithm.

Advantages of FDBSCAN:

1. performance is better than the DBSCAN
2. Runtime complexity and computational time is better than the basic DBSCAN algorithm.

Disadvantages of FDBSCAN:

1. Object loss is higher than the basic DBSCAN.

ODBSCAN (Optimized density based clustering algorithm)

ODBSCAN[3] algorithm, number of Region Query call has been reduced as well as some Region Query calls speed has been improved. For reducing the Region Query function calls, FDBSCAN Algorithm selected representative objects as seed objects approach during the cluster expansion has been used. As the Region Query retrieves the neighbor objects which belong inside the Eps radius, Circle lemmas are given and which can be directly used in the Region Query optimization. Inputs are same as original DBSCAN algorithm and all the data are initialized as UNCLASSIFIED in the beginning. All the border objects have been considered for the clustering process. But there are few possibilities to miss the core objects and which causes some loss of objects. ODBSCAN gives better result than the FDBSCAN Algorithm.

Advantages of ODBSCAN:

1. ODBSCAN algorithm performance is better than the existing algorithm and computation time is also less than the DBSCAN and FDBSCAN.
2. object loss is less than the FDBSCAN as well as border object are applicable in this ODBSCAN algorithm

Disadvantages of ODBSCAN:

1. All the border objects have been considered for the clustering process. But there are few possibilities to miss the core objects and which causes some loss of objects.

VDBSCAN(Varied Density Based Spatial Clustering of Applications with Noise)

VDBSCAN [4] algorithm detects cluster with varied density as well as automatically selects several values of input parameter Eps for different densities. Even the parameter Minpts is automatically generated based on the characteristics of the datasets. VDBSCAN provides a set of different Eps for a user-specified MinPts that can recognize clusters with varying density.

In general the algorithm has two steps, choosing parameters Eps and cluster with varied densities. The procedure for this algorithm is as follows; (i) it calculates and stores k-dist for each project and partition the k-dist plots. (ii) The number of

density is given by k-dist plot. (iii) The parameter Eps is selected automatically for each density. (iv) Scan the dataset and cluster different densities using corresponding Eps (v) Display the valid cluster with respect to varied density.

Advantages of VDBSCAN:

1. VBSCAN provide set of different Eps for user specified parameter that recognize clusters with varying density.
2. VDBSCAN has the same time complexity as DBSCAN and can identify clusters with different density which is not possible in DBSCAN algorithm.

Disadvantage of VDBSCAN:

1. VDBSCAN requiring K as an input from users deteriorates the accuracy of this algorithm.

ST-DBSCAN (Spatial- Temporal Density Based Clustering)

ST-DBSCAN [5] algorithm is constructed by modifying DBSCAN algorithm. In contrast to existing density-based clustering algorithm, ST-DBSCAN algorithm has the ability of discovering clusters with respect to non-spatial, spatial and temporal values of the objects.

The three modifications done in DBSCAN algorithm are as follows:

1. ST-DBSCAN algorithm can cluster spatial-temporal data according to non-spatial, spatial and temporal attributes.
2. In order to solve the conflicts in border objects it compares the average value of a cluster with new coming value.
3. DBSCAN does not detect noise points when it is of varied density but this algorithm overcomes this problem by assigning density factor to each cluster.

Advantages of ST-DBSCAN:

1. Spatial-temporal data refers to data which is stored as temporal slices of the spatial dataset.
2. The knowledge discovery in spatial-temporal data is complex than non-spatial and temporal data.
3. This algorithm can be used in many applications such as geographic information systems, medical imaging and weather forecasting.

Disadvantages of ST-DBSCAN:

1. In this algorithm input parameter are not generate automatically.

Incremental DBSCAN algorithm:

Incremental DBSCAN [6] algorithm is capable of adding points in to bulk to existing set of clusters. In this algorithm data points are added to the first cluster using DBSCAN algorithm and after that new clusters are merged with the existing clusters to come up with the modified set of clusters. In this algorithm Clusters are added incrementally rather than adding points incrementally. In this algorithm R*- tree is use as data structure.

Algorithm Steps for the Incremental DBSCAN:

1. New points added are clustered using DBSCAN.
2. New data points which intersect with old data points are determine.
3. For each intersection point from the new data set use incremental DBSCAN algorithm to determine new cluster membership.
4. Clusters membership of the remaining new points then updated.

Advantages of Incremental DBSCAN:

1. Allow to see the clustering pattern of the new data along with existing cluster pattern.
2. In this algorithm clusters can be merged.

IV. CONCLUSION

This paper gives detail study of the various density based algorithm like DBSCAN, FDBSCAN, ODBSCAN, VDBSCAN, ST-DBSCAN, Incremental DBSCAN based on the different parameters and essential requirement for clustering algorithm in spatial data. Comparison according to different input parameters, arbitrary shapes and varied density is given in table-1.

Table-1: Comparison Table for density based algorithms:

Algorithm:	Input Parameter:	Arbitrary Shape:	Varied Density:
DBSCAN	Radius and Minpts should be given.	YES	NO
FDBSCAN	Radius and Minpts should be given.	YES	NO
ODBSCAN	Number of identical circles, radius and Minpts.	YES	NO
VDBSCAN	Automatically generated.	YES	YES
ST-DBSCAN	Three parameters are given by the users.	YES	NO
Incremental DBSCAN	In initial phase Radius and Minpts should be given. Later phase already generated clusters.	YES	YES

References

1. Khushali Mistry, Swapnil Andhariya, Prof. Sahista Machchhar" NDCMD: A Novel Approach Towards Density Based Clustering UsingMultidimensional Spatial Data". International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 www.ijert.org IJERTIJERT Vol. 2 Issue 6, June – 2013
2. SHOU Shui-geng, ZHOU Ao-ying JIN Wen, FAN Ye andQIAN Wei-ning.(2000) "A Fast DBSCAN Algorithm" Journal of Software: 735-744.
3. J. Hencil Peter, A. Antonysamy" An Optimised Density Based Clustering Algorithm" International Journal of Computer Applications (0975 – 8887) Volume 6– No.9, September 2010
4. Wei Wang, Shuang Zhou, Bingfei Ren, Suoju He"IMPROVED VDBSCAN WITH GLOBAL OPTIMUM K" ISBN: 978-0-9891305-0-9 ©2013 SDIWC
5. Derya Birant, and Alp Kut. ST-DBSCAN: An algorithm for clustering spatial-temporal data Data Knowl. Eng. (January 2007)
6. Navneet Goyal, Poonam Goyal, K Venkatramaiah, Deepak P C, and Sanoop P S" An Efficient Density Based Incremental Clustering Algorithm in Data Warehousing Environment" 2009 International Conference on Computer Engineering and Applications IPCSIT vol.2 (2011) © (2011) IACSIT Press, Singapore.

