International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 15
May 2018

K-means Based Convex Hull Triangulation Clustering Algorithm

Y. Ratna Rao & Dr. D.B.K. Kamesh
[1]Schlor, Shri Venkateshwara University, [2]Professor, St. Martin's Engineering College, Secunderabad

Abstract:

As the variety of accessible web content grows, it becomes harder for users finding documents relevant to their interests. cluster is that the classification of an information set into subsets (clusters), so the info in every set (ideally) share some common attribute - usually proximity per some outlined distance live. we introduce Kmeans-Based Convex Hull Triangulation clustering algorithm (KBCHT) a new clustering algorithm that studies the given dataset to find the clusters. KBCHT algorithm can detect clusters without pre-determination of clusters number in datasets which contain complex non-convex shapes, different sizes, densities, noise and outliers.

Keywords: Cluster, K-means, KBCHT algorithm

Introduction

KBCHT has three phases of operations. The first phase obtains initial groups from running Kmeans algorithm just once, the second phase analyzes these initial groups to get sub-clusters and the last one merges the sub-clusters to find the final clusters in the dataset. As shown in Algorithm, KBCHT performs K-means algorithm on the dataset x given the number of clusters k.

Proposed System

The use of k is just to run K-means as we will notice by further study of the effect of k in means that the first run of K-means algorithm despite its bad initialization conditions has an initial set of clusters iCi where i is from 1 to N the number of obtained clusters from K-means. The set i_c with index i contains data points from dataset x which belong to the initial cluster i. Lines 3 to 7 describe the process of how we analyze these initial clusters to obtain a set of sub-clusters. In line 4, we construct a set of vertices which represents each initial clusters iC. This set of vertices is obtained from the convex hull of each initial clusters iC. The set iV handles these vertices which contains two indexes i and j as shown in line 4. In which the index i indicates that these vertices belong to the initial cluster i and the index j represents the vertex number of convex hull of initial cluster i in a counterclockwise order. After obtaining the vertices from the convex hull, these vertices need to be shrunk by adding new vertices from the belonged initial clusters set iC. Thus, we begin with vertices drawn from a convex hull and finish with vertices of a polygon. The shrunk vertices are handled in the set sV as shown in line 5 of Algorithm 3.1. Line 6 takes the shrunk vertices sV and processes them to obtain a set of sub-clusters s, the number of these sub-clusters S and the average distance between data points of each of the sub-clusters in sC (sCaD) using the Delaunay triangulation [1] as will be explained later. The sub-clusters are formed by searching for closed loops vertices in the sV set. The set sC has indexed from 1 to S in which sCi contains data points of dataset x that belong to sub-cluster i. Some of these sub-clusters could be merged together to form the result of clusters C as shown in line 8[2][3][4].

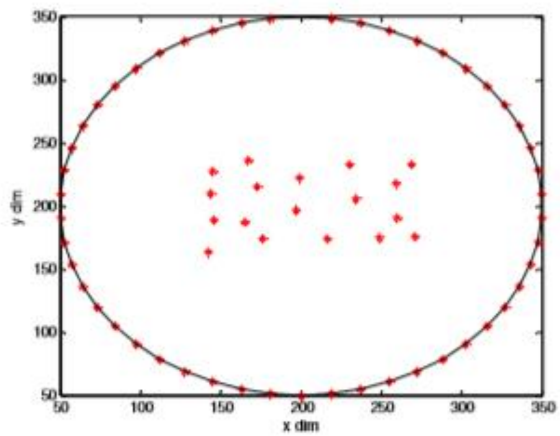
Algorithm:

1. Begin initialize k,x,iC={}, sC={}, iV={}, sV={}, sCaD={}, N=0,S=0
2. iC, N ← Kmeans(x,k)
3. for i=1 to N
4. iV(i,j) ← construct convexhull for cluster iC_i
5. sV(i,k) ← shrinkVexhull ($iC_{i\ ,\ iV(I,:)}$)

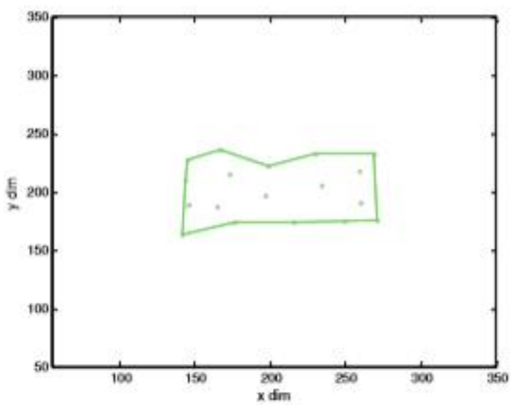
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 15
May 2018

6. sC, S,sCaD ← findSubCluster(iCi,sV(I,:))
7. End_for
8. c ← merging(sC,S, sCaD)
9. Return C
10. end

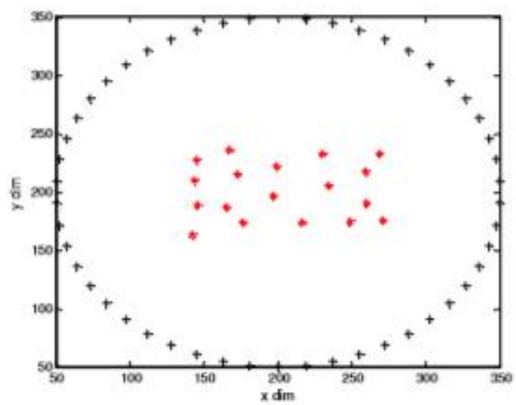
Results:



(a)

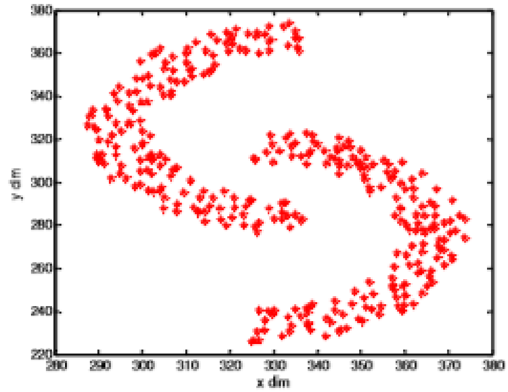


(b)

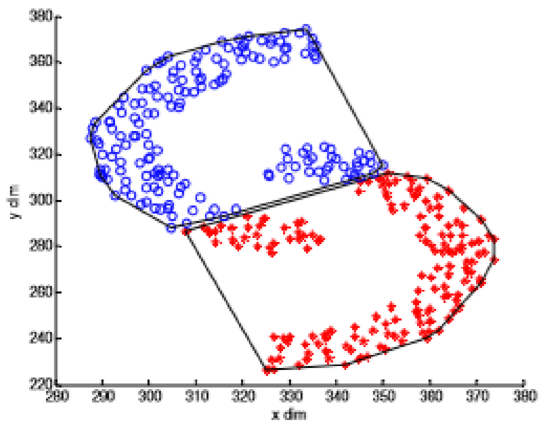


(C)

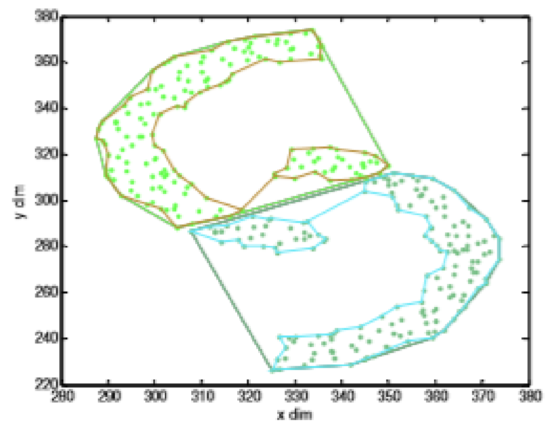
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 15
May 2018



(a)



(b)



(C)

Conclusion and Future Scope:

To prove the effectiveness and the strength of our proposed KBCHT algorithm, we construct experiments for measuring the performance in case of time cost and clustering accuracy according to the visual results

References:

[1]. Shakes, M. Langheinrich and 0. Etzioni. Ahoy! the home page finder. In Proceedings of the 6th World Wide Web Conference, 1997.
[2]. D.Hochbaum, and D. Shmoys, "A Best Possible Heuristic for the K-Center Problem," Math. Oper. Res., Vol. 10, No. 2, pp. 180-184, 1985.
[3]. Gan, Ch. Ma, and J. Wu, "Data Clustering: Theory, Algorithms, and Applications," ASA-SIAM series on Statistics and Applied Probability, SIAM, 2007.
[4]. Hansen and B. Jaumard, "Cluster analysis and mathematical programming," Mathematical Programming, pp. 191-215, 1997.