

BOOSTING CLUSTERING BY ACTIVE CONSTRAINT SELECTION

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 user queries to define a set of must-link or cannot-link constraints. Some works on active constraint learning have already been proposed but they are mainly applied to K-Means like clustering algorithms which are known to be limited to spherical clusters, while we are interested in clusters of arbitrary sizes and shapes. The novelty of our approach relies on the use of a k-nearest neighbor graph to determine candidate constraints coupled with a new constraint utility function. Comparative experiments conducted on real datasets from machine learning repository show that our approach significantly improves the results of constraints based clustering algorithms.