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 Connectivity Based 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 and located within a specified interval of distance.

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 $\left(\frac{m^2 - m}{2}\right)$ 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 define Proximity Contrast Ratio and Proximity Full Contrast Ratio.

Proximity Contrast Ratio

We define the Proximity Contrast Ratio as $\delta_{\frac{P(p,q)}{P(l,m)}} = \frac{P(p,q)}{P(l,m)}$ with only those values of $P(p,q) \neq 0$ and $P(l,m) \neq 0$ with $p, q, l, m = 1\ to M$. Furthermore, $p \neq l$ and $q \neq m$ simultaneously. That is, simply put $P(p,q) \neq P(l,m)$.

Proximity Full Contrast Ratio

We now define the *Proximity Full Contrast Ratio* $\delta_{\frac{Min}{Max}} = \frac{Min(P(i,j))}{Max(P(i,j))}$ with only those values of $P(i,j) \neq 0$. Also, $i, j = 1\ to M$.

Clustering Analysis

We now find all the Clusters based on connectivity definition of a Cluster wherein the connectivity distance is given by the following distances.

$$0 < x < r_1 + \left(\delta_{\frac{Min}{Max}}\right) r_1$$

$$r_1 + \left(\delta_{\frac{Min}{Max}}\right) r_1 < x < r_2 + \left(\delta_{\frac{Min}{Max}}\right) r_2$$

$$r_2 + \left(\delta_{\frac{Min}{Max}}\right) r_2 < x < r_3 + \left(\delta_{\frac{Min}{Max}}\right) r_3$$

……………………………………..

………………………………………..

………………………………………………

And so on so forth till we are exhausted with the given points in the act of clustering. Similarly, we repeat the above analysis using each of $\delta_{\frac{P(p,q)}{P(l,m)}}$ in place of $\delta_{\frac{Min}{Max}}$. 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