

基于稀疏Parzen窗密度估计的快速自适应相似度聚类方法

钱鹏江, 王士同, 邓赵红

1. 江南大学信息工程学院 无锡 214122

2. 江南大学数字媒体学院 无锡 214122

收稿日期 2010-5-11 修回日期 2010-7-30 网络版发布日期 接受日期

摘要

相似度聚类方法(Similarity-based clustering method, SCM)因其简单易实现和具有鲁棒性而广受关注. 但由于内含相似度聚类算法 (Similarity clustering algorithm, SCA)的高时间复杂度和凝聚型层次聚类 (Agglomerative hierarchical clustering, AHC) 的高空间复杂度, SCM不适用大数据集场合. 本文首先发现了SCM和核密度估计问题的本质联系, 并以此入手, 通过快速压缩集密度估计器(Fast reduced set density estimator, FRSDE)和基于图的松弛聚类(Graph-based relaxed clustering, GRC)算法提出了快速自适应相似度聚类方法(Fast adaptive similarity-based clustering method, FASCM). 相比于原SCM, 该方法的主要优点是: 1)其总体渐近时间复杂度与样本容量呈线性关系; 2)不依赖于人工经验的干预, 具有了自适应性. 由此, FASCM适用于大数据集环境. 该方法的有效性在图像分割应用中进行了验证.

关键词 [相似度聚类](#) [密度估计](#) [时间复杂度](#) [图像分割](#)

分类号

Fast Adaptive Similarity-based Clustering Using Sparse Parzen Window Density Estimation

QIAN Peng-Jiang, WANG Shi-Tong, DENG Zhao-Hong

1. School of Information Technology, Jiangnan University, Wuxi 214122

2. School of Digital Media, Jiangnan University, Wuxi 214122

Abstract

Similarity-based clustering method (SCM) has received much attention because it is robust and can be implemented simply and easily. However, because of its high time complexity of the embedded similarity clustering algorithm (SCA) and high space complexity of the embedded agglomerative hierarchical clustering (AHC), SCM is impractical for large data sets. In this paper, the relationship is revealed between SCM and the kernel density estimation of samples, a novel fast adaptive similarity-based clustering method (FASCM) is accordingly proposed by adopting fast reduced set density estimator (FRSDE) and graph-based relaxed clustering (GRC). The distinctive advantages of FMSSC over MSSC exist in: 1) its asymptotic linear time complexity with the data size; 2) independent on artificial experience and its adaptability. Thus, FASCM is practical for large datasets. Its effectiveness has also been demonstrated in image segmentation examples.

Key words [Similarity-based clustering](#) [density estimator](#) [time complexity](#) [image segmentation](#)

DOI: 10.3724/SP.J.1004.2011.00179

通讯作者 钱鹏江 qianpjiang@126.com

作者个人主页 钱鹏江; 王士同; 邓赵红

扩展功能
本文信息
▶ Supporting info
▶ PDF (2758KB)
▶ [HTML全文] (OKB)
▶ 参考文献[PDF]
▶ 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
相关信息
▶ 本刊中 包含“相似度聚类”的 相关文章
▶ 本文作者相关文章
<ul style="list-style-type: none"> • 钱鹏江 • 王士同 • 邓赵红