

# 基于自然邻居的 聚类分析和离群检测 算法研究

Clustering Analysis and  
Outlier Detection  
based on Natural Neighbor

程东东 黄金龙 朱庆生 著



上海交通大学出版社  
SHANGHAI JIAO TONG UNIVERSITY PRESS

# 基于自然邻居的 聚类分析和离群检测算法研究

Clustering Analysis and Outlier Detection  
based on Natural Neighbor

程东东 黄金龙 朱庆生 著

 上海交通大学出版社  
SHANGHAI JIAO TONG UNIVERSITY PRESS

## 内容提要

本书是有关数据挖掘中基于邻域概念的聚类分析和离群检测算法的专著。主要内容如下:在自然邻居概念的基础上提出了局部核心点、准聚类中心、自然核心点等概念;在局部核心点、准聚类中心和自然核心点概念的基础上提出了适用于复杂流形结构数据集的基于中心的聚类算法、层次的聚类算法和最小生成树聚类算法;对聚类评价标准概念及其现状进行分析,并提出了新的适用于复杂流形数据的聚类评价标准;还提出了无参的离群点检测算法,并利用互为邻域图的信息,提出了离群簇检测算法。本书可供具有一定数据挖掘基础,并对聚类离群感兴趣的学生和研究者参考。

## 图书在版编目(CIP)数据

基于自然邻居的聚类分析和离群检测算法研究 / 程  
东东, 黄金龙, 朱庆生著. —上海: 上海交通大学出版  
社, 2019  
ISBN 978-7-313-22070-7

I. ①基… II. ①程… ②黄… ③朱… III. ①数据采  
集—研究 IV. ①TP274

中国版本图书馆 CIP 数据核字(2019)第 230986 号

## 基于自然邻居的聚类分析和离群检测算法研究

Ji Yu Ziran Linju De Julei Fenxi He Liqun Jiance Suanfa Yanjiu

著 者: 程东东 黄金龙 朱庆生

出版发行: 上海交通大学出版社

邮政编码: 200030

印 制: 江苏凤凰数码印务有限公司

开 本: 710 mm×1000 mm 1/16

字 数: 254 千字

版 次: 2019 年 11 月第 1 版

书 号: ISBN 978-7-313-22070-7

定 价: 58.00 元

地 址: 上海市番禺路 951 号

电 话: 021-64071208

经 销: 全国新华书店

印 张: 14.75

印 次: 2019 年 11 月第 1 次印刷

版权所有 侵权必究

告读者: 如发现本书有印装质量问题请与印刷厂质量科联系

联系电话: 025-83657309

## 目录 | CONTENTS

第1章 绪论 .....	1
1.1 数据挖掘的研究背景和意义 .....	1
1.2 聚类分析和离群检测的研究现状 .....	5
1.2.1 聚类分析研究现状 .....	5
1.2.2 离群检测研究现状 .....	9
1.3 本章小结 .....	12
第2章 聚类分析和离群检测 .....	13
2.1 聚类分析 .....	13
2.1.1 聚类分析概述 .....	13
2.1.2 聚类算法概述 .....	15
2.1.3 聚类结果评估 .....	25
2.2 离群检测 .....	30
2.2.1 离群的定义 .....	30
2.2.2 离群检测算法概述 .....	31
2.2.3 离群检测评价 .....	35
2.3 本章小结 .....	36
第3章 基于自然邻居的局部核心点和准聚类中心 .....	37
3.1 自然邻居 .....	37
3.2 基于自然邻居的局部核心点 .....	41
3.3 准聚类中心及自然核心点 .....	45



3.3.1	准聚类中心 .....	46
3.3.2	自然核心点 .....	51
3.4	本章小结 .....	53
<b>第 4 章</b>	<b>基于中心的聚类算法 .....</b>	<b>54</b>
4.1	现有的基于中心的聚类算法存在的问题 .....	54
4.2	基于局部核心点的 DP 算法 DPLORE .....	56
4.2.1	基于局部核心点的 DP 算法基本思想 .....	56
4.2.2	实验结果及分析 .....	59
4.3	基于准聚类中心的聚类算法 QCC .....	71
4.3.1	QCC 算法的基本思想 .....	71
4.3.2	QCC 算法实验及分析 .....	74
4.4	本章小结 .....	89
<b>第 5 章</b>	<b>层次聚类算法 .....</b>	<b>91</b>
5.1	现有的层次聚类算法存在的不足 .....	91
5.2	基于自然邻居的层次聚类算法 Hi-CLUBS .....	92
5.2.1	Hi-CLUBS 算法基本思想 .....	92
5.2.2	实验及分析 .....	97
5.3	基于局部核心点的层次聚类算法 HCLORE .....	101
5.3.1	HCLORE 算法基本思想 .....	101
5.3.2	实验及分析 .....	106
5.4	基于噪声去除的层次聚类算法 HCBNR .....	122
5.4.1	HCBNR 算法的基本思想 .....	122
5.4.2	实验及分析 .....	123
5.5	本章小结 .....	126
<b>第 6 章</b>	<b>最小生成树聚类算法 .....</b>	<b>127</b>
6.1	基于最小生成树的聚类算法存在的问题 .....	127
6.2	基于局部核心点的最小生成树算法 LORE-MST .....	129
6.2.1	LORE-MST 算法基本思想 .....	129

6.2.2	实验及分析	133
6.3	基于自然核心点的最小生成树聚类分析算法 NCP-MST	145
6.3.1	NCP-MST 算法过程	146
6.3.2	实验分析	150
6.4	本章小结	158
第 7 章	聚类结果内部评估方法	160
7.1	现有的聚类结果内部评估方法存在的不足	160
7.2	基于局部核心点的聚类结果评估方法	161
7.2.1	局部核心点之间基于图的距离	162
7.2.2	LCCV 指标	162
7.2.3	LCCV 指标分析	163
7.2.4	基于 LCCV 指标的层次聚类算法	165
7.3	实验结果及分析	166
7.3.1	人工数据集上的实验结果	166
7.3.2	真实数据集上的实验结果	174
7.3.3	LCCV 指标与其他聚类算法结合的实验结果	176
7.4	讨论	179
7.4.1	局部核心点的合理性	179
7.4.2	噪声点的影响	180
7.5	本章小结	182
第 8 章	离群检测	183
8.1	现有离群检测算法分析	183
8.2	基于自然邻居的无参离群检测算法 NOF	186
8.2.1	NOF 算法及相关定义	186
8.2.2	实验结果及分析	189
8.3	基于互为邻域图的离群簇检测算法 ROCF	197
8.3.1	ROCF 算法及相关定义	197
8.3.2	实验结果及分析	203
8.4	本章小结	210

第 9 章 总结和展望 .....	211
9.1 总结 .....	211
9.2 展望 .....	213
参考文献 .....	215
索引 .....	225

巍巍交大 百年书香  
www.jiaodapress.com.cn  
bookinfo@sjtu.edu.cn

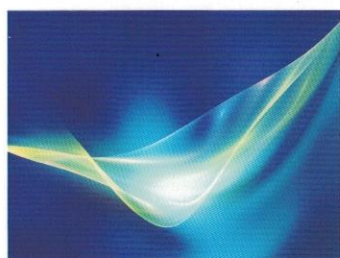


责任编辑 杨迎春  
蒋可玉  
封面设计 朱 懿



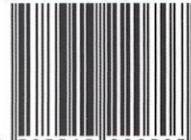
扫描二维码  
关注上海交通大学出版社  
官方微信





基于自然邻居的  
聚类分析和离群检测算法研究

ISBN 978-7-313-22070-7



9 787313 220707 >

定价: 58.00元