

数据库、信息处理

基于Seed集的半监督核聚类

李昆仑, 张超, 曹铮, 刘明

河北大学 电子信息工程学院, 河北 保定 071002

收稿日期 2009-1-9 修回日期 2009-4-1 网络版发布日期 2009-7-9 接受日期

摘要 提出了一种新的半监督核聚类算法——SKK-均值算法。算法利用一定数量的标记样本构成seed集, 作为监督信息来初始化K-均值算法的聚类中心, 引导聚类过程并约束数据划分; 同时还采用了核方法把输入数据映射到高维特征空间, 并用核函数来实现样本之间的距离计算。在UCI数据集上进行了数值实验, 并与K-均值算法和核-K-均值算法进行了比较。

关键词 [半监督聚类](#) [seed集](#) [核方法](#) [K-均值](#)

分类号

Semi-supervised kernel clustering algorithm based on seed set

LI Kun-lun, ZHANG Chao, CAO Zheng, LIU Ming

College of Electronic and Information Engineering, Hebei University, Baoding, Hebei 071002, China

Abstract

This paper presents a novel semi-supervised kernel clustering algorithm called Seed Kernel K-Means (SKK-Means) algorithm. It uses labeled data to generate initial seed clusters to guide the clustering process and data partition, and uses kernel method to map the input data into a high-dimensional feature space and calculates the distance between data points with a kernel function. The algorithm is compared with the other clustering algorithms such as K-Means and Kernel K-Means, on UCI databases in some numeric experiment.

Key words [semi-supervised clustering](#) [seed set](#) [kernel method](#) [K-means](#)

DOI: 10.3778/j.issn.1002-8331.2009.20.046

通讯作者 李昆仑 likunlun@hbu.cn

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF\(636KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [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“半监督聚类”的相关文章](#)

▶ [本文作者相关文章](#)

· [李昆仑](#)

· [张超](#)

· [曹铮](#)

· [刘明](#)