

**Greedy Optimization for K-Means-Based Consensus Clustering**

**Abstract:**

Consensus clustering aims to fuse several existing basic partitions into an integrated one; this has been widely recognized as a promising tool for multi-source and heterogeneous data clustering. Owing to robust and high-quality performance over traditional clustering methods, consensus clustering attracts much attention, and much efforts have been devoted to develop this field. In the literature, the K-means-based Consensus Clustering (KCC) transforms the consensus clustering problem into a classical K-means clustering with theoretical supports and shows the advantages over the state-of-the-art methods. Although KCC inherits the merits from K-means, it suffers from the initialization sensitivity. Moreover, the current consensus clustering framework separates the basic partition generation and fusion into two disconnected parts. To solve the above two challenges, a novel clustering algorithm, named Greedy optimization of K-means-based Consensus Clustering (GKCC) is proposed. Inspired by the well-known greedy K-means that aims to solve the sensitivity of K-means initialization, GKCC seamlessly combines greedy K-means and KCC together, achieves the merits inherited by GKCC and overcomes the drawbacks of the precursors. Moreover, a 59-sampling strategy is conducted to provide high-quality basic partitions and accelerate the algorithmic speed. Extensive experiments on 36 benchmark datasets demonstrate the significant advantages of GKCC over KCC and KCC++ in terms of the objective function values and standard deviations and external cluster validity.