

doi:10.1631/FITEE.1700262

**题目：**一种基于锚点的谱聚类方法

**概要：**谱聚类是模式识别、机器学习和数据挖掘中最流行最重要的聚类方法之一。然而，高计算复杂度妨碍了谱聚类在大规模数据集的应用。对于具有  $n$  个样本的聚类问题，谱聚类需  $O(n^3)$  时间复杂度计算图拉普拉斯矩阵特征向量。为解决该问题，提出一种新的基于锚点谱聚类方法 (anchor-based spectral clustering, ASC)。首先，在数据集中选择  $m$  ( $m \ll n$ ) 个可以基本保持原始数据内在 (流形) 结构的锚点。然后，构造原始数据与锚点之间的映射矩阵，并证明该映射矩阵能保持数据的聚类结构。基于该映射矩阵，近似得到原始数据谱嵌入。ASC 方法复杂度与数据集大小呈线性关系。将该方法与经典谱聚类方法和两种最新谱聚类加速方法，即能量迭代聚类 (power iteration clustering, PIC) 和基于地标聚类 (landmark-based spectral clustering, LSC)，在 10 个真实数据集上比较。实验结果表明，ASC 算法比经典谱聚类算法具有更快聚类速度，在效率和有效性上与现有方法相当或优于现有方法。

**关键词：**聚类；谱聚类；图拉普拉斯；锚点