

模糊测度和模糊积分

及在分类技术中的应用

王熙照◎著



科学出版社
www.sciencep.com

模糊测度和模糊积分及在 分类技术中的应用

王熙照 著

科学出版社

北京

内 容 简 介

本书从模糊积分作为一种分类技术的角度予以论述,内容涵盖了模糊测度与积分的基础知识,模糊积分作为一种分类器和模糊积分作为一种分类器的融合工具等三个部分,重点强调了模糊积分在分类问题求解过程中的应用.

本书结构清晰,内容翔实,可作为应用数学、机器学习、模式识别、自然语言处理、智能控制等专业的高年级本科生、研究生的教材或参考书.

图书在版编目(CIP)数据

模糊测度和模糊积分及在分类技术中的应用/王熙照 著. —北京:科学出版社, 2008

ISBN 978-7-03-020631-2

I. 模… II. 王… III. 模糊数学-应用-分类机-研究 IV. TP322

中国版本图书馆 CIP 数据核字(2008)第 031081 号

责任编辑:范庆奎 房 阳 / 责任校对:桂伟利

责任印制:赵德静 / 封面设计:王 浩

科 学 出 版 社 出 版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencecp.com>

新 蕾 印 刷 厂 印 刷

科学出版社发行 各地新华书店经销

*

2008 年 3 月 第 一 版 开本: B5(720 × 1000)

2008 年 3 月 第 一 次 印 刷 印张: 15 1/4

印数: 1—3 000 字数: 283 000

定 价: 48.00 元

(如有印装质量问题,我社负责调换〈环伟〉)

前 言

自 Zadeh 于 1965 年提出模糊集合的概念以来, 模糊集理论的研究至今已形成了一个完整的体系. 模糊技术已经在众多的领域 (如模式识别、图像处理、医疗诊断、决策支持、知识处理、自动控制等) 有了广泛而深入的应用. 模糊测度和模糊积分作为模糊集理论的一个分支, 最早形成于 20 世纪 70 年代. 它作为经典测度和积分的延拓, 重点研究了一种非可加情况. 这种从可加到非可加的推广, 首先在数学理论上有着重要意义, 它进一步丰富了非线性分析理论. 其次, 由于可加性是一种理想状态, 大多数复杂的实际问题都是非可加的情况, 故使用模糊测度和模糊积分来作为数学模型比使用经典的测度和积分更为符合实际. 模糊测度和模糊积分作为描述非可加和非线性的数学模型有着许多应用, 本书主要论述这种数学模型在处理分类技术 (更详细地讲在单分类器设计和多分类器融合) 中的应用.

分类和聚类是知识处理的核心问题. 众所周知, 知识表示和处理、推理、搜索和学习构成当今人工智能领域的 4 个主题内容, 而分类技术从始至终贯穿这 4 个主题. 关于分类问题的研究, 迄今已有超过一个世纪的历史, 已有大量的分类算法问世. 例如, 遗传算法、文本分类算法、贝叶斯分类算法、遥感分类算法、聚类分类算法、数据挖掘分类算法、 k -NN 分类算法、SVM 分类算法、指纹分类算法等, 但由于分类问题从理论上讲是一个极复杂的函数延拓问题, 不存在一种最优的分类算法适用于各种不同的情况, 故至今仍有许多分类算法陆续出现. 随着模糊集的出现, 模糊技术开始与各种分类算法结合用于处理分类过程中经典分类算法不能处理的认知问题. 这种结合可分为两个方面: 一是用模糊集隶属函数表示与人类的思维和感觉有关的不确定因素; 二是将模糊测度和模糊积分作为一种新的分类工具用于分类器设计和融合. 本书的主要内容属于第二个方面.

随着计算技术的飞速发展, 传统分类问题中的一些不能解决的问题 (如求解大规模二次规划问题) 现在可在高性能计算机上得以实现. 然而, 由于大部分出现在分类问题中的学习搜索问题都是 NP 困难, 它们的精确求解至今尚无法实现. 探索各种启发式算法去近似求解大规模复杂问题仍是一种有意义的工作. 本书中的模糊测度的数值确定就是这样一类问题, 由于问题的复杂性和解的精确性始终是一对不可调和的矛盾, 各种不同类型的模糊测度的表示在二者之间起到了一个折中的作用.

本书是在研究国家自然科学基金课题“加权模糊规则的泛化能力研究”过程中形成的, 并受到河北省首届百名优秀人才支持计划的帮助. 参加本书撰写和讨论的

有冯慧敏、陈俊芬、王晓君、张素芳、沈君、王旭光、赵连保、赵丽、陈爱霞、刘丽荣、李少红、杨峰、张国防. 本书的部分内容参阅了王震源先生的《模糊测度论》和法国学者 Grabisch 的《基于模糊积分的分类》以及参考文献中所列出的其他作者的相关工作, 作者在此表示谢意. 本书的第一稿来自于 2003 级研究生讨论班“模糊测度和模糊积分在分类技术中的应用”, 第二稿曾使用于 2004 级数学专业研究生讨论班, 现在的版本是本书的第三稿.

由于水平所限, 书中遗漏和错误在所难免, 恳切希望读者指正.

王熙照

2007 年 9 月于河北大学

目 录

前言

符号说明

第 1 章 引言	1
1.1 什么是分类问题	1
1.2 分类问题的一种公式描述	7
1.3 分类器的训练、设计及模糊积分的应用	10
参考文献	13
第 2 章 模糊测度	14
2.1 经典测度的一些基本概念	14
2.2 非可加测度	18
2.3 g_λ 模糊测度	21
2.4 拟测度	26
2.5 信任测度和似然测度	29
2.6 可能性测度和必要性测度	34
2.7 k 可加模糊测度	39
2.8 模糊集合的模糊值测度	43
2.9 基于三角模的模糊测度	47
参考文献	51
第 3 章 模糊积分	53
3.1 Sugeno 模糊积分的定义及性质	53
3.2 对称 Sugeno 模糊积分	57
3.2.1 对称序结构	57
3.2.2 默比乌斯变换	59
3.3 模糊数值的 Sugeno 模糊积分	63
3.4 Choquet 模糊积分	68
3.5 Choquet 模糊积分的层次分解	72
3.6 推广的 Choquet 模糊积分	77
3.7 Sugeno 积分与 Choquet 积分的比较	84
3.7.1 一般模糊积分的定义	84
3.7.2 一般意义下的共同性质	85

3.7.3	有限集上 Sugeno 积分与 Choquet 积分的比较	86
3.8	(N) 模糊积分和广义模糊积分	89
3.8.1	(N) 模糊积分	89
3.8.2	广义模糊积分	92
3.9	Zhenyuan 模糊积分	100
3.9.1	Zhenyuan 模糊积分的定义	100
3.9.2	Zhenyuan 模糊积分的性质	101
3.9.3	Zhenyuan 模糊积分的计算	103
3.10	可能性积分	105
3.11	泛积分	110
3.12	基于集合划分的非线性积分	114
3.13	几种模糊积分的比较	120
3.13.1	几种模糊积分与 Lebesgue 积分的联系	120
3.13.2	Sugeno 模糊积分与 Choquet 模糊积分	121
3.13.3	泛积分与 Zhenyuan 模糊积分	122
3.13.4	Choquet 模糊积分与 Zhenyuan 模糊积分	123
3.13.5	Choquet 模糊积分与泛积分	124
3.13.6	Sugeno 模糊积分与 (N) 模糊积分	125
3.13.7	对称 Sugeno 积分	125
	参考文献	127
第 4 章	模糊积分分类器	129
4.1	模糊积分分类器的基本类型	129
4.2	模糊积分分类器中属性间的交互作用	136
4.2.1	三个工人的例子	136
4.2.2	属性间交互作用的类型	138
4.2.3	模糊测度表示交互作用的能力	139
4.2.4	Shapley 值和交互指标	140
4.2.5	实验数据分析	144
4.2.6	基于属性交互作用的特征选取方法	146
4.3	Choquet 模糊积分分类器中一般模糊测度的学习	148
4.3.1	基于二次规划的模糊测度的确定	149
4.3.2	二次规划的推广	153
4.3.3	基于遗传算法的非可加集函数的确定	155
4.3.4	确定模糊测度的最小训练样本数	161
4.4	Choquet 模糊积分分类器中特殊结构的模糊测度的学习	162

4.4.1	用神经网络确定 g_{λ} 模糊测度	163
4.4.2	用 Shapley 值和交互指标确定 2 可加模糊测度	171
4.4.3	线性规划学习 2 可加模糊测度	173
	参考文献	176
第 5 章	模糊积分在分类器融合中的应用	178
5.1	融合工具: 从 OWA 算子到模糊积分	178
5.1.1	分类器融合的基本方法	179
5.1.2	分类器融合的模糊积分方法	186
5.1.3	OWA 算子与 Choquet 模糊积分的关系	190
5.2	基于模糊积分的多分类器融合中模糊测度的作用	193
5.2.1	分类器的重要性及之间的交互作用	193
5.2.2	模糊测度的单调性对融合系统的影响	195
5.2.3	模糊积分融合方法的纠错能力	197
5.2.4	融合过程中模糊测度的确定	199
5.2.5	注记	202
5.3	基于模糊积分的同一类型分类器的融合实验分析	203
5.3.1	基于模糊积分的多神经网络分类器融合	203
5.3.2	基于模糊积分的多 k -NN 分类器融合	207
5.3.3	基于模糊积分的多决策树融合	209
5.4	融合过程中分类器多样性的作用	210
5.4.1	什么是多样性	211
5.4.2	多样性的几种度量方法	211
5.4.3	融合过程中多样性的作用	215
5.5	融合结果对不同模糊测度或模糊积分的敏感性	217
5.6	基于模糊积分融合与其他融合技术的比较	221
	参考文献	225
	索引	229

符号说明

大写字母, 如 A, B, X, E, F 等表示集合

X : 论域

\emptyset : 空集

\mathbb{R} : 实数集

\mathbb{R}^+ : 非负实数集

\mathcal{F} : X 上的集类

$\mathcal{P}(X)$: X 上的幂集

\tilde{E}, \tilde{F} : 模糊集

$\{E_n\}$: 集序列

\cup : 集合并

\cap : 集合交

$E\Delta F$: 集合 E, F 的对称差

$E - F$: 集合 E, F 的差集

E^c : 集合 E 的补集

$|E|$: 集合 E 的势 (基数)

μ, ν : 模糊测度

g_λ : g -lambda 测度

ℓ : Lebesgue 测度

g_i : g -lambda 测度的概率密度函数

μ_A : 模糊集 \tilde{A} 的隶属函数

A_λ : 模糊集 \tilde{A} 的 λ 截集

$\mathcal{F}^*(X)$: X 上的所有模糊子集的全体构成的集类

\tilde{a}, \tilde{b} : 模糊数

$\tilde{\rho}$: 模糊数的模糊距离

\mathcal{F}^* : 模糊数的全体

\mathcal{F}_-^* : 负模糊数的全体

\mathcal{F}_+^* : 正模糊数的全体

$\text{supp}(\tilde{A})$: 模糊子集 \tilde{A} 的支撑

$(X, \mathcal{F}^*(X))$: 模糊可测空间

\tilde{f} : 模糊值函数

$\int f d\mu$: 模糊积分

(c) $\int f d\mu$: Choquet 积分

(s) $\int f d\mu$: Sugeno 积分

(ss) $\int f d\mu$: 对称 Sugeno 积分

(w) $\int f d\mu$: Zhenyuan 积分

(D) $\int f d\mu$: 泛积分

(N) $\int f d\mu$: N 积分

\int_{τ} : 集于划分的积分

ω : 权重向量

(X, \mathcal{F}) : 可测空间

(X, \mathcal{F}, μ) : 测度空间

\mathbb{F} : 可测函数集

f, g : 可测函数

F_{α} : α 截集

$F_{\alpha+}$: α 强截集

χ_A : 集合 A 的特征函数

\mathcal{B}_+ : $[0, +\infty)$ Borel 集类

ε : 任意小的正数

(L^+, \leq) : 线性有序集

$\nabla, \bar{\wedge}$: 对称取大、取小算子

sgn: 符号函数

$E(f)$: 自然扩张

\oplus : 泛可加算子

\odot : 泛乘算子

τ : 可测划分

$s(x)$: 泛单点可测函数

$\prod_{i=1}^n a_i$: 连乘

$\sum_{i=1}^n a_i$: 连加

\lim_n : 极限

$D = \{D_1, D_2, \dots, D_L\}$: 分类器集合

$a := b$: 用 b 定义 a