Abstract

The outlier detection problem in some cases is similar to the classification problem. For example, the main concern of clustering-based outlier detection algorithms is to find clusters and outliers, which are often regarded as noise that should be removed in order to make more reliable clustering.
In this article, we present an algorithm that provides outlier detection and data clustering simultaneously. The algorithm improves the estimation of centroids of the generative distribution during the process of clustering and outlier discovery. The proposed algorithm consists of two stages. The first stage consists of improved genetic k-means algorithm (IGK) process, while the second stage iteratively removes the vectors which are far from their cluster centroids.

Reference

knearestneighbour graph. In 17th International Conference on Pattern Recognition (ICPR 2004), Cambridge, United Kingdom, pp.430-433.

**Index Terms**

Computer Science | Data Mining

**Key words**

Outlier detection | Genetic algorithms | Clustering

K-means algorithm

Improved Genetic K-means (IGK)