

Jordan University of Science and Technology

A Novel Approach for Initializing the Spherical K-means Clustering Algorithm

Authors: Rehab Duwairi, Mohammed Abu Rehmeh

Abstract: In this paper, a novel approach for initializing the spherical K-means algorithm is proposed. It is based on calculating well distributed seeds across the input space. Also, a new measure for calculating vectors' directional variance is formulated, to be used as a measure of clusters' compactness. The proposed initialization scheme is compared with the classical Kmeans where initial seeds are specified randomly or arbitrarily on two datasets. The assessment was based on three measures: an objective function that measures intra cluster similarity, cluster compactness and time to converge. The proposed algorithm (called initialized K-means) outperforms the classical (random) K-means when intra cluster similarity and cluster compactness were considered for several values of k (number of clusters). As far as convergence time is concerned, the initialized K-means converges faster than the random K-means for small number of clusters. For a large number of clusters the time necessary to calculate the initial clusters' seeds start to outweigh the convergence criterion in time. The exact number of clusters at which the proposed algorithm starts to change behavior is data dependent (=11 for dataset1 and = 15 for dataset2).