Holistic Non-Unique Clustering

ISSN 1751-3030

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 have presented a novel method to find all Possible Clusters given a set of M points in N Space.

Theory

Given M number of points $\bar{x}_i \in R^N, i = 1 \; to \; M$, each belonging to R^N, we first find the Proximity Matrix P_{ij} for each (M number of) point with each of all (M Number of points) points, inclusive of itself. The Proximity can be found using Euclidean distance or using the concept stated in [1]. We now find the *Proximity Full Contrast Ratio* $\delta_{\frac{Min}{Max}} = \frac{Min(P_{ij})}{Max(P_{ij})}$ *with only those values of* $P_{ij} \neq 0$. Now, we consider any $P(i,j)$ which are $\left(\frac{M^2 - M}{2}\right)$ in number as The Proximity Matrix is Symmetric and also all the diagonal elements are equal to zero, and compute the distance $d\left\{P(i,j), \delta_{\frac{Min}{Max}}\right\} = P(i,j) \pm \left(\delta_{\frac{Min}{Max}}\right)(P(i,j))$. Now, we consider any point $\bar{x}_i \in R^N$ and find all points (inclusive of $\bar{x}_i$) that have at least one neighbouring point within the distance $d\left\{P(i,j), \delta_{\frac{Min}{Max}}\right\}$, considered among themselves. We say that all such points form one

Cluster. In this fashion, we can find at most $\left(\frac{M^2-M}{2}\right)$ number of overlapping Clusters where the membership of a point may not be unique to a given Cluster. We call this type of Clustering as Holistic Non-Unique Clustering. Also, we can consider, all possible Proximity Contrast Ratio's among the $\left(\frac{M^2-M}{2}\right)$ number of unique elements in the Proximity Matrix and can get at most $2\left(\frac{M^2-M}{2}\right)$ number of overlapping Clusters for each of the ${}^{\left(\frac{M^2-M}{2}\right)}C_2$ number of possible Proximity Contrast Ratio's Possible. Therefore, we can see at most $2\left\{{}^{\left(\frac{M^2-M}{2}\right)}C_2\right\}\left(\frac{M^2-M}{2}\right)$ number of clusters for the given Set of M Points.

Here, the *Proximity Contrast Ratio* can be defined as *Proximity Contrast Ratio*

$$\delta_{\frac{i,j}{l,m}} = \frac{P(i,j)}{P(l,m)} \textit{ with only those values of } P_{ij} \neq 0 \textit{ and } i = 1 \textit{ to } M, j = 1 \textit{ to} M \textit{ with}$$

$$(i,j) \neq (l,m), \textit{ i.e., } i \neq l \textit{ and } j \neq m \textit{ simultaneously. Non simultaneously, they can be equal.}$$

References

1. Bagadi, R. (2017). Using the Appropriate Norm In The K-Nearest Neighbours Analysis. ISSN 1751-3030. *PHILICA.COM Observation number 173.* http://www.philica.com/display_observation.php?observation_id=173
2. http://www.philica.com/advancedsearch.php?author=12897
3. http://www.vixra.org/author/ramesh_chandra_bagadi