Abstract

Organizing data into semantically more meaningful is one of the fundamental modes of understanding and learning. Cluster analysis is a formal study of methods for understanding and algorithm for learning. K-mean clustering algorithm is one of the most fundamental and simple clustering algorithms. When there is no prior knowledge about the distribution of data sets, K-mean is the first choice for clustering with an initial number of clusters. In this paper a novel distance metric called Design Specification (DS) distance measure function is integrated with K-mean clustering algorithm to improve cluster accuracy. The K-means algorithm with proposed distance measure maximizes the cluster accuracy to 99.98% at P = 1.525, which is determined through the iterative procedure. The performance of Design Specification (DS) distance measure function with K-mean algorithm is compared with the performances of other standard distance functions such as Euclidian, squared Euclidean, City Block, and Chebshew similarity measures deployed with K-mean algorithm. The proposed method is evaluated on the engineering materials database. The experiments on cluster analysis and the outlier profiling show that these is an excellent improvement in the performance of the proposed method.
- Jiawei Han, Micheline Kamber and Jian Pie (2012) "Data Mining: Concepts and Techniques", Margan Kaufmann Publishers, 3rd edition.
- JSTOR, 2009, JOTOR