Holistic Unique Clustering

Author:
Ramesh Chandra Bagadi
Data Scientist
INSOFE (International School Of Engineering), Hyderabad, India.
rameshcbagadi@uwalumni.com
+91 9440032711

Technical Note

Abstract

In this research Technical Note the author has presented a novel method to find all Possible Clusters given a set of M points in N Space.

Theory

Definition of a Cluster

We define a Cluster as follows:

A Cluster is a collection of Points (or objects) wherein they are scattered (their property is distributed) in such a fashion that, for a specified distance (measured in appropriate Metric of concern using appropriate Norm of concern) every point of this cluster has at least one neighbouring point also belonging to this cluster located within

(i) this specified distance* [1]

(ii) a certain small neighbourhood of this this specified distance, measured from the aforementioned point of concern.

Proximity Matrix

Given M number of points $\bar{x}_i \in R^N$, $i = 1 \; to \; M$, each belonging to R^N, we find the Proximity Matrix P for each (M number of) point with each of all other (M Number of points) points, inclusive of itself. The Proximity can be found using Euclidean distance or using the concept stated in [1].

$$P = \begin{bmatrix} d(1,1) & d(1,2) & d(1,3) & . \; . \; . \; . & d(1,(m-1)) & d(1,m) \\ d(2,1) & d(2,2) & d(2,3) & . \; . \; . \; . & d(2,(m-1)) & d(2,m) \\ d(3,1) & d(3,2) & d(3,3) & . \; . \; . \; . & d(3,(m-1)) & d(3,1) \\ . & . & . & . \; . \; . \; . & . & . \\ . & . & . & . \; . \; . \; . & . & . \\ . & . & . & . \; . \; . \; . & . & . \\ . & . & . & . \; . \; . \; . & . & . \\ d((m-1),1) & d((m-1),2) & d((m-1),3) & . \; . \; . \; . & d((m-1),(m-1)) & d((m-1),m) \\ d(m,1) & d(m,2) & d(m,3) & . \; . \; . \; . & d(m,(m-1)) & d(m,m) \end{bmatrix}..$$

We now note that the aforementioned Proximity Matrix is a Symmetric Matrix with all its diagonal elements equal to zero. Therefore, there are $\left(\dfrac{m^2-m}{2}\right)$ number of Proximity values. We now order these in an ascending order. Let these be $P=\left\{r_1, r_2, r_3, \ldots\ldots\ldots, r_{\left(\frac{m^2-m}{2}\right)-1}, r_{\left(\frac{m^2-m}{2}\right)}\right\}$.

We now find all the Clusters which satisfy the specified distance of $0<x<r_1$ that characterizes a Connectivity based cluster (as specified in the definition already). That is, we find all the clusters which satisfy the following property:

A Cluster is a collection of Points (or objects) wherein they are scattered (their property is distributed) in such a fashion that, for a specified distance (measured in appropriate Metric of concern using appropriate Norm of concern) every point of this cluster has at least one neighbouring point also belonging to this cluster located within a certain small neighbourhood of this specified distance, (say for example $0<x<r_1$) measured from the aforementioned point of concern.

Similarly, we repeat the above analysis for each of $r_1<x<r_2$, $r_2<x<r_3$,........, $r_{\left(\frac{m^2-m}{2}\right)-1}<x<r_{\left(\frac{m^2-m}{2}\right)}$. In this fashion, we find all Possible Clusters.

References

1. http://www.philica.com/advancedsearch.php?author=12897
2. http://www.vixra.org/author/ramesh_chandra_bagadi