

AN EFFICIENT METHOD FOR ACTIVE SEMI-SUPERVISED DENSITY BASED CLUSTERING

Vũ Việt Vũ,

TÓM TẮT:

Semi-supervised clustering algorithms relies on side information, either labeled data (seeds) or pairwise constraints (must-link or cannot link) between data objects, to improve the quality of clustering. This paper proposes to extend an existing seed-based clustering algorithm with an active learning mechanism to collect pairwise constraints. My new semi-supervised algorithm can deal with both seeds and constraints. Experiment results on real data sets show the efficient of my algorithm when compared to the initial seed-based clustering algorithm.