Abstract

Fuzzy clustering techniques handle the fuzzy relationships among the data points and with the cluster centers (may be termed as cluster fuzziness). On the other hand, distance measures are important to compute the load of such fuzziness. These are the two important parameters governing the quality of the clusters and the run time. Visualization of multidimensional data clusters into lower dimensions is another important research area to note the hidden patterns within the clusters. This paper investigates the effects of cluster fuzziness and three different distance measures, such as Manhattan distance (MH), Euclidean distance (ED), and Cosine distance (COS) on Fuzzy c-means (FCM) and Fuzzy k-nearest neighborhood (FkNN) clustering techniques, implemented on Iris and extended Wine data. The quality of the clusters is assessed based on (i) data discrepancy factor (i.e., DDF, proposed in this study), (ii) cluster size, (iii) its compactness, (iv) distinctiveness, (v) execution time taken, and (vi) cluster fuzziness (m) values. The study observes that FCM handles the cluster fuzziness better than FkNN. MH distance measure yields best clusters with both FCM and FkNN. Finally, best clusters are visualized using a Self Organizing Map (SOM).
Comparative Study of Fuzzy k-Nearest Neighbor and Fuzzy C-means Algorithms

Comparative Study of Fuzzy k-Nearest Neighbor and Fuzzy C-means Algorithms


Index Terms

Computer Science
Artificial Intelligence
Keywords
Fuzzy clusters  FkNN  FCM  Cluster fuzziness  Data discrepancy factor (DDF)