AN EFFICIENT ACTIVE CONSTRAINT SELECTION ALGORITHM FOR CLUSTERING

Vũ Việt Vũ, Nicolas Labroche, and Bernadette Bouchon-Meunier

TÓM TẮT:

In this paper, we address the problem of active query selection for clustering with constraints. The objective is to determine automatically a set of queries and their associated must-link and can-not link constraints to help constraintsbased clustering algorithms to converge. Some works on activeconstraints learning have already been proposed but they are only applied to K-Means like clustering algorithms which are known to be limited to spherical clusters while we are interested in constraints-based clustering algorithms that deals with clusters of arbitrary shapes and sizes (like Constrained-DBSCAN, Constrained-Hierarchical Clustering . . .). Our novel approach relies on a k-nearest neighbors graph to estimate the dense regions of the data space and generates queries at the frontier between clusters where the cluster membership is most uncertain. Experiments show that our framework improves the performance of constraints based clustering algorithms.