ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 4, March 2017

A REVIEW ON PIXEL BASED CLUSTERING ALGORITHMS IN IMAGE PROCESSING SYSTEMS

FRIZILIN.R
Research Scholar, Bharathiar University, Coimbatore

DR.A.MUTHUKUMARAVEL
Professor, Head, Dept.of MCA, Bharath Institute of Higher Education and Research, Selaiyur, Chennai

***Abstract*— Cluster is defined as a collection of articles which has the effects of similarity between them and the effects of dissimilarity between other clusters.**

Clustering Algorithms are more powerful in designing unlabeled patterns more efficiently. This paper deals with a survey of all related clustering approaches which describes the functionality, required parameters, time and space complexity needed for clustering of images.

It can be termed as a well chosen algorithm for unsupervised learning problem. It plays a major role in all research issues related to variety of applications data setsnamely image processing, data mining, computational biology, mobile communication and medicine.

Traditional clustering algorithms mainlyconsider clusters with spherical shapes and similar sizes, but later stage clustering algorithms deals with dissimilar sizes and shapes.

Keywords*— *Clustering, unsupervised learning, pixel based clustering,K-means, Fuzzy-C means clustering.

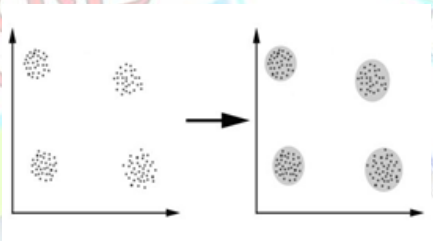
I. INTRODUCTION (*HEADING 1*)

Clustering in image processing is defined as the technique of identifying similar image primordial groups[1]. It mainly groups the pixels of an image together. Clustering is a pure classification technique[3]. It is classified into supervised and unsupervised clustering demands. come into existence by arranging a clustering technique with accordant, for a given clustering problem. It separate the large amount of data into clusters[10]. Moreover, it is well known that there is no adequate clustering method to handle all sorts of cluster architecture(shape, size and density).

II. CLUSTERING

Clustering is a process of standardizing the articles into groups based on its attributes. A cluster maintains similarity within them and dissimilarity between other clusters.

Clustering mainly described here as a classification of similar images in the database. It is done based on different contents of an image like size, color and texture.[2]. An image is arranged by a keyword or by its content. In keyword based clustering, a keyword is used which describes about the image keyword. The identical featured images are assembled to form a Cluster and assigning value to each feature. In content based clustering[6],[7],[8], a content refers to shapes or textures or any other knowledge based information that can be congenital from the imageitself. Clustering basically investigates the search for definite groups in the featurespace. These clustering groups have different framework and they can be clearly discriminated. Inclustering,project separates the data into number of partitions,



(Fig I:showing four clusters framed from unlabeled set of data)

Clustering algorithms can be broadly grouped into two categories namely

A) Supervised Clustering Algorithm and

B) Unsupervised Clustering Algorithms Equations

A. Supervised Clustering

For every input, we train our task with the corresponding target. It is called supervised learning, which will provide adequate target for any new input after sufficient training. Our algorithm follows a function from inputs to the respective expected targets. If the targets are expressed in some classes, it is called classification problem.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 4, March 2017

Alternatively, if the target space is continuous, it is called regression problem.

B. Supervised Clustering

If our task is trained with a set of inputs, then it is called unsupervised learning, which will be more helpful to find the architectural relationships between different inputs. Most important unsupervised learning is clustering, which in turn creates different clusters as inputs and will be able to put any new input in appropriate cluster.

Unsupervised clustering is further classified as Unsupervised linear clustering algorithms and Unsupervised non-linear clustering algorithms.

B1. Unsupervised linear Clustering

It has limited modeling capacity. only certain dynamics can be modelled. If the classes of the training vectors are known, supervised methods are used to build the classifiers. If class information does not exist, one has to hideout to unsupervised methods.

III. CLUSTERING TECHNIQUES

Clustering aims on natural partitioning of data.There are different approaches for image segmentation like threshold based, edge based, cluster based, pixel based and neural network based[5]. From the different approaches one of the most adequate methods is the clustering method.

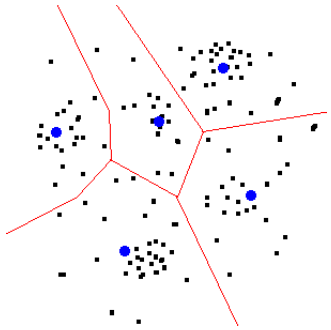
Based on that there are different types of clustering in which we focus on pixed based clustering techniques namely K-means clustering and Fuzzy C-means clustering.One of most used clustering algorithm is k-means clustering In this paper, we just focus on unsupervised learning algorithms namely K-Means and Fuzzy C-Means clustering algorithms.

A. K-Means Clustering Algorithm

K-Means clustering groupsarticles into K groups where K belongs to a variety of pre-chosen groups. The grouping is performed by reducing the sum of Euclidean distances

(ie.squared distances) between articles and the corresponding centrical part(centroid).

A centroid denotes the t center of a geometric article of uniform density. In this paper, we consider mean vectors as centroids.



(Fig II: k-means cluster groups with representation of data points)

K-Means is one of the easiest and simplest unsupervised learning algorithms which solves the prominent clustering problems. The procedure classifies a given data set with a certain number of clusters especially K clusters.

The main aim is to characterize k centers with one center for each cluster. These centers should be placed in a tricky way because based on the location of the centers, the results differ. So, placing the centers of each clusters far away from each other will yield better results.

From a given set of data, each point is chosen and associated to the nearest center. At a particular level, when there is no specific point is pending, then we conclude that the first step is done successfully. At this level, we need to recollect k new centroids of the clusters that results from the previous step.

With this k new centroids, we need to bind the data set points with the nearest new center. This generates a loop. As a result of this loop we may consider that the k centers change their location step by step until no more changes are done or in other words centers do not move any more. Finally, this algorithm reduced to an objective function know as squared error function given by:

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} \left(\left\| x_i - v_j \right\| \right)^2$$

where

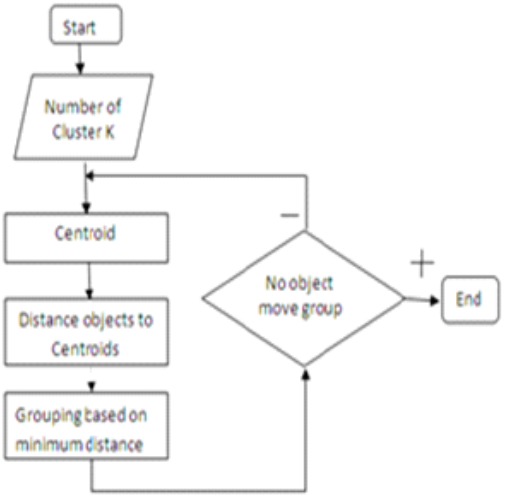
||xi - vj|| is the Euclidean distance between xi and vj.,ci represents the number of data points in ith cluster.

c denotes the number of cluster centers.

1. Flowchart

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 4, March 2017



2. *Algorithm*

Let D = {d1,d2,d3,……..,dn} be the set of data points and C = {c1,c2,…….,cnc} be the set of centers.

1) Randomly choose 'k' data items as initial centroids.

2) Calculate the distance between each data point and cluster centroids.

3) Assign each data point to the cluster center so that the distance from the cluster center is minimal of all the cluster centers..

4) Recollect the new cluster center with suitable formulas.

5) The distance among each data point and newly obtained cluster centers are recalculated.

6) If no data point is reassigned, stop the process. Otherwise repeat from step 3.

3. *Strength*

- *The major advantage of k-means clustering is its fastness and easy understandable.*
- *Gives the best outcome when data sets are well separated from each other.*

4. *Weakness*

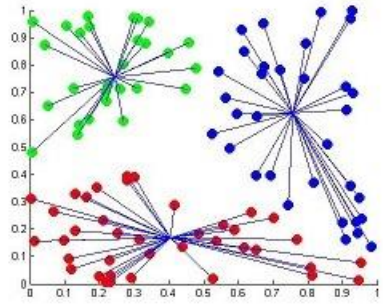
- *A prior information regarding the number of cluster center should be known in advance.*
- *No proper resolution obtained when there is overlapping of data.*
- *Algorithm produces different results for different way of representing the same data set.*
- *Selection of cluster centroid randomly, does not yield good outcomes.*
- *Noisy data cannot be handled properly*

C. Fuzzy C-Means Clustering Algorithm

Fuzzy C-means (FCM) is a method which defines a data point belonging to clusters rather than belonging to one whole cluster. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2$$

where m is any real number greater than 1.
uij is the degree of membership of xi in the cluster j
xi is the ith of d-dimensional measured data, cj is the d-dimension center of the cluster.

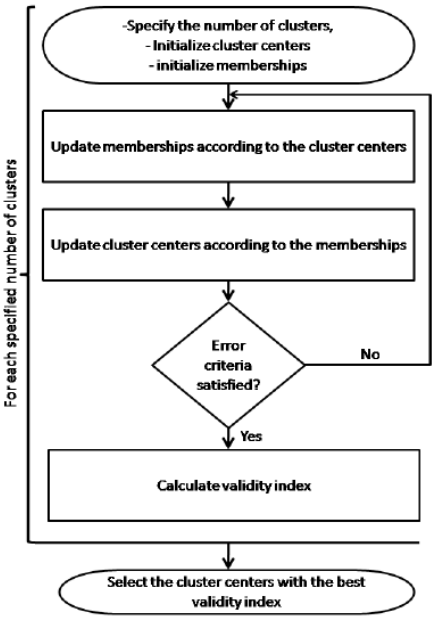


(Fig III: Fuzzy C-Means cluster groups with representation of data points)

a) Flowchart

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 4, March 2017



b) Algorithm

*1)*The algorithm is composed of the following steps:
2)Choose randomly of atleast 2 centroids.
3)Calculate the dissimilarity between the data point and the cluster center.
4)Repeat steps till the centroids are not changing.

c) Strength

Clusters all kinds of noisy data.

d) Weakness

Coincidence of clusters may occur due to independent rows and columns of data.

IV. RELATED WORK

1. Santanu Bhowmik, Viki Datta published a journal paper on "A Survey on Clustering Based Image Segmentation" were a comprehensive survey was done on different clustering techniques that helps to increase the efficiency of the image retrieval process.
2. Vinodhini published an international paper on "Survey on Clustering Algorithms" were the efficiency is higher when it is implemented with the tradeoff weighted fuzzy factor and the kernel metric.
3. Swati Harkanth, Prof. B. D. Phulpagar published an international journal paper on "A Survey on Clustering Methods and Algorithms" were the time complexity of different kinds of clustering algorithms are compared.
4. Ka-Chun Wong presented a paper on "A Short Survey on Data Clustering Algorithms" were to verify the algorithm's competitiveness, different types of performance metrices were defined and reviewed.
5. Devarshi Naik , Pinal Shah published an international journal paper on "A Review on Image Segmentation Clustering Algorithms" were the maximum memory usage in each clustering algorithms were discussed and compared and concluded that K-means and self organizing maps are fastest among other algorithms.
6. Priyansh Sharma and Jenkin Suji published an international paper on "A Review on Image Segmentation with its Clustering Techniques" were reviews related to image segmentation and problem identification were discussed and covered.

IV. CONCLUSION

Clustering deals with high dimensionality of data. A comparative study is required between different clustering techniques and a better knowledge about its strength and weakness. No particular technique is fixed for particular distribution of data set. Each clustering approaches differs based on the types of data the project is dealing with.

A complete comparative analysis survey has been made for image segmentation with pixel based clustering techniques. For a particular set of data point, a particular method only can be suitable and one clustering approach cannot suit for all types of data[4]. Through clustering algorithms, image segmentation can be done in an effective way[9].

References

[1] Puzicha, J., Hofmann, T. and Buhmann, J. M.,"HistogramClustering for Unsupervised Image Segmentation"., ComputerVision and Pattern Recognition , Vol.2.IEEEpress,602-608,2000.

[2] A Survey on Clustering Based ImageSegmentation, SantanuBhowmik, VikiDatta ISSN: 2278 – 1323;International Journal of Advanced Research in Computer Engineering &Technology,Volume 1, Issue 5, July 2012

[3] A Review on Image Segmentation ClusteringAlgorithmsDevarshiNaik , Pinal Shah Department of Information Technology, Charusat UniversityCSPIT, Changa, di.Anand, GJ,IndiaDevarshiNaik et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 3289 – 3293

[4] Swati Harkanth et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 4 (5) , 2013, 687-691; A Survey on Clustering Methods and Algorithms; Swati Harkanth, Prof. B. D. PhulpagarComputer Science Department, PES Modern College of EngineeringPune University, Pune, India

[5]KhaledHammouda, A Comparative study of Data Clustering technique. Department of System Design Engineering, University of Waterloo,Canada.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 4, March 2017

[6] Yixin Chen, James Z.Wang, Robert Krovetz, "ContentBased Image Retrieval byClustering",Proc of the 5th ACMSIGMM I'ntl workshop on Multimedia information retrieval,NewYork,ACM press,pp-193-200,2003.

[7] Huiyu Zhou, Abdul H. Sadka, Mohammad R. Swash, JawidAzizi and Abubakar S. Umar., "Content Based Image Retrievaland Clustering: A Brief Survey" school of Engineering and Design, Brunel University, Uxbridge, UB8 3PH, UK.

[8] Zhang, Q.,Goldman,S. A., Yu,W.,andFritts, J.E.."Content based image retrieval using multiple instance learning".InProc.ICML,2002.

[9] Priyansh Sharma1 and Jenkin Suji2 Dept. of EC, ITM University "A Review on Image Segmentation with its Clustering Techniques" International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.9, No.5 (2016), pp.209-218 http://dx.doi.org/10.14257/ijsip.2016.9.5.18

[10] Ka-Chun Wong , City University of Hong Kong "A Short Survey on Data Clustering Algorithms".