Image Clustering Technique

Submitted by

Ahmed Elnagar[1] ,Heba Mohammed[2], Fatma Abdelsalam[3], Magda Ahmed[4], Yasmine Mohamed[5], Mohrail Thrwat[6].

[1]Assistant Professor at Faculty of Computers & Artificial Intelligence Beni Suef University, Email: elngar_7@yahoo.co.uk
[2]Teaching Assistant at Faculty of Computers & Artificial Intelligence Beni Suef University, Email: heba mohamed@fcis.bsu.edu.eg
[3]Computer Science Student at Beni Suef University, E-mail: **fatmaabdelsalam328_sd@fcis.bsu.edu.eg**
[4]Computer Science Student at Beni Suef University, E-mail :magdaahmed330_**sd@fcis.bsu.edu.eg**
[5]Computer Science Student at Beni Suef University, E-mail: yasminemohamed391_sd@fcis.bsu.edu.eg
[6]Computer Science Student at Beni Suef University, E-mail: mehrailtharwat374_**sd@fcis.bsu.edu.eg**

Abstract: In order to the find the close association between the density of data points, in the given data set of pixels of an image, clustering provides an easy analysis and proper validation. In this paper various clustering techniques along with some clustering algorithms are described. Further k-means algorithm, its limitations and a new approach of clustering called as M-step clustering that may overcomes these limitations of k-means is included. .

Keywords: Clustering , k-means clustering , Clustering Techniques for Digital Image Segmentation,

1. Introduction

Computer vision tries to understand scene with the help of image processing and machine learning. Image processing is a process an image with the help of a processor**[2]**.

Image processing manipulates an image to analyze, to better understand, and to achieve required results. The centre of attention of digital image processing is to make a digital system with the help of efficient algorithm and techniques which is capable of processing an image**[2]** .

In 1854, in London during cholera outbreak, John Snow had plotted the diseased reported cases using a special map. After the creation of the map it was observed that there was close association between the density of disease cases and a single well located at a central street. With the above map it is very simple

to identify association between phenomena, which is very difficult to analyze in actual. This was the first known application of clustering analysis for many researchers **[3]** .
Since then cluster analysis has been widely used in many fields such as statics, marketing, engineering, medical and other social sciences, for clustering large data sets into natural groups, the number of clustering algorithms had been proposed for performing clustering task.

1.1. Applications of Clustering

market research:can help marketers discover distinct groups in their customer base**,** characterize their customer groups based on the purchasing patterns, **patternrecognition, data analysis, and image processing,biology**: it can be used to derive plant and animal taxonomies, categorize genes with similar functionalities and gain insight into structures inherent to populations[6].

2. Objectives of the Research

functions in clustering formalize the goal of attaining high intra-cluster similarity (documents within a cluster are similar) and low inter-cluster similarity (documents from different clusters are dissimilar). This is an internal criterion for the quality of a clustering[7].

3. Methodology

This clustering method helps grouping valuable data into clusters and picks appropriate results based on different techniques. For example, in information retrieval, the results of the query are grouped into small clusters, and each cluster has irrelevant results. By Clustering techniques, they are grouped into similar categories, and each category is subdivided into sub-categories to assist in the exploration of queries output,

3.1 Partitioning Clustering

The most popular example of this algorithm is the KNN algorithm. This is how a partitioning clustering algorithm looks like[8]

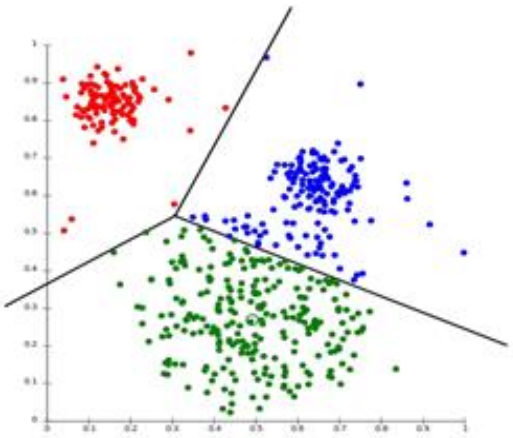
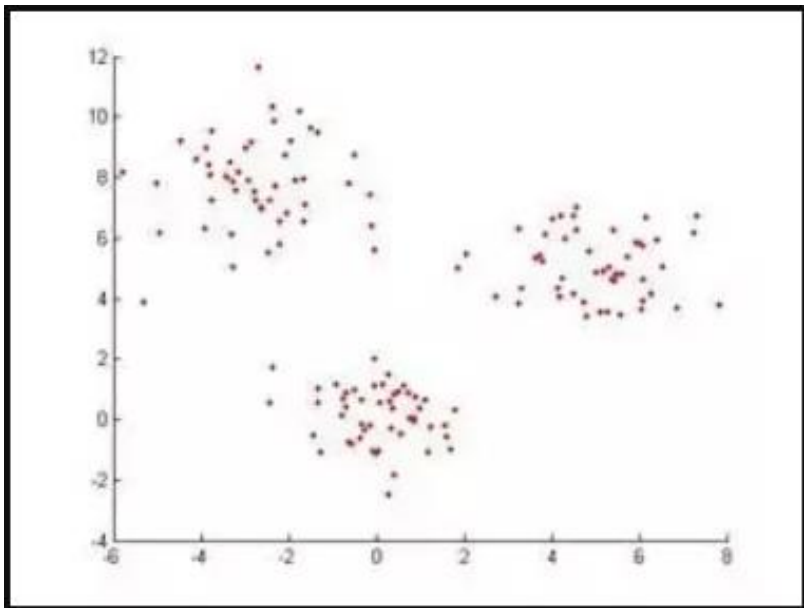


Fig. 3.0

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
file = "/path/to/the/dataset"
#Push dataset into Pandas dataframe
dataset = pd.read_csv(file)
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.30)
from sklearn.neighbors import KNeighborsClassifier
classifier = KNeighborsClassifier(n_neighbors=3)
classifier.fit(X_train, y_train)
y_pred = classifier.predict(X_test)
```

The results of A coding example in Listing 3.1.1



3.2 Hierarchical Clustering

It is a clustering technique that divides that data set into several clusters, where the user doesn't specify the number of clusters to be generated before training the model[8]

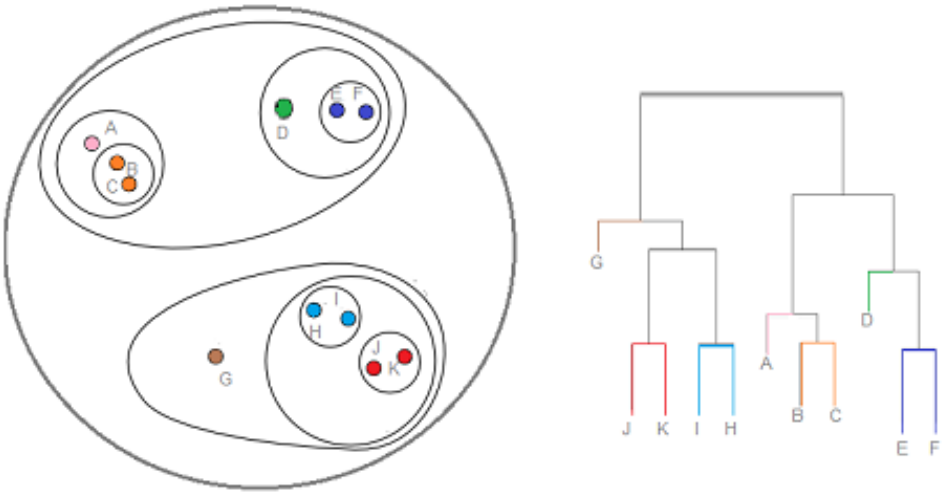


Fig. 3.2

3.3 Density-Based Clustering

Density-Based Spatial Clustering and Application with Noise (DBSCAN) is the most used algorithm in this type of technique

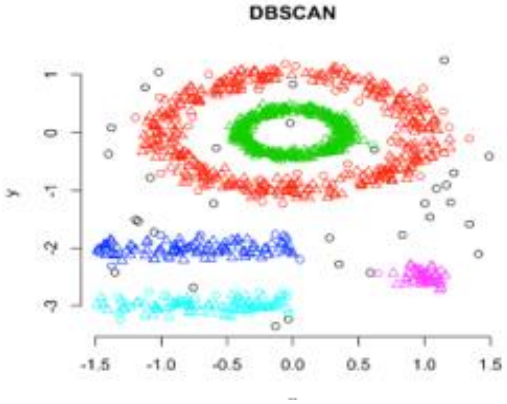


Fig. 3.3

3.4 Distribution Model-Based Clustering

In this type of clustering, technique clusters are formed by identifying the probability of all the data points in the cluster from the same distribution (Normal, Gaussian).

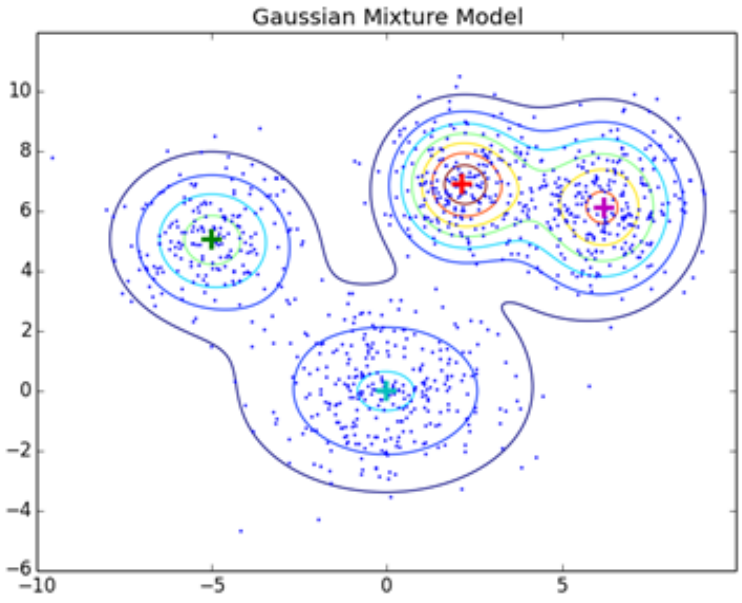


Fig. 3.4

6. K-mean algorithm

is an unsupervised clustering algorithm. Based on the inherent distances between data points, k-mean classified the input data points into multiple classes or clusters [9].

7.KNN

is firstly introduced by E. Fix and J. Hodges researchers in their paper Discriminatory Analysis: Nonparametric Discrimination: Consistency Properties, in 1951 [10].

$$d = \sqrt{\sum_{i=0}^{i=n} (X_i - Y_i)^2}$$

Advantages of Clustering	**Disadvantages of Clustering**
• The main advantage linked to clustering computers is that if one computer fails, any other computer within the cluster can assume the load of that failed computer. Clustering therefore fosters better availability and users of such a system normally see limited or no disruption of access. • Clustering computers ensures that your business applications do not rely on one computer only. • 3. Improves flexibility:Clustering offers a cohesive computing resource that provides IT personnel more options with regards to the configuration of the infrastructure.	• Cost is high. Since the cluster needs good hardware and a design, it will be costly comparing to a non-clustered server management design. Being not cost effective is a main disadvantage of this particular design. • Since clustering needs more servers and hardware to establish one, monitoring and maintenance is hard. Thus increase the infrastructure.[11]

• 4. Enhances network performance: Through clustering servers, users benefit from an improved network performance.	

8. Achievements

An achievement cluster is a term used in entrepreneurship to describe the characteristics needed for entrepreneurs.

1. Opportunity-seeking
2. Persistence
3. Commitment to work contract
4. Risk-taking
5. Demand for quality and efficiency[13].

9. Conclusion

In this paper various techniques and parameter are discussed for clustering methods. The shortcoming of k-means clustering algorithm to find optimal k value and initial centroid for each cluster is discussed.

References

[1]International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-1, March 2012.

[2] International Journal of Scientific & Engineering Research, Volume 7, Issue 12, December-2016 55 ISSN 2229-5518

[3] Wenbing Tao, Hai Jin and Yimin Zhang," Color Image Segmentation Based on Mean Shift and Normalized Cuts", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 37, NO. 5, OCTOBER 2007, pp. 1382-1389.

[4] D. I. P. Tutorialspoint, "Tutorialspoint,"

2016. [Online]. [Accessed 4 April 2016].

[5] L. Hao, "National University Of Singapor," July 2006.

[6] http://www.tutorialspoint.com/data_mining/dm_cluster_analysis.htm

[7]https://nlp.stanford.edu/IR-book/html/htmledition/evaluation-of-clustering

[8] Types of Clustering,By Priya PedamkarPriya Pedamkar :https://www.educba.com/types-of-clustering/

[9] S. Tatiraju And A. Mehta, "UCI Donald Bren, School Of Information And Computer Science," 2016. [Online]. Available: Http://Www.Ics.Uci.Edu/~Dramanan/Teaching/Ics273a_Winter08/Projects/Avim_Report.Pdf. [Accessed April 2016].

[10] E. K. Mower, "Computing Research Association," 2016. [Online]. Available: "http://archive2.cra.org/Activities/Craw_Archive/Dmp/Awards/2003/Mower/KNN.Html" [Accessed April 2016]

[11] https://www.tecmint.com/ ch-cman Documentation

[12] https://www.tecmint.com/ Mod Cluster Documentation

[13] https://entrepreneurspecialist.blogspot.com/ https://entrepreneurspecialist.blogspot.com/2010/12/achievement-cluster.html" HYPERLINK "https://entrepreneurspecialist.blogspot.com/2010/12/achievement-cluster.html"/achievement-cluster.html