Abstract

Patterns and useful trends in large datasets has attracted considerable interest recently, and one of the most widely studied problems in this area is the identification and formation of clusters, or densely populated regions in a dataset. Prior work does not adequately address the problem of large datasets and minimization of I/O costs. The objective of this paper is to present a fuzzy logical approach towards clustering to refine the results obtained from the previous approach; Triangle-density based clustering technique (TDCT) [1], which was proposed in an earlier research paper in 2012. We hence name this algorithm as Soft Triangle Density Based Clustering Technique (STDCT). This algorithm incorporates soft clustering and is capable of identifying embedded clusters of arbitrary shapes as well as multi-density clusters over large spatial datasets with precision. Experimental results are reported to establish the superiority of the technique in terms of cluster quality and complexity. Fig. 1 depicts the formation of clusters of similar data.
- M. Ester, H. P. Kriegel, J. Sander and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise", in International Conference on Knowledge Discovery in Databases and Data Mining (KDD-96), Portland, Oregon, 1996, pp. 226-231.
- G. Sheikholeslami, S. Chatterjee and A. Zhang, "Wavecluster: A Multiresolution Clustering approach for very large spatial database", in SIGMOD&apos;98, Seattle, 1998.

- L. Ertoz, M. Steinbach and V. Kumar, "Finding Clusters of Different Sizes, Shapes, and Densities in Noisy, High Dimensional Data", in SIAM International Conference on Data Mining (SDM&apos;03), 2003.


**Index Terms**

Computer Science | Artificial Intelligence

**Keywords**

Clustering; Density-based; Triangle Density; Polygon Approach; Fuzzy Clustering