

图书基本信息

书名：<<多目标线性规划分类方法业绩分析与改进研究>>

13位ISBN编号：9787563819126

10位ISBN编号：7563819126

出版时间：2011-6

出版时间：首都经济贸易大学出版社

作者：朱梅红

页数：129

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

## 内容概要

社会发展离不开数据，而数据必须使用统计方法来加以分析。自威廉·配第《政治算术》始，历史上几乎每一次对社会经济发展的深刻理解都是建立在统计分析方法变革的基础上的。正是这种变革所提供的各种数据分析工具加深了人们对社会经济本质的理解，使得人们的认识能够还原真实世界并与之无限接近。统计学数百年的发展历经两次方法上的“革命”：从最初不完整的全面调查方法到大样本统计推断，是统计方法的第一次革命；以大样本统计推断方法为基础，进一步发展出小样本统计推断方法，是统计方法的第二次革命。这两次革命都是施于用样本数据推断总体特征这一思想，而抽样误差的干扰导致统计方法日益复杂，使其应用受到限制。目前，以数据挖掘方法为代表的统计学的第三次革命即将到来。数据挖掘是在继承已有统计理论的基础上，与计算机技术紧密结合，充分发挥计算机运算速度快、存储量大的特点，将统计方法从抽样推断向海量数据分析推进，是统计学、计算机技术、仿真计算、机器学习、人工智能甚至哲学思想相融合的新学科，体现了科学发展“螺旋式上升”的哲学内涵。

## 书籍目录

## 1 绪论

- 1.1 研究背景
- 1.2 问题的提出
- 1.3 基本概念
- 1.4 本书的研究内容与方法
- 1.5 本书的结构安排
- 1.6 本书的特色与贡献

## 2 文献综述

- 2.1 几种线性规划分类模型
- 2.2 多目标线性规划分类模型
- 2.3 分类方法业绩改进的一般技术
- 2.4 本章主要结论

## 3 MCLP的偏差和方差分析

- 3.1 关于MCLP三个特性的一般理论
- 3.2 期望预测误差的分解
- 3.3 数据准备与实验安排
- 3.4 实验结果与分析
- 3.5 本章主要结论

## 4 MCLP在不平衡数据集上的业绩分析

- 4.1 分类业绩评价标准及选择
- 4.2 数据不平衡对分类方法业绩影响机制的一般结论
- 4.3 数据不平衡对MCLP业绩影响机制的分析
- 4.4 数据不平衡对MCLP业绩影响的实证分析
- 4.5 最优类分布结论的稳定性分析
- 4.6 本章主要结论

## 5 组合分类器方法对MCLP的业绩改进分析

- 5.1 Bagging和Adaboost程序
- 5.2 数据准备与实验安排
- 5.3 两种基本组合方法对MCLP的业绩改进分析
- 5.4 一种SmoothBoosting方法对MCLP的业绩改进分析
- 5.5 一种Sequential : Bagging方法对MCLP的业绩改进分析
- 5.6 随机子空间方法对MCLP的业绩改进分析
- 5.7 本章主要结论

## 6 不平衡数据处理方法对MCLP的业绩改进分析

- 6.1 MCLP分类中对不平衡数据的一般处理
- 6.2 数据准备与实验安排
- 6.3 基于MCLP分类结果的数据集特性分析
- 6.4 随机上抽样和随机下抽样方法对MCLP的业绩改进分析
- 6.5 一种改进的单边抽样方法对MCLP的业绩改进分析
- 6.6 改进的单边抽样+上抽样方法对MCLP的业绩改进分析
- 6.7 一种正类加权的MCLP模型及其业绩改进分析
- 6.8 不同方法在信用卡数据集上的业绩比较分析
- 6.9 本章主要结论

## 7 总结与展望

- 7.1 主要研究结论

7.2 需要进一步研究的问题

参考文献

章节摘录

安全点的识别, 需要我们从MCLP对数据集的初步分类结果中进行仔细研究。

直观上看, 如果剔除靠左的即分值较小的负类点, 相当于新的训练集内负类点分布偏右, 此时训练集的线性可分程度增加。

但是, MCLP为使目标函数最优, 会综合考虑正负两类, 相应的分类边界也就会右移。

训练集上的分类结果, 必定是正类正确率提高。

如果将训练集的解代人测试集, 也必定会使测试集的负类正确率明显降低, 因为分类边界右移, 而测试集内负类点密集分布在边界处。

左边的负类点剔除得越多, 这种后果越严重。

相反, 如果剔除靠右的即分值较大的负类点, 相当于新的训练集内负类点分布偏左。

MCLP为使目标函数最优, 会综合考虑正负两类, 相应的分类边界就会左移。

训练集上的分类结果, 必定是正类正确率降低。

如果将训练集的解代人测试集, 也必定会使测试集的正类正确率明显降低, 因为分类边界左移, 而测试集内正类点也密集分布在边界处。

同样, 右边的负类点剔除得越多, 这种后果越严重。

所以, 只有小心地选择安全点, 才能使测试集上两类的分类结果比较均衡, AUC的值比较理想。

.....

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>