

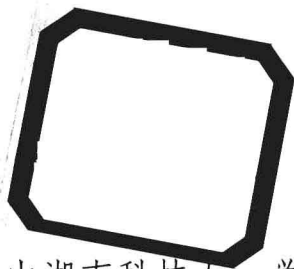


多光谱遥感图像 土地利用分类方法

林 剑 彭顺喜 邓吉秋 廖秀英 编著



中南大学出版社
www.csupress.com.cn



本书由湖南科技大学学术著作出版基金资助

多光谱遥感图像 土地利用分类方法

林 剑 彭顺喜 编著
邓吉秋 廖秀英



中南大学出版社
www.csupress.com.cn

图书在版编目(CIP)数据

多光谱遥感图像土地利用分类方法/林剑等编著. —长沙:中南大学出版社,2011.5

ISBN 978-7-5487-0227-6

I. 多... II. 林... III. 多光谱遥感-应用-土地分类
IV. F301

中国版本图书馆 CIP 数据核字(2011)第 039642 号

多光谱遥感图像土地利用分类方法

林 剑 彭顺喜 编著
邓吉秋 廖秀英

☐责任编辑 陈海波

☐责任印制 文桂武

☐出版发行 中南大学出版社

社址:长沙市麓山南路

邮编:410083

发行科电话:0731-88876770

传真:0731-88710482

☐印 装 长沙利君漾印刷厂

☐开 本 720 × 1000 B5 ☐印张 14 ☐字数 273 千字

☐版 次 2011 年 5 月第 1 版 ☐2011 年 5 月第 1 次印刷

☐书 号 ISBN 978-7-5487-0227-6

☐定 价 35.00 元

图书出现印装问题,请与出版社调换

内容简介

Introduction

本书介绍了多光谱遥感图像土地利用分类方法，主要包括：多光谱遥感图像土地利用类别分析方法及其图像特征，FasART 神经网络多光谱特征分类方法，纹理特征分开扩张分类方法，光谱特征与纹理特征基于模糊积分融合分类方法，区域多中心分类方法，遥感图像分类方法用于城市变化检测等。本书包含了实现主要算法的 C++ 程序，全书还附上大量的参考文献以便读者深入学习。

本书可供高等院校本科生、研究生以及工程技术人员学习参考。

前言

Foreword

随着对地观测技术的迅速发展,遥感图像在社会生活和经济建设中发挥着越来越重要的作用。实现遥感图像土地利用的计算机分类,对利用遥感技术进行国土资源管理、生态环境监测和土地生态评价等方面具有重要的意义。

第1章绪论,介绍了遥感图像土地利用的基本概念,概述了模式识别的基本内容,分析了国内外遥感图像土地利用分类的发展状况和趋势。

第2章遥感图像土地利用类别的多光谱特征,概述了随机变量分布的理论,介绍了一种基于聚类的多光谱像元空间距离分析方法,对土地利用类别的多光谱特征的分布规律进行分析。

第3章模糊自适应神经网络(FasART)分类方法,对土地利用类别像元进行了多中心与单一聚类中心的类内、类间距的计算分析,探讨了C-均值法、最大似然法和模糊神经网络分类方法的分类机理,实现了FasART神经网络分类算法。

第4章基于纹理特征的分类,概述了常用的纹理图像特征提取方法,分析了遥感图像多波段的特点,介绍了模糊纹理分析方法提取纹理特征,指出了分开-合并和多分辨率图像锥对多光谱遥感分割所存在的问题,针对多光谱遥感图像纹理特征的不规则性,实现了一种基于区域的多光谱遥感图像纹理分割方法——分开-扩张方法。

第5章模糊积分融合纹理特征与光谱特征分类,概述了数据融合的一般知识,分析了数据融合方法在遥感图像处理中的应用现状,简述了基于决策层融合的D-S理论,根据多光谱遥感图像光谱信息和纹理信息的特点,介绍了一种基于决策层的模糊积分融合方法,并且探讨了特征矢量的维数对分类器的输出的影响。

第6章基于区域多中心的多光谱遥感图像土地利用分类,分析了土地利用类别区域多中心特征,实现了一种基于区域多中心

的多光谱遥感图像分类方法,该方法把区域特征和多光谱特征有机地结合起来,较好地处理了区域特征和多光谱特征融合中矢量维数过多和设计分类器困难的问题,有效地提高了土地覆盖/土地利用分类精度。

第7章遥感图像分类方法在城市扩展变化检测中的应用,分析了遥感图像变化检测技术的发展现状和趋势,介绍了一种先比较后分类的城市扩展变化检测方法,该方法还减少了先分类后比较方法的分类次数,有效地解决了分类误差积累的问题,概述了RS与GIS相结合的长株潭城市遥感监测系统,对长沙、株洲、湘潭城市扩展进行了分析。

需要说明的是,本书中包括的方法实用于多光谱、高光谱和中等分辨率的遥感图像。

本书主要由湖南科技大学林剑博士和湖南省水库移民开发管理局彭顺喜博士等编写。彭顺喜博士负责第2章、第6章和第7章的编写,林剑博士负责第1章、第3章、第4章、第5章的编写,并对全书进行统编,此外,中南大学GIS中心邓吉秋参与了第7章中的RS与GIS应用系统的大部分编写,湖南科技大学廖秀英参与了第6章的部分工作。湖南科技大学研究生赵会芳、谭勇鸿、钟迎春、李乐、曾毅和刘兵等参与了基础材料和文字的整理工作,特此致谢。

本书是以林剑和彭顺喜的博士论文为基础,指导老师中南大学鲍光淑教授倾注了大量的心血,在此表示衷心地感谢!国防科技大学王润生教授对本书提出了重要的修改意见,特此感谢!

本书得到国家自然科学基金(编号:60673119)、湖南省自然科学基金(编号:04JJ3046,07JJ5066)和湖南省发展改革委员会基金(湘财企指[2008]149号)的资助。同时得到湖南科技大学知识处理与网络化制造湖南省普通高校重点实验室的大力支持,特此致谢。

本书中的主要算法均已实现,并附录了较为完整的VC++程序,但由于编者水平有限,书中难免存在一些缺点和错误,恳请专家和读者批评指正。

编者

2011年4月

第 1 章 绪论	1
1.1 土地利用	1
1.2 资源卫星及其多光谱遥感图像	2
1.3 模式识别技术概述	5
1.3.1 基本模型	6
1.3.2 特征归一化、提取与选择	6
1.3.3 分类器选择的原则	7
1.3.4 分类器融合	9
1.4 遥感图像分类国内外研究现状	11
1.4.1 分类特征的提取	11
1.4.2 分类器的选择	13
第 2 章 遥感图像土地利用类别的多光谱特征	16
2.1 随机变量(矢量)的基本理论	16
2.1.1 特征矢量和特征空间	16
2.1.2 随机矢量的描述	17
2.1.3 正态分布	20
2.2 聚类分析基础	21
2.2.1 模式相似性测度	22
2.2.2 类的定义与类间距离	24
2.2.3 聚类的准则函数	25
2.2.4 最大最小聚类法	28
2.3 遥感图像多光谱特征空间距离分析方法	29
2.3.1 空间距离分布	29
2.3.2 类内、类间距分析	31
2.3.3 土地利用多光谱特征	32

2 / 多光谱遥感图像土地利用分类方法

2.3.4 土地利用类别类内、类间距分析	39
2.4 本章小结	39

第3章 模糊自适应神经网络(FasART)分类方法 41

3.1 改进的C-均值法和最大似然法	41
3.1.1 改进的C-均值法	41
3.1.2 最大似然法	44
3.2 模糊数学基础	45
3.2.1 模糊集合	45
3.2.2 模糊关系	47
3.2.3 模糊推理	49
3.2.2 模糊语言	51
3.3 ART模型及模糊ARTMAP神经网络	53
3.3.1 ART模型原理及模糊ART网络	54
3.3.2 模糊ARTMAP	56
3.4 FasART神经网络及其简化	58
3.4.1 新型的模糊神经网络模型FasART	58
3.4.2 FasART模型的简化及其算法	61
3.5 最大似然法和FasART网络的土地利用分类	62
3.5.1 最大似然法分类	62
3.5.2 FasART网络分类	63
3.6 本章小结	66

第4章 基于纹理特征的分类 67

4.1 纹理特征的提取	67
4.1.1 常见的纹理特征提取方法	67
4.1.2 模糊纹理分析	69
4.1.3 多光谱遥感图像的模糊纹理分析及算法的改进	70
4.1.4 纹理特征参数	73
4.2 多光谱遥感图像纹理特征的不规则性	74
4.2.1 同一类别不同大小的测量区域的模糊纹理光谱特性	74
4.2.2 不同类别不同测量区域的模糊纹理光谱特性	75
4.2.3 同一类别不同波段的模糊纹理光谱特性	76
4.3 基于纹理特征的分割	77
4.3.1 分开-合并与多分辨率图像锥图像分割法	77

4.3.2 分开 - 扩张方法	82
4.4 实验	86
4.4.1 多分辨率图像锥与分开 - 扩张分割	86
4.4.2 分开 - 合并与分开 - 扩张分割	87
4.5 本章小结	89
第 5 章 模糊积分融合纹理特征与光谱特征分类	91
5.1 数据融合的基本概念和融合方法	91
5.1.1 数据融合的基本概念	91
5.1.2 数据融合方法	93
5.2 基于 D - S 理论的决策层融合	96
5.2.1 D - S 理论简介	96
5.2.2 基于 D - S 理论的融合	97
5.3 基于模糊积分的决策层融合	98
5.3.1 模糊测度和模糊积分	98
5.3.2 模糊积分的融合模型	102
5.3.3 维数差异对融合影响分析	103
5.3.4 特征维数调整	104
5.3.5 实验	105
5.4 本章小结	108
第 6 章 基于区域多中心的多光谱遥感图像 土地利用分类	109
6.1 区域多中心分类方法	110
6.1.1 类别模式的表征	110
6.1.2 分类方法及初步分析	111
6.1.3 方法流程及步骤	112
6.2 类别模式的选择	113
6.2.1 训练样品的选取原则	113
6.2.2 模式中心数量的确定	114
6.3 区域像元类别百分比和模式中心数门限值的确定	115
6.3.1 不同训练样品的区域像元类别百分比和类别 模式中心数	116
6.3.2 区域单元大小不同的区域像元类别百分比和 类别模式中心数	117

4 / 多光谱遥感图像土地利用分类方法

6.4 实验	118
6.4.1 不同百分比和模式中心数的分类	118
6.4.2 不同训练样品及模式中心数的分类	119
6.4.3 单一类别的提取	121
6.5 本章小结	122

第7章 遥感图像分类方法在城市扩展变化检测中的应用 124

7.1 遥感图像城市扩展变化检测概述	124
7.1.1 变化检测的基本概念	124
7.1.2 常用的变化检测方法	125
7.1.3 城市扩展变化检测存在的问题	128
7.2 先比较后分类变化检测方法	129
7.2.1 分类后比较方法	129
7.2.2 先比较后分类方法	130
7.3 遥感图像变化检测在长株潭遥感监测系统中的应用	132
7.3.1 系统结构及主要功能	132
7.3.2 遥感图像分类后数据	137
7.3.3 长株潭城市扩展变化检测	140
7.4 基于遥感监测系统的长株潭城市扩展分析	143
7.4.1 城市空间扩展特征	143
7.4.2 空间扩展的因素	144
7.4.3 空间扩展的问题	145
7.4.4 空间扩展的趋势及发展战略	145
7.5 本章小结	146

附录A 用C++实现FasART网络	147
--------------------	-----

附录B 用C++实现分开-扩张算法	164
-------------------	-----

附录C 用C++实现区域多中心分类方法	181
---------------------	-----

参考文献	206
------	-----

第1章 绪论

随着对地观测技术的迅速发展,遥感技术已广泛渗透到国民经济的各个领域,特别在推动经济建设、环境改善、国防建设等方面起着越来越重要的作用。空间对地观测技术成为20世纪80年代世界高技术发展的重要成就之一。

遥感技术是一门跨学科的综合技术,它包括三个方面:遥感平台、传感器和遥感信息处理。遥感平台是运载工具,分航空和卫星平台。传感器是安装在遥感平台上的遥感测量仪器。遥感信息处理是遥感应用的一个关键环节。在数字化、可视化、智能化和网络化方面有了很大发展的今天,遥感信息的处理还不能充分满足广大用户的需求,日益丰富的遥感信息还没有被充分发掘和处理。据估计,空间遥感获取的图像数据,经处理的还不足5%。

遥感图像分类是模式识别技术在遥感领域中的具体应用,是遥感信息处理技术中的一项关键技术。随着数字信号处理技术、计算机技术以及通信技术的迅猛发展,遥感信息处理在军事、民用等领域发挥着越来越重要的作用。虽然模式分类技术的研究历史比较长,但随着遥感技术的发展,用户的要求也越来越高,越来越需要遥感分类技术能够提供更好的结果,同时也促进了模式分类技术的发展。

1.1 土地利用

IGBP和IHDP将土地覆盖定义为“地球陆地表层和近地面层的自然状态,是自然过程和人类活动共同作用的结果”,美国生态学会将土地覆盖定义为“土地表面的生态状态和自然表现”,土地利用指的是人类施加于地表的活动。全球不同区域土地覆盖的性质主要取决于自然因素,但目前的土地覆盖状况则主要是人类对土地的利用和整治活动造成的。

土地利用分类系统是土地利用变化研究的核心,也是研究成果的表达方式。土地利用类型是什么、是否变化,都与定义的分类系统有密切关系。例如,土地利用分类系统是三级系统,在10年尺度上,在第一级系统上土地利用没有发生变

化,始终是农田,但在较详细的第三级系统上却有重大变化,如水田变成旱地。另一种情况是,土地利用分类系统是一级系统,但系统中土地利用类型的数目对于土地利用及其变化的定义有很大影响,如土地利用类型数目越多,土地利用变化就越容易发生。土地利用及变化研究是相对于一定的土地利用分类系统而言的。

国内外现存分类体系主要有 6 种:美国 USGS(ANDERSON)分类法,欧盟 CORINE 分类法,FAO(联合国粮农组织)分类法,LUCC 土地分类法,中国科学院土地分类法以及国土资源部的土地分类法。

中国科学院进行资源调查所采用的土地利用分类体系采用三级分类体系:一级分为耕地、林地、草地、水体、城乡工矿和居民用地及未利用土地 6 类,主要根据土地的自然生态和利用属性;二级分为 25 个类型,主要根据土地经营特点、利用方式和覆盖特征;三级主要根据地形、坡度等土地的自然属性特征,将耕地进一步分为山区、丘陵、平原和坡度大于 25° 几种类型。

1.2 资源卫星及其多光谱遥感图像

资源卫星是用于陆地资源和环境探测的卫星,又称之为陆地卫星。依据不同的指标和方法,陆地卫星有多种分类方法,按综合分类为陆地卫星,高分辨率陆地卫星,高光谱卫星和合成孔径雷达等 4 类。主要的陆地卫星系列有 Landsat(美)、SPOT(法)、CBERS(中)、IRS(印)、ALOS(日)、IKONS(美)等,卫星的主要参数见表 1-1。

本书主要以地面分辨率为 10~30 m 的 CBERS-1、LANDSAT-7 资源卫星多光谱遥感图像为研究的对象。这类图像的特点如下:

- (1) 空间分辨率 10~30 m,有较好的地物细节的空间特征表达;
- (2) 时间分辨率小于 30 天,可用于灾害、环境变化等方面动态监测;
- (3) 具有 4 个以上的光谱波段;
- (4) 可以应用在农业、林业、土地、城市、环境、灾害、地质、海洋、测绘等领域。

各种土地类型在中巴资源卫星的(4、3、2 波段)图像的颜色特征见表 1-2 和表 1-3。

表 1-1 资源卫星主要参数表

项目	中国	美国	法国	印度
	CBERS-1	LANDSAT-7	SPOT-4	IRS-1D
轨道	高度: 778 km 倾角: 98.5° 降交点地方时: 10:30 运行周期: 100.26 min/圈 14.3 圈/天 轨道重复周期: 26 d WFI 重复周期: 4~5 d	高度: 705 km 倾角: 98.9° 降交点地方时: 10:00 ± 15 运行周期: 98.9 min/圈 轨道重复周期: 16 d	高度: 832 km 倾角: 98.7° 降交点地方时: 10:30 ± 15 运行周期: 101.4 min/圈 14.2 圈/天 轨道重复周期: 26 d VI 重复周期: 1~2 d	高度: 817 km 倾角: 98.69° 降交点地方时: 10:30 运行周期: 100.35 min/圈 14 圈/天 LISS-3 轨道重复周期: 26 天 WIFS, PAN 轨道重复周期: 5 天
主要有效载荷配置	CCD 多光谱相机 IRMSS 红外光谱扫描仪 巴西 WFI 宽视场 CCD 相机	7 号改进增强型 ETM +	2 台高分辨率可见光—近红外相机 HRVIR, 1 台植被仪 VI	LISS-3 四谱段 CCD 相机; WIFS 广角遥感器 PAN 全色相机(具有 ± 26° 测试能力, 5 天重复周期)
扫描宽度/km	CCD: 113 IRMSS: 119.5 WFI: 890	ETM + : 183	HRVIR: 2 × 60 VI: 幅宽 2250	LISS-3: 142/148 WIFS: 774 PAN: 70
有效扫描视场/(°)	CCD: 8.32° (具有 ± 32 侧视能力) IRMSS: 8.8° WFI: 60°	ETM + : 15°	具有 ± 27° 侧视能力	LISS-3: ± 5° WIFS: 54° PAN: ± 2.5° 可在 ± 27° 侧摆

续表 1-1

项目	中国	美国	法国	印度
	CBERS-1	LANDSAT-7	SPOT-4	IRS-1D
光谱范围 /μm	CCD: B1 0.45~0.52 20 B2 0.52~0.59 20 B3 0.63~0.69 20 B4 0.77~0.89 20 B5 0.51~0.73 20	ETM+: 0.45~0.52 30 0.52~0.60 30 0.63~0.69 30 0.76~0.90 30 1.55~1.75 30 2.08~2.35 30 10.4~12.5 60 另新增—15m分辨率全色 波段 0.50~0.90μm	HRVIR: 0.50~0.59 20 0.61~0.68 10 0.78~0.89 20 1.50~1.70 20 VI: 0.43~0.47 1000 0.61~0.68 1000 0.78~0.80 1000 1.58~1.75 1000	LISS-3: 0.52~0.59 23.5 0.62~0.68 23.5 0.77~0.86 23.5 1.55~1.75 70.5 WIFS: 0.62~0.68 188 0.77~0.86 188 PAN: 0.50~0.75 5.8
分辨率 /m	IR-MSS: B6 0.5~0.9 80 B7 1.55~1.75 80 B8 2.08~2.35 80 B9 10.4~12.5 160 WFI: B10 0.63~0.69 256 B11 0.77~0.89 256			
发射日期	1999年10月14日	1999年4月15日	1998年3月24日	1997年

表 1-2 城区建筑和土地类型 CCD 图像颜色特征

类别	亚类	颜色特征
建筑类型	高层建筑	阴影呈黑色斑点及黑色斑块
	中层建筑	阴影呈深蓝色斑点及深蓝色斑块
	低层建筑	房顶呈蓝色斑点及蓝色斑块
	简易建筑	房顶呈翠绿色斑点及翠绿色斑块
土地类型	无植物裸露地	白色斑点及白色斑块
	布满绿树土地	鲜红色斑点及鲜红色斑块
	未布满绿树土地	白色斑点与鲜红色斑点组成的杂斑
	枯草覆盖土地	淡黄绿色斑点及淡黄绿色斑块

表 1-3 城郊农业用地 CCD 图像颜色特征

类别	亚类	遥感解译标志(综合)
房屋	矮楼	蓝色、浅蓝色斑点及斑块(矮楼群)
	平房	淡蓝色斑点及斑块(平房群、村落)
农田	菜地	粉红色斑块
	稻田、绿草田	浅粉红色、粉红色斑块
	枯草地	淡黄绿色斑块
池塘	水塘	深蓝色斑点、斑条(块)
	养鱼池	深蓝色的四边形,池边及池间的浅蓝白色线为池堤,白色的四边形为无水干池

1.3 模式识别技术概述

模式识别是一门以应用为基础的学科,目的是将对象进行分类。这些对象与应用领域有关,他们可以是图像、信号波形或者任何可测量的且需要分类的对象。模式识别具有悠久的历史,但在 20 世纪 60 年代以前,模式识别主要是统计学领域中的理论研究。同其他事物一样,计算机的出现提高了对模式识别实际应用的需求,而反过来又对理论发展提出了更高的要求。就像我们的社会从工业化到后工业化阶段一样,工业生产中的自动化以及信息处理和检索的需求变得日益重要,这种趋势把模式识别推向今天的工程应用和研究的高级阶段。在大多数机

器智能系统中,模式识别是用于决策的主要部分。

1.3.1 基本模型

鉴于模式识别的许多技术涵盖了多个学科的发展,自然会出现不同学科对相同术语不同的甚至相反界定的情形。在此,我们将“模式(样本)”表示为 p 维数据向量 $\mathbf{x} = (x_1, \dots, x_p)^T$, 其中 $\mathbf{x}$ 表示被分类的对象, p 表示用于分类的特征变量的数量, T 表示向量的转置, x_i 表示第 i 个特征的观测值。若识别问题含有 C 个类, 记为 $\omega_1, \dots, \omega_c$, 则关于每个模式 $\mathbf{x}$ 的类别变量记为 z , z 表示 $\mathbf{x}$ 的类别, 即若 $z = i$, 则模式 $\mathbf{x}$ 属于 ω_i , $i \in \{1, \dots, C\}$ 。

图 1-1 为模式分类器的示意图。一旦分类器的参数确定下来, 分类器便能对给定的样本产生某种意义上的最佳响应, 该响应通常是对样本所属类别的估计。一组类别已知的样本 $\{(x_i, z_i), i = 1, \dots, n\}$ 可以作为分类器的训练集或设计集, 分类器的设计就是用训练集确定分类器的内部参数, 由此形成的分类器可用于估计未知样本 $\mathbf{x}$ 的属性。

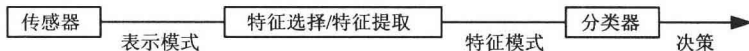


图 1-1 模式分类器

分类器的形式取决于多种因素。其一是训练数据的分布以及对数据分布所作的假设；其二是错分代价——做出错误决策的代价。在很多实际应用中，错分代价包括金钱、时间和其他主观性判断等多个方面，是很难量化的。

图 1-1 大致勾画了模式分类的过程。期间, 需要经过几个独立的数据变换阶段。这些变换有时也叫做预处理、特征选择或特征提取, 变换的结果通常导致模式维数的降低(减少特征数)、多余的和无关信息的删除, 并将数据模式转换成更适合于后续分类工作所需的形式。

1.3.2 特征归一化、提取与选择

特征归一化、提取与选择均是对表述特征的处理方法。

在很多情况下, 特征值具有不同的动态范围。在代价函数中, 大值特征比小值特征的影响更大, 但在分类器设计中, 这并不能反映它们所具有的重要程度。为了使这些特征值位于相似范围, 消除数值大小的影响, 特征归一化一般采用如下方法: 一是用各自的均值和方差的估计值做归一化; 二是通过合适的比值将特征值限制在 $[0, 1]$ 或 $[-1, 1]$ 范围内; 三是当数据不是均匀分布在均值周围时, 使用非线性方法, 如对数或 Sigmoid 函数。

通常的归一化的方法只是从数值的大小上对特征进行处理,并没有涉及属性的差异,即属性对分类的重要程度。

为了提高分类性能,使来自同一类别的不同样本的特征值非常接近,而来自不同类别的样本的特征值有很大的差异,删除表述特征多余的或不相关的信息,从 p 维表述特征中选择(或提取) d 个特征,这个过程称为特征选择或特征提取。对于特征选择来说,是在从 p 个可能的测量值 $x_1, \dots, x_p$ 的所有可能的大小为 d 的子集 I_d 的集合上进行的,寻求子集 $\tilde{X}_d$, 使:

$$J(\tilde{X}_d) = \max_{X \in I_d} J(X)$$

J 为最优化准则函数。准则函数通常是基于距离测度或分布差异测度的。

特征提取是通过变换将原始数据(对所有变量)变换成变量个数减少的数据集。如果 $\mathbf{x}$ 是测量样本的 m 维向量,将其变换成另一个 l 维向量 $\mathbf{y}$,所使用的类可分测量准则最优。定义线性变换为

$$\mathbf{y} = A^T \mathbf{x}$$

其中: A^T 是一个 $l \times m$ 矩阵。优化过程的复杂度很大程度上依赖所选择的准则。

凡直接用于分类的特征称之为分类特征。一般情况下表述特征需经归一化、提取与选择等处理才用于分类,因此,经选择所得到的特征子集 $\tilde{X}_d$ 或经变换所形成的数据集均称之为分类特征。

对于存在多种属性参数的表述特征,在进行特征提取时,通常对整个表述特征各参数进行统一的数学变换,并未考虑源自不同属性的特征子集的差异,往往会造成信息量的损失;而经特征选择所形成的分类特征仍然存在不同属性的特征子集的问题,这些都会影响分类的精度。

1.3.3 分类器选择的原则

模式识别系统中,分类器的作用是根据分类特征给每个被测对象赋予一个类别标记,因为完美的分类性能通常是不可能获得的,更一般的任务是确定每一个可能类别的概率,类别概率的高低反映分类器的置信度。

模式识别的许多领域都会面临分类器的选择问题,即应该允许分类器具有怎样的复杂程度。这里,复杂度是根据待估计的自由参数的个数来度量的。分类器的最优复杂度取决于训练数据的质和量。一个过于复杂的分类器能够很好地模拟训练数据(甚至也包含训练数据的噪声),但对于与训练数据同分布的未知数据来说,其一般性能很可能会变差(分类器会对数据“过度拟合”)。如果分类器不够复杂,则不能充分模拟数据的结构。因此,分类器选择本身就是确定最优模型参数过程的一部分。在这种情况下,分类器的复杂度是一个待定参数。因此,许