Density based spatial clustering of applications with noise (DBSCAN)

Digital Image Processing

University Institute of Technology

The University of Burdwan

Group Members:

Biswajit Roy (2016-1031)

Ankita Pandey (2016-1018)

Subhajit Mondal (2016-1008)

Rajib Minj (L2017-1072)

Project Guide

Sir. Anamitra Basu

Introduction

DBSCAN is a data clustering algorithm proposed by **Martin Ester**, **Hans-Peter Kriegel**, **Jörg Sander** and **Xiaowei Xu** in 1996. It is a density-based clustering algorithm which finds a number of clusters starting from the estimated density distribution of corresponding nodes. The definition of DBSCAN is a cluster based on the notion of **density reachability**.

According to the Webster dictionary:

- A number of similar things growing together or of things or persons collected or grouped closely together: Bunch

A cluster is closely packed group of people or things.

Objective : Choosing DBSCAN over other clustering

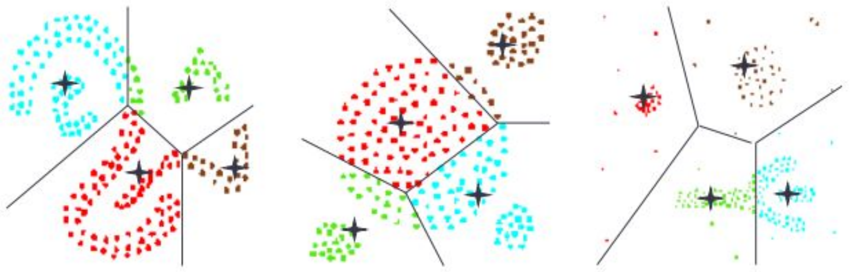
Partitioning methods (e.g. K-means) and hierarchical clustering work for finding spherical-shaped clusters or convex clusters.

In other words, they are suitable only for compact and well separated clusters.

Moreover, they are also severely affected by the presence of noise and outliers in the data.

Real life data may contain irregularities, like -

- Clusters can be of arbitrary shape.
- Data may contain noise.
- KMeans result for K=4 in fig. doesn't look satisfying.



DBSCAN over other clustering algorithms (Advantages)

- DBSCAN doesn't require one to specify the number of clusters in the data.
- It can find arbitrarily shaped clusters; even find a cluster surrounded by a different cluster.
- It can find noise points and is robust to outliers.
- It requires just two parameters.

Conventional DBSCAN Algorithm

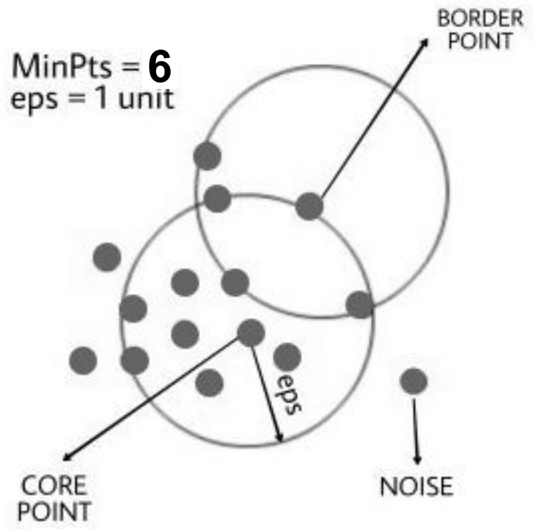
eps : It defines the neighbourhood around a data point

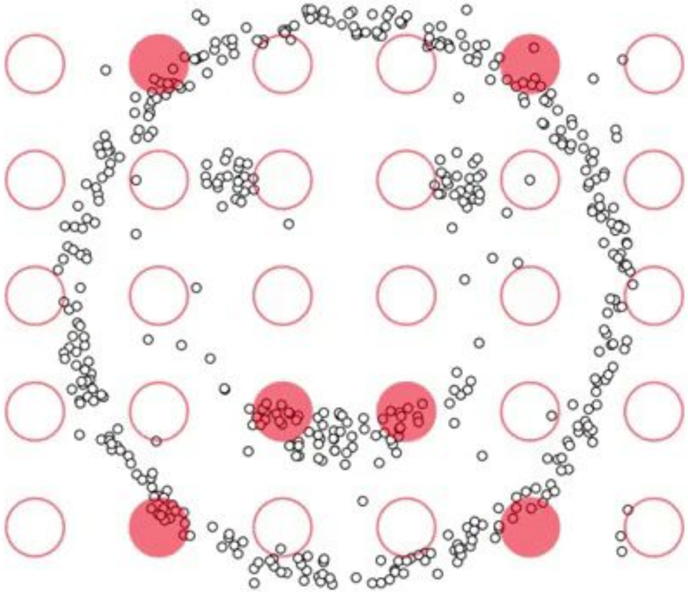
MinPts : Minimum number of neighbours (data points) within ‘eps’ radius.

Find all the neighbour points within eps and identify the core points or visited with more than MinPts neighbours.

- For each **core point** if not assigned to a cluster, create new **cluster**.
- Recursively find all its density connected points and assign them to the same cluster as the core point.
- Iterate through the remaining unvisited points in the dataset. Those points that do not belong to any cluster are **noise**.

Conventional DBSCAN Algorithm (Graphical)



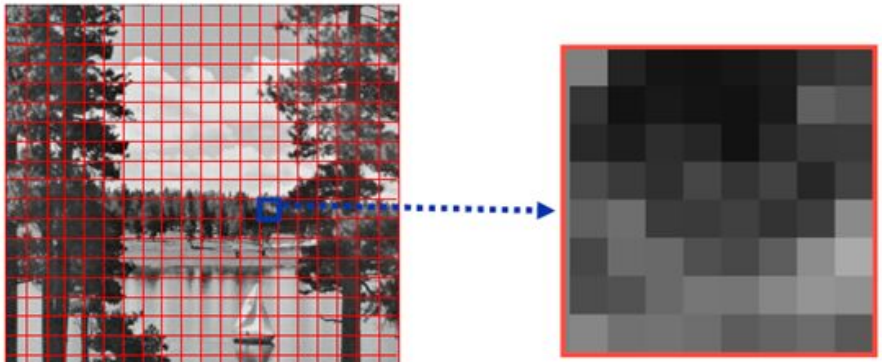


DBSCAN Algorithm (Disadvantages)

- The process cannot be partitioned for multiprocessor systems.
- Datasets with altering densities are tricky.
- Sensitive to clustering parameters, MinPts and Eps.
- Fails to identify clusters if density varies and if the dataset is too sparse.
- Sampling affects density measure.

Graphical discussion on proposed algorithm

- An image is captured and gray value of image is computed into a matrix.
- 8-neighbour distance of a pixel P with respect to pixel Q is found out using the algorithm.
- Density of a pixel is computed.



Input image

8x8 block

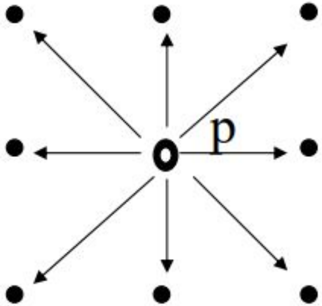
$$
\begin{bmatrix}
52 & 55 & 61 & 66 & 70 & 61 & 64 & 73 \\
63 & 59 & 55 & 90 & 109 & 85 & 69 & 72 \\
62 & 59 & 68 & 113 & 144 & 104 & 66 & 73 \\
63 & 58 & 71 & 122 & 154 & 106 & 70 & 69 \\
67 & 61 & 68 & 104 & 126 & 88 & 68 & 70 \\
79 & 65 & 60 & 70 & 77 & 68 & 58 & 75 \\
85 & 71 & 64 & 59 & 55 & 61 & 65 & 83 \\
87 & 79 & 69 & 68 & 65 & 76 & 78 & 94
\end{bmatrix}
$$

Graphical discussion on proposed algorithm

$D_i(P,Q) = (\,|(u-v)|\,)$; *where u & v is gray value of pixel P & Q resp.*

$N_8(P) = max\,(D_0, D_1, D_2, D_3, D_4, D_5, D_6, D_7)$

$$Density\ of\ pixel(P) = \frac{number\ of\ pixel\ with\ gray\ value\ > 0}{N_8(P)\ distance}$$

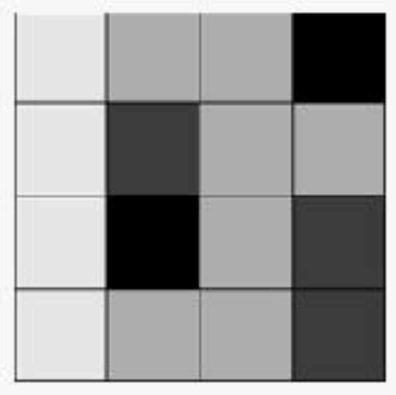


$N_8(p)$

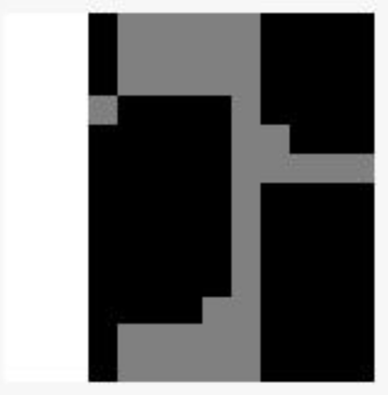
Project results (conventional DBSCAN)

Using conventional DBSCAN over grid image for

(a) Eps=5 and MinPts=10

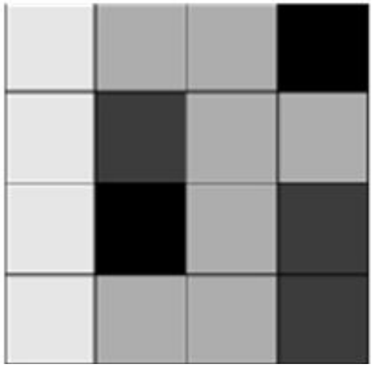


Original grid.jpg



Clustered image

(b) Eps=5 and MinPts=20



Original grid.jpg

Clustered image

Project results (Novel DBSCAN Algorithm)

Input image

Clustered image

The novel algorithm does not depend upon any parameters as conventional DBSCAN and KMeans. Hence, output for an image is consistent

Conclusion

In this presentation, a novel method of performing DBSCAN is proposed.

- The proposed algorithm will help in performing effective DBSCAN clusters even without **Eps** and **MinPts**
- The parameters are replaced by density of regions in the image which is computed automatically according to image.
- Algorithm can form clusters even when the density is varying in image.

Hence, the image can be clustered effectively using the proposed algorithm.

References

[1]https://towardsdatascience.com/how-dbscan-works-and-why-shouldi-use-it-443b4a191c80

[2]https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf

[3]https://www.geeksforgeeks.org/dbscan-clustering-in-ml-densitybased-clustering/

[4]https://webdocs.cs.ualberta.ca/~zaiane/courses/cmput690/slides/Chapter8/index.html

[5]http://www.cse.iitb.ac.in/infolab/Data/Courses/CS632/1999/clusterin g/dbms.html

[6]https://apple.github.io/turicreate/docs/userguide/clustering/dbscan.html

Thank You