

数据库、信号与信息处理

## 半监督的改进K-均值聚类算法

汪军<sup>1, 2</sup>, 王传玉<sup>2</sup>, 周鸣争<sup>1</sup>

1.安徽工程科技学院 计算机科学与工程系, 安徽 芜湖 241000

2.安徽工程科技学院 应用数理系, 安徽 芜湖 241000

收稿日期 2009-4-8 修回日期 2009-6-11 网络版发布日期 2009-9-29 接受日期

**摘要** K-均值聚类算法必须事先获取聚类数目, 并且随机地选取聚类初始中心会造成聚类结果不稳定, 容易在获得一个局部最优值时终止。提出了一种基于半监督学习理论的改进K-均值聚类算法, 利用少量标签数据建立图的最小生成树并迭代分裂获取K-均值聚类算法所需要的聚类数和初始聚类中心。在IRIS数据集上的实验表明, 尽管随机样本构造的生成树不同, 聚类中心也不同, 但聚类是一致且稳定的, 迭代的次数较少, 验证了该文算法的有效性。

**关键词** [半监督学习](#) [K-均值聚类](#) [标签样本](#) [最小生成树](#)

**分类号** [TP391.4](#)

## Semi-supervised improved K-means clustering algorithm

WANG Jun<sup>1, 2</sup>, WANG Chuan-yu<sup>2</sup>, ZHOU Ming-zheng<sup>1</sup>

1.Department of Computer Science & Engineering, Anhui University of Technology and Science, Wuhu, Anhui 241000, China

2.Department of Math & Physics, Anhui University of Technology and Science, Wuhu, Anhui 241000, China

### Abstract

K-means clustering algorithm acquires the number of clusters in advance. The random selection of the initial cluster centers will result in the instability and K-means clustering algorithm will be terminated in access to a local optimum value. In order to solve the problem, the improved K-means clustering algorithm based on semi-supervised learning theory obtains the number of clustering and initial clustering centers after building minimum spanning tree used by few label samples and splitting it iteratively. Although minimum spanning tree making up of random samples and initial clustering centers are different, the clustering is consistent and stable; the iteration is less than traditional K-means algorithm. It proves that the semi-supervised improved K-means algorithm is effective.

**Key words** [semi-supervised learning](#) [K-means clustering](#) [labeled sample](#) [minimum spanning tree](#)

DOI: 10.3778/j.issn.1002-8331.2009.28.041

通讯作者 汪军 [wangjun@auts.edu.cn](mailto:wangjun@auts.edu.cn)

### 扩展功能

#### 本文信息

▶ [Supporting info](#)

▶ [PDF\(577KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [参考文献](#)

#### 服务与反馈

▶ [把本文推荐给朋友](#)

▶ [加入我的书架](#)

▶ [加入引用管理器](#)

▶ [复制索引](#)

▶ [Email Alert](#)

▶ [文章反馈](#)

▶ [浏览反馈信息](#)

#### 相关信息

▶ 本刊中 [包含“半监督学习”的相关文章](#)

▶ [本文作者相关文章](#)

· [汪军](#)

·

· [王传玉](#)

·

· [周鸣争](#)