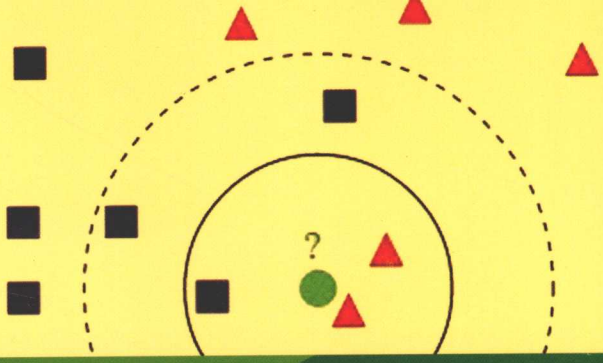


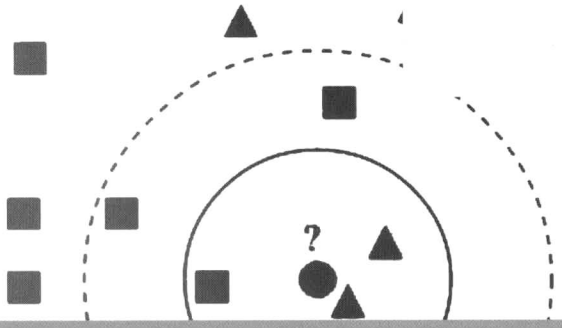
近邻分类方法及其应用

[下册]

郭躬德 陈黎飞 李南 ◎著

Nearest Neighbour Classification Method and its Applications

厦门大学出版社 XIAMEN UNIVERSITY PRESS
国家一级出版社
全国百佳图书出版单位



近邻分类方法及其应用

[下册]

郭躬德　陈黎飞　李南　◎著

厦门大学出版社 国家一级出版社
XIAMEN UNIVERSITY PRESS 全国百佳图书出版单位

图书在版编目(CIP)数据

近邻分类方法及其应用．下册/郭躬德，陈黎飞，李南著．—厦门：厦门大学出版社，2014.4
ISBN 978-7-5615-5056-4

Ⅰ.①近… Ⅱ.①郭…②陈…③李… Ⅲ.①数据采掘-算法分析 Ⅳ.①TP311.131

中国版本图书馆 CIP 数据核字(2014)第 068045 号

厦门大学出版社出版发行
(地址：厦门市软件园二期望海路 39 号　邮编：361008)
http://www.xmupress.com
xmup @ xmupress.com
厦门市明亮彩印有限公司印刷
2014 年 4 月第 1 版　2014 年 4 月第 1 次印刷
开本：787×1092　1/16　印张：15
插页：2　字数：341 千字
定价：38.00 元
本书如有印装质量问题请直接寄承印厂调换

前 言

计算机技术的普及应用给人类社会带来了深刻的变革，也使得我们所拥有的数据以前所未有的速度膨胀。随着大数据时代的到来，越来越多的人开始关注数据挖掘这一项大数据分析和处理的重要技术。作为数据挖掘的一种主要方法，分类(classification)——用于发掘隐藏在历史数据中的类别模式进而对未知事件做出预测或判断的技术——由于其很强的实用性成为了许多数据分析处理系统的基本构件。在机器学习领域，分类是有监督学习的代表性方法，其应用也已深入到信息检索、生物信息、客户关系管理等社会经济生活的方方面面。

近邻分类技术源来已久，可以追溯到早期对 k_n-NN 规则(E. Fix 等人，1951 年)和 k-NN 规则(T. Cover 等人，1967 年)的研究。随后发展起来的 kNN 分类算法由于其原理简单、易于实现、可扩展性好、可解释性强等优点，备受青睐，位列 2005 年 ICDM 国际会议遴选的 10 大最有影响力的数据挖掘算法之一。如今，近邻分类方法无论在理论模型、算法还是在领域应用方面都吸引了众多的研究者和实践者，呈现出蓬勃发展的态势，涌现了一大批改进型新算法，也提出了一些基于近邻分类思想的新模型、新方法。“欲穷千里目，更上一层楼”，跟踪了解近年来取得的这些新进展才能进一步推进近邻分类技术的研究和深入应用，这也是本书出版的首要目的所在。

著者之一郭躬德教授系英国 Ulster 大学归国学者，长期从事近邻分类方法的研究与应用，早于 2003 年提出称为 kNNModel 的近邻分类新方法，赢得业界热烈反响。近年来，在郭教授的带领下，福建师范大学数据挖掘与网络内容安全实验室开展了近邻分类理论方法与应用方面的系统研究，取得了一系列成果。在理论方法方面，研究团队提出了基于近邻思想的相似性度量新方法并将之推广到类属型数据，提出了增量学习、多代表点学习和

子空间近邻分类等新方法；应用研究涵盖了毒性物质预测、特征选择、文本分类以及数据流分类等近邻分类的新应用领域。本书将有关研究成果集结成册，以飨读者。

本书共六章，第一章介绍近邻分类算法及近年来的研究新进展，第二章至第六章每章介绍近邻分类的一类新方法。郭躬德主要编写本书的第二、三章，并参与编写第四、六章部分章节，约 20 万字；陈黎飞主要编写第四、五、六章，约 30 万字；李南主要编写第一章，并参与编写第四章部分章节，约 5 万字。研究生黄彧、陈红、辛轶、黄杰、李南、张健飞、卢伟胜、兰天，访问学者陈雪云等参加了有关研究工作和部分章节写作。在写作过程中，参考了大量的国内外文献资料，在此一并表示感谢。

本书内容有误或不妥之处，欢迎读者批评指正。

郭躬德　陈黎飞　李南
福建师范大学数学与计算机科学学院
2014 年 1 月

目 录

上 册

下　册