

IMPROVING CONSTRAINED CLUSTERING WITH ACTIVE QUERY SELECTION

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

TÓM TẮT:

In this article, we address the problem of automatic constraint selection to improve the performance of constraint-based clustering algorithms. To this aim we propose a novel active learning algorithm that relies on a k -nearest neighbors graph and a new constraint utility function to generate queries to the human expert. This mechanism is paired with propagation and refinement processes that limit the number of constraint candidates and introduce a minimal diversity in the proposed constraints. Existing constraint selection heuristics are based on a random selection or on a min-max criterion and thus are either inefficient or more adapted to spherical clusters. Contrary to these approaches, our method is designed to be beneficial for all constraint-based clustering algorithms. Comparative experiments conducted on real datasets and with two distinct representative constraintbased clustering algorithms show that our approach significantly improves clustering quality while minimizing the number of human expert solicitations.