

研究生教学用书

教育部学位管理与研究生教育司推荐

现代模式识别

(第二版)

Modern Pattern Recognition

(Second Edition)

孙即祥 编著



高等教育出版社

研究生教学用书

教育部学位管理与研究生教育司推荐

现代模式识别

(第二版)

Modern Pattern Recognition

(Second Edition)

孙即祥 编著



高等教育出版社

内容提要

本书系统地论述了各类经典的模式识别的理论与方法,同时还较全面地反映了本学科的新近科技成果。本书讨论的主流模式识别技术有:统计模式识别、模糊模式识别、神经网络技术、人工智能方法、子空间模式识别及结构模式识别等。

全书共 17 章。第一章为引论;第二章至第七章介绍统计模式识别,包括:聚类分析、判别域代数界面方程法、统计判决、统计决策中的学习与错误率估计、最近邻法和特征提取与选择;第八章为模糊模式识别方法;第九章介绍神经网络技术;第十章信息融合主要论述识别与决策中的有关融合技术;第十一章为结构模式识别;第十二章智能化方法侧重讨论不确定推理;第十三章阐述决策树;第十四章论述支持矢量机;第十五章讨论隐马尔可夫模型识别方法;第十六章为子空间模式识别方法;第十七章介绍最小风险设计。

本书可供电子科学与技术、信息与通信工程、控制科学与工程、计算机科学与技术及其他领域的有关专业和研究方向的研究生、本科高年级学生作为关于信息分析、检测、识别的教材或教学参考书,也可以供相关专业的科研人员参考。

图书在版编目(CIP)数据

现代模式识别 / 孙即祥编著. —2 版. —北京:高等教育出版社, 2008. 10

ISBN 978-7-04-020587-9

I. 现… II. 孙… III. 模式识别 - 高等学校 - 教材
IV. TP391.4

中国版本图书馆 CIP 数据核字(2007)第 068882 号

策划编辑	刘 英	责任编辑	刘 英	封面设计	李卫青	责任绘图	朱 静
版式设计	史新薇	责任校对	刘 莉	责任印制	韩 刚		

出版发行 高等教育出版社
社 址 北京市西城区德外大街 4 号
邮政编码 100120
总 机 010-58581000

经 销 蓝色畅想图书发行有限公司
印 刷 北京中科印刷有限公司

购书热线 010-58581118
免费咨询 800-810-0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.landaco.com>
<http://www.landaco.com.cn>
畅想教育 <http://www.widedu.com>

开 本 787×1092 1/16
印 张 45.5
字 数 1 100 000

版 次 2002 年 1 月第 1 版
2008 年 10 月第 2 版
印 次 2008 年 10 月第 1 次印刷
定 价 58.00 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

物料号 20587-00

前言

模式识别是研究分类识别理论和方法的科学技术,是一门综合性、交叉性学科。在理论上它涉及代数学、矩阵论、函数论、随机数学、模糊数学、图论、最优化理论、信号处理、计算机科学、神经物理学等众多学科的知识;在应用上它又与其他许多领域的知识及工程技术密切相关;其内涵可以概括为信息处理、分析与决策,它既是人工智能研究领域的重要分支,又是实现机器智能必不可少的技术手段。该学科的理论任务是运用相关科技研发分类识别的理论和方法,而其应用目标是创造能进行分类识别决策的智能机器系统以代替人类的分类识别工作。自20世纪70年代以来,该学科受到了学术界和各应用领域的极大重视,与模式识别相关的理论专著、论文、科研成果层出不穷,使得该学科得以丰富和发展,形成了许多大类的模式识别理论、方法。但是现在多数论著只涉及一至两类模式识别知识的介绍,多学科、多视角、多层次地介绍该学科知识的著作不多,能兼顾教学使用和科研参考的高校教材也较少,因此有必要将该学科涉及的基本理论、基本方法以及当代发展成熟的理论技术进行沉淀、提炼、归纳、整合,让读者能较系统地学习本学科的理论精髓,较全面地了解和掌握相关技术,是我们撰写本书的初衷和希望。

本书是一本关于模式识别理论和方法的著作,是在原研究生课程讲义基础上结合该课程多年教学实践经验及相关的科研成果,参考了大量的专家著作及科技文献,依据教学、科研需要和学科发展趋势撰写而成,它是已出版的《数字图象处理》、《图象处理》、《图象分析》、《图象压缩与投影重建》及《模式识别中的特征提取与计算机视觉不变量》的姊妹篇。在本书的撰写过程中遵循以下三个原则:(1)在结构安排上尽量使知识表达体系与学科本身的体系相一致;(2)在内容阐述方式上遵循人的认知规律;(3)在选材上尽量使读者掌握经典和现代的重要学科知识,使读者学后提高解决实际问题的能力,融入学科发展潮流。我们的目标是使本书可读性好、学术性强、实用价值大。在内容深度与表述形式上,定位于教材与专著之间,兼顾理科与工科使用。

本书的第一版面世后,受到读者的好评,被许多高校和科研单位采用,多次重印、多次获奖,并于2005年被教育部有关部门评审为研究生教学推荐用书。在此基础上,依据教学、科研需要和学科发展,我们又增写了五章,这五章较深入地论述了一些模式识别的专题,并对第一版的内容进行了修订更新,增加了一些新的成果。

较详细地了解本书的特点,对于读好这本书,学习这门课程,掌握这门知识是有益的。具体地讲,本书具有下述的特点:

1. 内容广泛新颖。模式识别是一门相当活跃的重要学科,其发展非常迅速,它所涉及的理论十分广泛,方法十分丰富,新理论、新方法、新技术、新应用不断涌现。本书较全面系统地阐述了模式识别几乎所有主流领域的知识,除了传统、经典的重要内容外,还收入了经实践证明有重要现实意义和广阔应用前景的新理论、新方法和新技术。本书包括统计模式识别、模糊模式识别、神经网络技术、人工智能方法、子空间方法及句法模式识别。本书涵盖面广、内容多,有益于不同专业的教师、学生在教学时有更多的选择空间,可以满足教师课堂讲授信息量大的教学设想和学生课下更多地学习要求。本书某些内容超出了当前教学大纲,内容在质和量上都具有“超前

性、超前量”,这样既能开拓读者的视野和知识面,提高读者深入学习的兴趣,还能适应教学与时俱进、不断更新发展的需要,同时也为科技工作者提供解决实际问题的思路和方法。

2. 结构清晰合理。合理清晰的学科知识表述体系有益于读者对各种理论、方法的理解和记忆。本书内容在组织上呈现层次化、模块化,全书及各章节都尽量遵循由浅入深、先易后难、先具体后抽象的原则来安排。例如,将较简单直观的聚类分析、判别域代数界面法放在全书的前面,其后才是较抽象的统计识别理论与方法;模糊识别方法放在统计识别方法之后,因为模糊识别方法中的一些识别策略及具体算法与统计方法中的某些内容在形式上是类似的,这样有利于读者既能区分两者概念原理的不同,又能掌握两者方法形式上的联系。另外,本书主要是为研究生教学使用撰写的,但也兼顾到了其他层面的读者。本书的前面几章及其他一些章的前面部分可用于本科生高年级有关专业教学使用,前面几章的后面部分及中间一些章主要是面向硕士研究生,后面几章主要是为博士研究生撰写的。同时全书也适用于不同层面的从事相关科研的技术人员参考。

3. 选材考究精细。如前所述,这门学科的各种理论、方法、技术纷繁众多,而且还在不断地出现和发展。编著者根据多年科研与教学工作实践的体会和认识,对模式识别的各种理论、方法、技术及应用等成果,进行梳理、沉淀、提炼、归纳和总结,以是否具有当前或潜在的理论意义或应用价值为标准,在众多的知识中选取那些或具基础理论性,或具思维训练性,或具有实用性,或具思想启发性,或具方向前瞻性的有代表性的重要内容,并且处理好本书的内容与其他学科知识的关系。

4. 注重基础。打好基础是教育经验的总结,也是科技高速发展的需要,本书自始至终都非常注重强化基本概念、基本思想、基础理论、基本方法和基本技能。本书基于“新”与“恒”的对立与统一的观点和方法进行取材和论述,两者统一在知识的实用价值上。只有掌握了基本而有用的知识,读者才能在日后的学习和科研中有“后劲”,持续前进,才能以“不变”应“万变”,才能在浩瀚的知识海洋中畅游,始终处于科技潮流的最前沿。

5. 注重讲“理”,突出学术思想。为使读者真正掌握学科知识和实实在在提高解决问题的能力,本书在阐明知识时,不仅讲其然,还要讲其所以然,不是一些定理、结论、算法步骤的简单罗列。注意理论严密性与表述浅显性的统一,对一些通常认为较“简单”的内容,尽量挖掘其理论依据,使之有理论深度;而对于涉及较深奥理论的部分,在严格、严密的前提下,尽量用平实直白的语言进行论述,以避免不必要的符号猜解和复杂推导,使读者淡化更为重要的基本概念、学术思想和技术思路。力求本书各部分在深线上达到基本的统一。

6. 详略得当。由于本书涉及的知识面广,对于基本的、重要的、关键性的理论和方法要论述清楚,说深说透;而对于相关的、类似的或不“稳定”的方法,或知识层次限制的内容,则适当地论述或点到为止,以给出联系,给出启示,给出方向。这样既实现了知识在面上的广度,又达到知识点处的理论深度,读者通过知识的“内插”和“外推”,就可以形成巨大的“三维”知识体。为了益于学习和研究,有些内容详细阐述,当一些内容以应用为目的时,有关知识和结论则可能采用简单列写的形式。

以上努力的目的在于让读者对知识能够深刻理解、融会贯通、牢固记忆、全面掌握,花费较少的代价而学到较多有价值、有意义的知识;在学到知识的同时,提高自学能力、活用能力和创新能力,有益于读者综合素质的提高。

基于强化基础的思想,为了加深对正文的理解,在前面一些章节中给出了一些例题和图示;

随着学习的深入、水平的提高,更由于篇幅的限制,后面章节侧重于尽量在正文中详细、清楚地论述知识。

为了提高能力,在习题编选上基于如下的选题标准:(1)加深知识的理解,提高活用能力;(2)证实正文中的观点或结论;(3)进一步扩展知识;(4)引导研究方法和提高研究能力。

由于篇幅限制和计算机网络的普及,有关的重要文献没有一一列出,只是选写;在参考文献列写次序上,本着先“学习”,后“研究”的顺序。由于同样的原因,有关的重要应用建议在网络上查阅。

与模式识别直接密切相关或者说作为其先导课程的是:代数学、矩阵论、函数论、概率论与数理统计、最优化理论与方法、模糊数学、信号处理、人工智能原理、计算机软件设计等。对于某些读者,不可能也不必要花费大量的精力和时间去学完上述每门课程,由于前述的本书结构、选材及论述的特点,读者只要具备一些必要的理论基础和相关基本知识便可以顺利地学完每一章的主要内容。有些章前部给出了有关的数学基础和数学理论,若跳过去也不影响后面的学习。对于希望深入学习、进行学术理论研究或从事技术应用研究的读者,应先行修完上述课程,并在学习本课程时参阅其他相关的科技资料。

本书的内容除了包含我们的某些科研成果之外,还取材于国内外诸多学者、专家的研究成果、论文、著作,在此对他们表示由衷的感谢。感谢原国防科技大学校长郭桂蓉院士及院、系领导对本书的完稿给予的极大的支持。“现代模式识别”课程获得了湖南省研究生精品课程建设的资助,在这里对湖南省教育主管部门的支持表示感谢。感谢学校将“现代模式识别”课程作为“十一五”研究生重点建设课程的资助。在此也衷心感谢编辑的辛勤工作。

多位人员参加了本书编写的各项工作,感谢博士、硕士研究生季虎,姚伟,滕书华,谭志国,杜春,邵晓芳,毛玲,王学梅,王春光、赵健、陈明生,王亮亮,阮葵,吴琼等同学,他们向计算机录入了书稿并参与了书稿校对。另外,滕书华参与了全书习题的选编、录入,参考文献的选录以及插图的制作;杜春对一些公式进行了推导验证、参考文献的审核并极其细致地校对了全书校样;姚伟对全书的电子版进行了规格化工作和制作了插图。刘纲钦、邹刚等副教授也给予了支持和帮助。本书的第一章至第十七章由孙即祥撰写,其中,第十章由孙即祥和王宏强合写,刘雨、张帆参与了第十一章撰写,全书由孙即祥统稿。

多位人员参加了本书的编写工作,这里要感谢博士、硕士研究生季虎,谭志国,姚伟,滕书华,邵晓芳,毛玲,阮葵,吴琼,王亮亮,杜春等同学,他们录入了大量的扩充后的书稿内容以及制作了许多插图。本书的第一章至第十七章由孙即祥撰写,其中,第十章由孙即祥和王宏强合写;刘雨、张帆参与了第十一章部分内容的撰写;张帆参与了第十二章部分内容的撰写;滕书华参与了全书习题的选编、录入工作,杜春对一些公式进行推导并对部分错误作了更正。全书由孙即祥统稿,姚伟、杜春对全书的电子版进行了规格化工作并制作了许多插图。

由于模式识别是一门不断发展的学科,新的理论、方法和技术,新的应用成果不断涌现,再加上我们的学识水平及时间有限,可能没有完全达到我们所希望的目标,也不可避免地存在各种错误和疏漏,敬请读者给予批评指正。

孙即祥

2007年11月于国防科技大学

第一版序

随着社会发展的需要和科技的不断进步,可以断言,未来的人造系统或产品的显著特点之一是智能化,模式识别正是研究人类识别机理和实现识别功能的科学,它是由数学、控制理论、信息处理、计算机技术、生物生理学和心理学等等多门理学和工学学科相互交叉融合而产生的。现在已形成了内涵丰富的理论体系,产生了一系列有效的技术手段,并已广泛地应用于诸多领域,产生了巨大的效益。它是一门发展非常活跃的学科,新的理论、技术、方法和新应用层出不穷。自它诞生之日起,就引起了国内外各领域专家、学者的高度重视,吸引了众多科技人员投入极大的研究和应用热情。

孙即祥教授根据多年的教学经验和科研成果,撰写了这本既面向高校教学使用又适用于科研参考的优秀著作。该书蕴涵的充实、严谨的基础性理论,有助于学生打下坚实的理论功底,同时书中介绍的一般性和较新的模式识别技术又为广大读者提供了广泛的解决实际问题的思路和方法。相信此书必将为模式识别的教学和相关科研工作的发展起到一定的促进和推动作用。

孙即祥 院士

2002年1月

目 录

第一章 绪论	1	上机练习	129
1.1 概述	1	参考文献	130
1.2 特征矢量和特征空间	6	第四章 统计判决	132
1.3 随机矢量的描述	6	4.1 最小误判概率准则判决	132
1.4 正态分布	9	4.2 最小损失准则判决	148
参考文献	15	4.3 最小最大损失准则	156
第二章 聚类分析	16	4.4 $N-P$ (Neyman-Pearson)判决	159
2.1 聚类分析的概念	16	4.5 序贯判决(SPRD)	162
2.2 模式相似性测度	18	4.6 Fisher 准则判决	167
2.3 类的定义与类间距离	25	4.7 特征数据缺损或被噪声 污染下的 Bayes 判决	168
2.4 准则函数	29	4.8 批对象的复合判决	170
2.5 聚类的算法	37	文献简评 应用简介	171
文献简评 应用简介	67	习题	171
习题	68	上机练习	175
上机练习	70	参考文献	176
参考文献	71	第五章 统计决策中的学习与错误率 估计	178
第三章 判别域代数界面方程法	73	5.1 统计推断概述	178
3.1 用判别域界面方程分类的概念	73	5.2 参数估计	180
3.2 线性判别函数	73	5.3 Bayes 学习	186
3.3 判别函数值的鉴别意义、权空间 及解空间	78	5.4 概密的窗函数估计法	189
3.4 Fisher 线性判别	80	5.5 有限项正交函数级数逼近法	198
3.5 线性可分条件下判别函数的权矢量 算法	85	5.6 用位势函数法逼近 Bayes 判决 函数	202
3.6 一般情况下的判别函数权矢量 算法	92	5.7 随机逼近方法求类的后验概率	205
3.7 线性规划方法	99	5.8 统计决策准则下线性判决函数的 训练生成	210
3.8 线性二分能力	101	5.9 错误率估计	215
3.9 广义线性判别函数	104	5.10 基于平均损失估计的学习及最小误 判概率的估计	224
3.10 二次判别函数	106	5.11 无监督估计(盲估计)	226
3.11 分段线性判别函数	108	5.12 期望最大化算法	232
3.12 位势函数分类法	116	5.13 集成学习	238
3.13 支持矢量机简介	120	文献简评 应用简介	241
3.14 最小最大概率机	122	习题	243
文献简评 应用简介	128		
习题	128		

上机练习	246	应用	385
参考文献	248	9.1 神经网络的基本知识	385
第六章 最近邻法	252	9.2 前向型人工神经网络	389
6.1 基本的最近邻法	252	9.3 BP 网的性能和学习改进	400
6.2 剪辑最近邻法	259	9.4 Hopfield 网络	415
6.3 引入拒绝决策的最近邻法	263	9.5 随机神经网络	424
6.4 最近邻法中的最佳距离及其 实际计算	265	9.6 自适应共振理论神经网络	435
文献简评 应用简介	269	9.7 自组织特征映射神经网络	438
习题	269	9.8 模糊神经网络	442
参考文献	270	9.9 概率神经网络	446
第七章 特征提取与选择	272	9.10 RCE 神经网络	447
7.1 概述	272	文献简评 应用简介	448
7.2 类别可分性判据	273	习题	449
7.3 基于可分性判据进行变换的 特征提取与选择	285	上机练习	450
7.4 最佳鉴别矢量的提取	296	参考文献	451
7.5 离散 K-L 变换及其在特征提取 与选择中的应用	299	第十章 信息融合	454
7.6 独立成分分析	310	10.1 概述	454
7.7 基于决策界的特征提取	316	10.2 融合技术层次性及融合系统 功能模块和结构	455
7.8 特征选择中的直接挑选法	323	10.3 关于信息融合的熵理论	462
7.9 多维尺度分析	329	10.4 观测不相关的分布式最小损失 准则下的检测与决策融合	467
文献简评 应用简介	331	10.5 观测相关的决策融合	475
习题	333	10.6 N-P 准则下的决策融合	479
参考文献	334	10.7 分布式检测决策融合全局最优概述 及某些约束条件下最优解	481
第八章 模糊模式识别	338	10.8 D-S 证据理论的融合算法	485
8.1 引言	338	文献简评 应用简介	491
8.2 普通集合与模糊集合	338	习题	492
8.3 普通集合上的关系及有关知识	351	参考文献	492
8.4 模糊关系与模糊变换	355	第十一章 结构模式识别	495
8.5 模糊度和特征提取与选择	359	11.1 结构模式识别概述	495
8.6 模糊识别的基本方法	362	11.2 形式语言	496
8.7 基于模糊相似矩阵的分类方法	366	11.3 高维文法与随机文法	500
8.8 模糊 C-均值聚类算法	367	11.4 模式的描述	507
8.9 最大树法模式识别	378	11.5 句法分析	512
8.10 几何图形的模糊识别	380	11.6 文法推断	532
文献简评 应用简介	382	文献简评 应用简介	548
习题	382	习题	548
参考文献	383	参考文献	554
第九章 神经网络在模式识别中的		第十二章 智能化方法	555

12.1 人工智能	555	15.2 一阶隐马尔可夫模型(HMM)	653
12.2 专家系统	555	15.3 可见序列概率估计	655
12.3 知识的表示	558	15.4 隐状态估计	658
12.4 智能推理技术	563	15.5 模型参数估计	659
12.5 不确定性推理	572	15.6 隐马尔可夫模型方法模式识别	661
文献简评 应用简介	599	文献简评 应用简介	662
习题	599	习题	662
参考文献	602	参考文献	664
第十三章 树分类器	604	第十六章 子空间模式识别方法	666
13.1 树分类器原理	604	16.1 概述	666
13.2 树分类器的设计原则	607	16.2 子空间 投影	667
13.3 树分类器的关键技术	608	16.3 子空间判别法	673
13.4 决策树生成算法	614	16.4 线性回归模型法	676
文献简评 应用简介	618	16.5 正交子空间法	676
习题	619	16.6 Kohonen 学习子空间法	679
上机练习	621	16.7 子空间的平均学习法	684
参考文献	622	文献简评 应用简介	686
第十四章 支持矢量机	624	习题	687
14.1 最优化的分析方法原理	624	参考文献	687
14.2 最优分类界面	629	第十七章 机器统计学习理论	689
14.3 广义最优分类界面	633	17.1 机器统计学习理论概述	689
14.4 最优界面与广义最优界面分类 性能的统计特性	636	17.2 经验风险最小化设计	690
14.5 支持矢量机(SVM)	637	17.3 经验风险最小化原则的一致 性条件	692
14.6 基于 Adaboost 的 SVM 组合	646	17.4 最优指示函数判决风险的界	699
文献简评 应用简介	650	17.5 训练序列的长度和识别率估计 精度的关系	704
习题	650	17.6 结构风险最小化	707
参考文献	650	文献简评 应用简介	711
第十五章 基于隐马尔可夫模型 识别方法	653	习题	711
15.1 一阶马尔可夫模型(MM)	653	参考文献	712

第一章 绪 论

1.1 概 述

1.1.1 模式识别

人们在日常生活、社会活动、科研生产以及学习工作中无时无刻不在进行着分类识别,分类识别是人类的基本活动之一。例如,儿童在认读识字卡片上的数字时,区分它们中的0~9,这是对数符号的识别;在读书看报时,人们进行的是文字识别活动;做某种实验时,对示波器显示波形的观察是一个波形识别过程;对科学研究、工程实验、场景监测等所摄取的图像进行内容分析是较复杂的图像识别工作;医生给患者诊断疾病需要对病情进行辨识;在人群中寻找某人是对人的形体及其他特征的识别行为。随着人类社会活动及生产、科研广泛而深入的发展,需要识别的对象种类越来越多,内容越来越深入复杂,要求也越来越高。为了改善工作条件、减轻工作强度,人们希望机器能代替人类完成某些繁重的识别工作;在某些环境恶劣、存在危险或人们根本不能接近的场合,就需要借助机器、运用分析算法进行识别工作。人们利用机器可以提高识别的速度、正确率及扩大应用的广度。这里所说的模式识别是指运用机器进行分类识别。以一个实例来说明机器的识别过程。模式识别的重要应用之一是计算机自动诊断疾病,其与医生的诊断过程相仿。首先要获得患者的有关情况,如测量体温、血压、心率,还可能对血液等进行化验,做X光透视、B超、心电图,甚至CT等检查。医生根据这些检查结果以及患者的病史等资料,运用自己的临床经验对患者进行诊断。而机器识别是将上述各种有用的资料信息输入计算机中,在这之前计算机已装入有关的分析算法,这些算法是专家知识、经验的总结和集成,其形式可以是规则、函数、数表等;通过计算机程序运用它们,对信息进行分析并作出判断。

下面首先较详细地介绍模式和模式识别这两个基本概念。为了能让机器执行和完成分类识别任务,必须首先将分类识别对象的有用信息输入计算机中,为此,应对分类识别对象进行科学的抽象,建立它的数学模型,用以描述和代替识别对象,称这种对象的描述为模式——pattern。pattern的原义是模范、模型、典型、样品、图案等,其内涵深刻、外延广泛。无论是自然界中物理、化学或生物等领域的对象,还是社会中的语言、文字等,都可以进行科学的抽象。具体地讲,我们对它们进行量测,得到表征它们特征的一组数据,为运用方便,将它们表示成矢量形式,称其为特征矢量;也可以将对象的基本特征元素作为基元用符号表示,从而将它们的结构特征描述成一个符号串、图或某个关系式。通俗地讲,模式就是事物的代表,是事物的数学模型之一,它的表示形式是矢量、符号串、图或关系式。对一类对象的抽象也称为该类的模式。

所谓模式识别(pattern recognition)是根据研究对象的特征或属性,利用以计算机为中心的机器系统,运用一定的分析算法认定它的类别,系统应使分类识别的结果尽可能地符合真实。

目前,模式识别理论和技术已成功地应用于工业、农业、金融、军事、公安、科研、生物医学、气

象、天文学等许多领域。模式识别成功应用的范例如我们熟知的信件分拣、遥感图片的机器判读、系统的故障诊断、生物特征识别(指纹、虹膜、脸等识别)、生物医学的细胞或组织分析、具有视觉的机器人、武器制导寻的系统、汽车自动驾驶系统以及文字与语言的识别等,并且现在正扩展到许多其他领域。当今时代科技发展的重要趋势之一是智能化,模式识别是人工智能的一个重要分支,模式识别还是知识发现、机器学习的主要技术之一。尽管现在机器识别的水平还远不如人脑,但随着模式识别理论以及其他相关学科的发展,可以预言,它的功能将会越来越强大,应用也会越来越广泛。

1.1.2 模式识别系统

前面介绍了模式识别的重要性以及两个首先接触的基本概念。一个功能较完善的识别系统在进行模式识别之前,首先需要进行学习。一个模式识别系统及识别过程的原理框图可以用图 1.1.1 表示,虚线的上部是分类、识别过程,虚线的下部是学习、训练过程。需要指出的是,应用的目的不同、采用的分类识别方法不同,具体的分类识别系统和过程也将会有所不同。一般而言,特征提取与选择、训练学习、分类识别是任何模式识别方法或系统的三大核心问题。

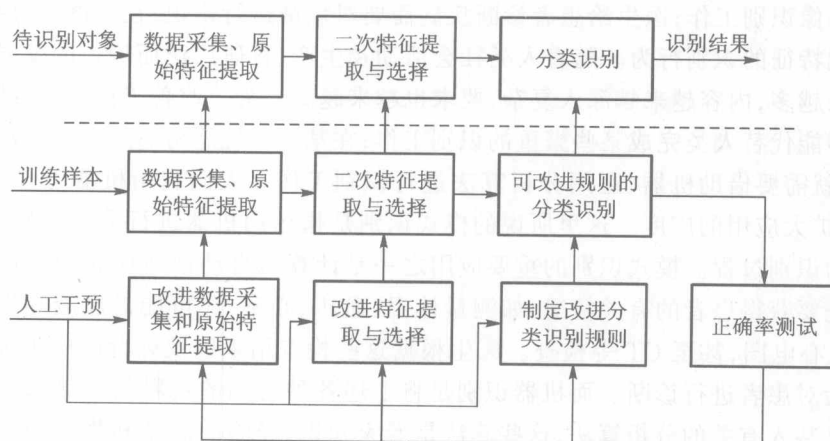


图 1.1.1 模式识别系统及识别过程的原理图

模式识别过程从信息层次、形态转换上讲,是由分析对象的物理空间通过特征提取转换为模式的特征空间,然后通过分类识别转换为输出的类别空间。

下面对识别系统的主要环节作简要的说明。

1. 特征提取

无论是识别过程还是学习过程,都要对分析对象固有的、本质的及重要的特征或属性进行量测并将结果数值(字)化,或将对象分解并符号化,形成特征矢量或符号串、关系图,从而产生代表对象的模式,模式类中的个体在有些场合中也称为样本。用于学习与训练的样本的类别通常是已知的。另外,在进行特征提取之前,一般还需要对目标的有关信息进行预处理。

2. 特征选择

通常能描述对象的特征或属性的种类很多,为了节约资源,节省计算机存储空间、运算机时、

特征提取的费用,有时更是为了算法的可行性,在满足分类识别正确率要求的条件下,按某种准则尽量选用对正确分类识别作用较大的特征,使得用较少的特征就能完成分类识别任务。这项工作表现为减少特征矢量的维数、符号串字符数或简化图的结构。

3. 学习和训练

为使机器具有分类识别功能,如同人类自身一样,人们应首先对它进行训练,将人类的识别知识和方法以及关于分类识别对象的知识输入机器中,产生分类识别的规则和分析程序,这也相当于机器进行学习。这个过程一般要反复进行多次,不断地修正错误、改进不足,这包括修正特征提取方法、特征选择方案、判决规则方法及参数,最后使系统正确识别率达到设计要求。目前,机器学习常需要人工干预,这个过程通常是人机交互的。

4. 分类识别

在学习、训练之后,所产生的分析规则及程序用于未知类别对象的分类识别。需要指出的是,输入机器的人类分类识别的知识和方法以及有关对象知识越充分,机器中的知识与待识对象越匹配,知识的运用越合理,这个系统的识别功能就越强、正确率就越高。有些分类过程(如聚类分析)似乎没有将有关对象的类别知识输入,但实际上我们在选择相似性测度、采用某种聚类方法时,已经隐含地用到了对象的一些知识,也在一定程度上加入了人类的知识。

1.1.3 模式识别的基本方法

由于分类识别活动是人类广泛而重要的活动,人们希望机器能代替人类进行分类识别工作,因此模式识别的理论和方法引起人们极大的兴趣并进行了长期的深入研究,现已发展成一门多学科的交叉学科。这门学科涉及的理论与技术相当广泛,涉及多种数学理论、神经心理学、神经物理学、信息论、控制论、计算机科学、信号处理等。从本质上讲,这门学科实际上是数据处理与信息分析,而从功能上讲,可以认为它是人工智能的一个分支。

针对不同的对象和不同的目的,可以运用不同的模式识别理论、方法,目前主流的技术是:统计模式识别、结构模式识别、模糊数学方法、人工神经网络方法、人工智能方法、子空间方法等,它们之间往往存在一定的联系和借鉴。

1. 统计模式识别

这类识别技术理论较完善,方法也很多,通常较为有效,现已形成了一个完整的体系。尽管方法很多,但从根本上讲,都是直接利用各类的分布特征,即利用各类的概率密度函数、后验概率等,或隐含地利用上述概念进行分类识别。其中基本的技术为聚类分析法、判别类域代数界面法、统计决策法、最近邻法等。在聚类分析中,基本的思想是利用待分类模式之间的“相似性”进行分类,较相似的作为一类,较不相似的不作为一类。在分类过程中不断地计算所分划的各类的中心,待分类模式与各类中心的距离作为对其分类的依据。这实际上是在某些设定下隐含地利用了概率分布概念,因常见的概率密度函数中,距期望值较近的点概密值较大。该类方法的另一种技术是根据待分类模式和已指判出类别的模式的距离来确定其类别,这实际上也是在一定程度上利用了有关的概念。在最近邻法中,是根据待分类模式的一个或 k 个近邻训练样本的类别而确定其类别,其中 k -近邻法实际上是在样本情况下运用最大后验概率规则判决。在判别类域代数界面法中,首先用已知类别的训练样本产生判别函数,这相当于学习或训练,然后根据待分类模式代入判别函数后所得值的正负确定其类别,判别函数提供了相邻两类判别域的界面,其

也相应于在一些设定下两类概密函数之差。在统计决策法中,在一些分类识别准则下严格地按照概率统计理论导出各种判决规则,这些判决规则可以产生某种意义上的最优分类识别结果,这些判决规则要用到各类的概率密度函数、先验概率或后验概率。可以通过训练样本对未知概密函数中的参数进行估计,或对未知的概密函数进行逼近。

2. 结构模式识别

结构模式识别也称为句法模式识别。在许多情况下,对于具有较复杂结构特征的对象仅用一些数值特征已不能较充分地对其描述与正确识别,这时可采用结构识别技术。结构识别技术是将对象分解成若干个基本单元,这些基本单元称为基元,用这些基元以及它们的结构关系来表征对象,基元以及这些基元的结构关系可以用字符串或图来表示,这些字符串或图称为语言的句子,然后根据代表类的文法运用形式语言理论与技术对该句子进行句法分析,根据其是否符合某一类的文法而决定其类别。

3. 模糊数学方法(模糊模式识别)

这类识别技术运用模糊数学的理论和解决方法解决模式识别问题,因此适用于分类识别对象本身或允许识别结果具有模糊性的场合。模糊模式识别方法的基本思想是:将模式或模式类作为模糊集,将模式的属性值或属性转化为隶属度,运用隶属函数、模糊关系或模糊推理进行分类识别。目前,模糊模式识别方法较多,应用较广。这类方法的有效性主要在于对象类的隶属函数建立得是否良好,对象间的模糊关系的度量是否良好。

4. 人工神经网络方法

人工神经网络是由大量简单的基本单元——神经元(neuron)相互连接而构成的非线性动力学系统,每个神经元的结构和功能比较简单,而由其构成的系统却可以非常复杂,具有生物神经网络的某些特性,在自学习、自组织、自适应、联想及容错方面具有较强的能力,可用于联想、识别和决策。在模式识别方面,与前述方法显著不同的特点之一是在学习过程中具有自动提取特征的能力。

5. 人工智能方法

众所周知,人类具有极完善的分类识别功能,人工智能是研究如何使机器具有人类智能的理论和方法,模式识别从本质上讲就是如何根据领域的知识和对象的知识进行类别的判断,因此可将人工智能中有关知识表示、推理、学习等技术用于模式识别。

6. 子空间法

子空间法是将代数学的基本理论与统计学基本理论进行综合应用于模式识别中,这类方法的基本思想是:根据各类训练样本的相关阵通过线性变换由原始模式特征空间产生各类对应的子空间,这些子空间由原始模式类的样本相关阵的主要特征矢量所张成,这些主要特征矢量反映了模式分布结构信息,每个子空间与每个类别一一对应。在子空间法中,主要的分类决策规则有三个类型:基于投影长度的比较法,基于表达熵的比较法和基于统计假设检验的方法。基于投影的识别方法是根据待识别模式在各个子空间的投影大小判定该模式类别。

上述的六类方法各有特点及其应用范围,现在来看,它们不能相互取代,只能共存,相互促进、借鉴、渗透及融合。除了上述的六类方法外,还有其他的一些类型方法,如协同模式识别、仿生模式识别等。一个较完善的识别系统很可能是综合利用上述各类识别方法的观点、概念和技术而构成的。

1.1.4 模式识别的基本原则

实践经验和理论研究表明,不能脱离应用与问题简单地讲一种学习算法或识别算法比其他的某些算法更好,任何一个算法的优劣都不是绝对的,而只是相对的,它们只是相对一些应用或问题表现优良,但对另一些应用或问题则可能表现较差。讨论一个算法的优劣应在一定的背景下进行,所谓背景就是针对某一类具体的应用问题在一些性能技术指标下来进行比较。一个学习算法或识别算法的优劣总是根据(或相对)某些具体的性能指标或标准、针对具体问题而确定的,常考虑的性能指标包括:计算复杂度、空间复杂度、先验知识可利用性、训练阶段的识别正确率、训练之后工作时的识别正确率即推广性等,其中最后一个性能指标尤其重要。应用问题主要是指:对象的先验信息、数据维数、数据分布、训练样本数量、采用的目标函数、是否经济方便等。上述的技术指标有的相对地独立于具体的训练样本和应用,如计算复杂度,有的与训练样本和应用密切相关,如推广性。对不同的应用,上述的各技术指标的重要性也不尽相同。如果我们的目的是得到更好的推广性,那么不存在任何一种与对象知识或运用无关的更好的学习算法或识别算法,如果某种算法对某个特定问题比另一个算法更好,那也仅仅是因为它与问题匹配,它更适合这一特定的模式识别任务。在模式识别的理论中,存在与物理学类似的“守恒定律”,如果想在一些方面上得到性能提高,就不可避免地要在另外一些方面付出性能代价,广义地讲,得到(正的)和付出(负的)达到某种意义上的平衡(零和)。“没有免费午餐定理”阐述了这一基本规律,说某一个学习或识别算法比另一个分类性能更好,总是相对某个相关的目标函数而言的,但它们的识别错误率对所有可能的目标函数的求和结果是相等的。所有可能的目标函数意味着所有可能的应用。“没有免费午餐定理”表明,在没有“假设”的前提下,我们不能泛泛地说哪种算法是优越的,没有理由偏爱某一学习或分类算法而轻视另外一个,这些假设就是前面谈到的应用背景和性能指标。“没有免费午餐定理”告诉我们,“掌握更多的不同种类的技术,是实践者面对任意新的分类问题时仍能保持从容不迫态度的最佳保证”^[8]。

类似地,上述关于算法的观点对于特征的优劣也是成立的。“丑小鸭定理”表明,在没有“假设”的前提下,也不存在“优越”或“更好”的特征表达,不存在与问题无关的优越或更好的特征集合或属性集合。

一个算法优良性总是因为所选用的判决规则,学习算法,所使用的训练样本集,以及训练时的精度(或何时停止)与实际要解决的问题相匹配。因此,要想设计一个优良的分类器,必须深入了解问题,尽可能多地利用对象的各种知识,掌握和试探较多的分类识别方法,所有的目的都是为了使方法和问题匹配。

设计一个优良的分类器应该遵循“最小描述原理”:使模型的算法复杂度与和该模型相适应的训练数据的描述长度之和最小。算法复杂度的度量应独立于程序语言种类,这样能可靠地进行复杂度比较,可提供数据的内在的固有的信息量。一个算法可以表达为一个过程,执行 y 产生 x , 记为 $U(y) = x$, 一个二进制串 x 的戈氏复杂度定义为能够输出 x 的最短程序串 y 的长度

$$K(x) = \min_y [|y| \mid U(y) = x]$$

算法或模型与给定分类问题的匹配程度可以用“偏差”和“方差”来度量。偏差是度量算法或模型与问题匹配的准确性,一个高的偏差表示一个坏的匹配,而方差度量匹配的精确性,一个大的方差意味着一个弱的匹配。考虑两类问题,设判别函数 $d(x)$ 只取 1 或 0, 即 $d(x) =$

$P(y=1|x)=1-P(y=0|x)$, y 为类别标记, 令 $y=d(x)+\epsilon$, 式中 ϵ 是均值为零、方差较小的均匀分布的随机变量, 显然 $d(x)=E[y|x]$ 。求 $d(x, X)$ 使均方差

$$E_X[(d(x, X)-y)^2] \Rightarrow \min$$

式中 X 为训练样本集。在一些简单的设定下, 经运算可以得到一些有意义的结果, 由此可以得出如下的见解: 分类错误率可以被看作是判决“边界误差”, 边界误差可以表示成边界“偏差”和“方差”的组合。偏差与方差不是独立的, 方差对边界偏差的关系是高度非线性乘性的, 而且当方差很小时, 对边界偏差的符号将十分敏感, 从而边界偏差的符号影响了方差在误差中的应用。要想得到理想的零偏差和零方差, 唯一的方法就是事先要知道真实的模型。这在实际中是不可能的, 只能用样本数 N 很大的训练集和对 $d(x)$ 的形式的精确的先验知识来降低偏差和方差。一般来讲, 为了得到所需要的较小的推广误差, 小的方差要比小的偏差重要得多, 小的方差对于精确分类是重要的, 而小的边界偏差则不然, 即只要能保持方差很小, 就不必特别在意估计是否有偏。设计者可以调整分类器的偏差和方差。对于给定的偏差, 方差将随着样本数 N 的增加而减小。

1.2 特征矢量和特征空间

设一个分析对象的 n 个特征量测值分别为 $x_1, x_2, \dots, x_n$, 由于它们来自同一个对象, 所以应将它们作为一个整体一起考虑, 同时为了便于数学上处理, 让它们构成一个 n 维特征矢量 x , 即 $x=(x_1, x_2, \dots, x_n)^T$, x 是原对象的一种数学抽象, 用来代表原对象, 即为原对象的模式。对某对象的分类识别实际上是对它的模式, 即它的特征矢量进行分类识别。各种不同取值的 x 的全体构成了 n 维空间, 这个 n 维空间称为特征空间, 不同场合的特征空间可记为 X^n 、 $\mathbb{R}^n$ 或 Ω 。特征矢量 x 便是特征空间中的一个点, 所以特征矢量有时也称为特征点。

经验表明, 对于同一个对象的某种特征进行多次量测所得到的结果往往是不同的 (如果量测设备精度足够高的话), 对同一类不同对象的某一特征的量测结果一般也是按某种规律分布的, 因此, 同一个对象或同一类对象的某特征量测值是随机变量, 这是由于量测系统随机因素的影响及同类不同对象的特征本身就是特征空间散布的。由随机分量构成的矢量称为随机矢量, 从而可知, 同一类对象的特征矢量在特征空间中是按某种统计规律随机散布的。应该指出的是, 不同类对象的某一特征的量测值或量测值统计特性有时可能是相同的, 但如果特征选择适当, 则它们的特征矢量应有尽可能多的分量具有不同的统计特性。

1.3 随机矢量的描述

1. 随机矢量的分布函数

设 $X=(X_1, X_2, \dots, X_n)^T$ 为随机矢量, $x=(x_1, x_2, \dots, x_n)^T$ 为确定性矢量。随机矢量 X 的联合概率分布函数定义为

$$F(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) \quad (1-3-1)$$

式中 $P(\cdot)$ 表示括号中事件同时发生的概率。写成矢量形式

$$F(x) = P(X \leq x) \quad (1-3-2)$$

这里约定,一个矢量小于(或大于)另一个矢量是指它的每个分量均小于(或大于)另一个矢量的对应分量。随机矢量 $\mathbf{X}$ 的联合概率密度函数定义为

$$p(x_1, x_2, \dots, x_n) \triangleq p(\mathbf{x}) = \partial^n F(x_1, x_2, \dots, x_n) / (\partial x_1 \partial x_2 \cdots \partial x_n) \quad (1-3-3)$$

设集合由 c 类模式组成,第 i 类记为 ω_i , ω_i 类模式特征矢量有其自己的分布函数和密度函数。 ω_i 类的模式特征矢量的分布函数及密度函数分别定义为

$$F(\mathbf{x} | \omega_i) = P(\mathbf{X} \leq \mathbf{x} | \omega_i) \quad (1-3-4)$$

$$p(\mathbf{x} | \omega_i) = \partial^n F(x_1, x_2, \dots, x_n | \omega_i) / (\partial x_1 \partial x_2 \cdots \partial x_n) \quad (1-3-5)$$

2. 随机矢量的数字特征

(1) 均值矢量(期望矢量)

n 维随机矢量 $\mathbf{X}$ 的数学期望 $\boldsymbol{\mu}$ 定义为

$$\boldsymbol{\mu} = E[\mathbf{X}] \triangleq \bar{\mathbf{X}} = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix} \triangleq \int_{X^n} \mathbf{x} p(\mathbf{x}) d\mathbf{x} \quad (1-3-6)$$

其中, $\boldsymbol{\mu}$ 的第 i 个分量

$$\mu_i = E[X_i] = \int_{-\infty}^{\infty} x_i p(x_i) dx_i = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i p(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n \triangleq \bar{X}_i \quad (1-3-7)$$

式中, $p(x_i)$ 是 $\mathbf{X}$ 的第 i 个分量 X_i 的边缘密度。随机矢量 $\mathbf{X}$ 的均值矢量 $\boldsymbol{\mu}$ 的各分量是相应的各随机分量的均值。

(2) 条件期望

在模式识别中,经常以类别 ω_i 作为条件,在这种情况下,随机矢量 $\mathbf{X}$ 的条件期望矢量定义为

$$\boldsymbol{\mu}_i = E[\mathbf{X} | \omega_i] = \int_{X^n} \mathbf{x} p(\mathbf{x} | \omega_i) d\mathbf{x} \quad (1-3-8)$$

(3) 协方差矩阵

随机矢量 $\mathbf{X}$ 的自协方差矩阵(以后矩阵常简称为阵)表征各分量围绕其均值的散布情况及各分量间的相关(协同变化)关系,其定义为

$$\boldsymbol{\Sigma} = E[(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})^T] = \int_{X^n} (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T p(\mathbf{x}) d(\mathbf{x}) \triangleq (\sigma_{ij}^2)_{n \times n} \quad (1-3-9)$$

式中, σ_{ij}^2 是 $\mathbf{X}$ 的第 i 个分量与第 j 个分量的协方差,当 $i = j$ 时, σ_{ii}^2 便是 X_i 的方差。

$$\sigma_{ij}^2 = E[(X_i - \bar{X}_i)(X_j - \bar{X}_j)] = \iint_{-\infty}^{\infty} (x_i - \mu_i)(x_j - \mu_j) p(x_i, x_j) dx_i dx_j \quad (1-3-10)$$

(4) 自相关矩阵

随机矢量 $\mathbf{X}$ 的自相关矩阵定义为

$$\mathbf{R} = E[\mathbf{X}\mathbf{X}^T] \quad (1-3-11)$$