

赵树芳

模式识别的 模糊数学方法

西北电讯工程学院出版社

模式识别的模糊数学方法

赵 树 芑

西北电讯工程学院出版社

1 9 8 7

内 容 简 介

《模式识别的模糊数学方法》一书用模糊数学理论研究模式识别，强调了不确定性下的模式识别方法。本书不是仅把模糊识别作为一种个别方法，而是作为研究模式识别的一种理论基础，并且和经典研究方法结合起来，渗透到全书各部分，尽可能给出模式识别的统一描述。

本书适合于计算机科学、自动化、信号与系统等有关专业的研究生及本科高年级学生使用，也适合有关科技工作者阅读。

模式识别的模糊数学方法

赵 树 芑

责任编辑 徐德源

西北电讯工程学院出版社出版发行

西北电讯工程学院印刷厂印刷

新华书店经销

开本 787×1092 1/16 印张 8 8/16 字数 200 千字

1987 年 12 月第 1 版 1987 年 12 月第 1 次印刷 印数 1—1 400

ISBN 7-5606-0028-X/TN·0008

统一书号：15322·94

定价：1.50元

序

模式识别是一门新兴的、发展迅速的学科，正在日益显示其理论价值和应用价值。就研究的方法而言，从经典的统计方法开始，接着有句法识别方法，继而出现模糊识别方法。前两种方法，经历了较长时间的研究，已经建立了比较完整的理论体系和方法论，有了不少论著。模糊识别则是近几年才开拓的新领域，还缺乏比较系统的论著。本书正是在这一引人注目，而且很有发展前途的新领域中绽开的一朵新花。

和其他模式识别论著不同的主要之点是，本书以模糊识别作为一条主轴贯穿始终。作者在论述中，不是仅把模糊识别作为一种方法，而是作为研究模式识别的一种理论基础，并且和两种经典的研究方法结合起来，渗透到全书的各个部分。这样，我们就有可能希望通过模糊识别的理论，给出模式识别的统一描述。这是本书的特点，也是作者用心良苦的大胆尝试。

作者多年来一直从事信息处理、字符识别等方面的研究工作；近年来，对模糊识别的基础理论、模糊测度、聚类分析等专题，作了较深入的探讨；结合这些研究工作，为研究生开设了有关课程。本书是在这样的研究与教学相结合的实践基础上写成的。因此，本书的内容不仅体现了对新的学科的理论方法的严谨论述，也充分考虑了读者学习的需求，行文流畅，论理明晰，是一本可读性很强的著作。

模式识别是一种普遍的智能行为，而识别过程一般总是在不确定的，给定的信息不完全的和有干扰的情况下进行判定的，这是模式识别理论研究中的难题，也是当前信息科学研究工作中的一个热点，同时也是世界各大国的高技术计划中寻求解决的重要课题。本书恰好强调了识别过程的不确定性，提出了一些建设性的意见，我觉得这也是值得重视的一个特点，是本书的重要贡献之一。

蔡希尧

1987年3月

前 言

模式识别技术的发表已经有 30 年的历史, 已经形成了一门内容丰富、影响广泛、在理论上较为系统的学科。模式识别的研究方法有三大类: 统计模式识别、句法(结构)模式识别与模糊模式识别。统计模式识别发展最早, 到本世纪 70 年代初已形成了系统的理论。句法模式识别是 70 年代发展起来的, 目前也已经有了较为系统的理论。模糊模式识别是模式识别的一个最新分支, 它的研究大约是从 70 年代中期开始的, 至今尚未形成系统的理论。但是, 由于模糊识别方法能更好地反映出模式识别中的不确定性质, 它的研究已经引起人们极大的注意。

本书是一本以模糊识别方法为主导的模式识别原理的著作。本书的特点是, 模糊模式识别不仅仅被作为模式识别的方法之一来讨论, 而是作为整个模式识别的理论基础被提出与研究的。模糊识别的观点贯穿在所有的章节之中, 把上述三大类模式识别方法联系起来, 希望能对模式识别理论给以统一的描绘。这是一个新的尝试, 也是个困难的尝试。正因为如此, 本书远远不能说是成熟的与完善的。如果它的问世能够给读者一些新的启发, 在同行中起到抛砖引玉的作用, 也就达到作者的目的了。

全书共有 11 章, 分作 3 篇。第 1 篇是概论, 包括第 1 章到第 3 章。其中第 2 章与第 3 章是本书的理论基础。在这两章里, 作者加入了一些新的观点和新的内容, 特别是以测度为基础的模糊集合运算与软匹配。第 2 篇是分类方法, 包括第 4 章到第 8 章。在这一篇里讨论了分割理论以及各种模式分类方法, 其中有模糊模式分类、经典的统计模式分类、句法模式分类和聚类方法。在所有这些方法的讨论中, 都强调了不确定性下的软匹配概念。第 3 篇是模板设计, 也就是学习或训练的问题。这一篇包括第 9 章到第 11 章。

在编写本书时, 得到了计算机系蔡希尧教授与学院甘圣予副院长的鼓励与支持, 任培以同志阅读与整理了全书的手稿。作者对他们表示衷心的感谢。

由于作者水平有限, 书中错误与不妥之处在所难免, 希望读者予以指正, 作者对此将不胜感激。

赵树芑识于西北电讯工程学院

1987 年 3 月

目 录

第1篇 概 论

第1章 什么是模式识别

§ 1.1 模式识别技术发展简史	1
§ 1.2 模式识别技术的重要应用	2
§ 1.3 模式识别的基本概念	4
参考文献	7

第2章 相似测度与模糊集合

§ 2.1 距离	8
§ 2.1.1 距离的一般定义	8
§ 2.1.2 常用的几种距离	9
§ 2.2 相似度	9
§ 2.2.1 相似度的定义	10
§ 2.2.2 相似度与距离的关系	11
§ 2.3 模糊集合的若干概念	12
§ 2.3.1 扎德的模糊集合定义	12
§ 2.3.2 以模糊测度为基础的模糊 集合定义	13
§ 2.3.3 相似度与隶属度	14
§ 2.3.4 条件隶属度	14
§ 2.4 模糊集合的运算	16
§ 2.4.1 扎德运算及其他	16
§ 2.4.2 测度运算	17
§ 2.4.3 测度运算的性质	17
§ 2.5 复合事件的模糊测度	21
§ 2.6 模糊关系	22
参考文献	26

第3章 模式识别方法论

§ 3.1 引言	27
§ 3.2 模板匹配	28
§ 3.3 软匹配	30
§ 3.4 多点匹配	31
§ 3.5 分类方法	32
§ 3.6 模板设计	33
参考文献	34

第2篇 分类方法

第4章 集合与分割

§ 4.1 特征空间与特征集合	35
§ 4.2 硬分割	36

§ 4.3 软的二分割	39
§ 4.4 软的 c 分割	40
§ 4.5 从软分割到硬分割的转化	40
§ 4.6 关于分割的再探讨	41
参考文献	43

第5章 隶属函数与鉴别函数

§ 5.1 直接模糊分类	44
§ 5.2 最小距离分类	45
§ 5.2.1 单模板情况	45
§ 5.2.2 多模板情况	46
§ 5.3 线性分类	47
§ 5.3.1 线性鉴别函数	47
§ 5.3.2 多维情况	47
§ 5.4 线性鉴别函数的性质	51
§ 5.4.1 模式空间与加权空间	51
§ 5.4.2 超平面的几何性质	53
§ 5.4.3 二分法能力	54
§ 5.5 距离鉴别函数与线性鉴别函数 的比较	56
参考文献	56

第6章 聚类分析

§ 6.1 用模糊关系矩阵聚类	57
§ 6.1.1 普通集合上的聚类	57
§ 6.1.2 模糊集合上的硬聚类	58
§ 6.1.3 模糊关系矩阵的建立	60
§ 6.1.4 直接由模糊相似矩阵聚类	62
§ 6.1.5 软聚类	63
§ 6.2 用目标函数聚类	64
§ 6.2.1 硬聚类	64
§ 6.2.2 软聚类	67
参考文献	69

第7章 统计模式识别

§ 7.1 贝叶斯决策理论	70
§ 7.2 正态分布情况	71
§ 7.3 模糊事件的概率测度	73
§ 7.4 正态分布情况	74
参考文献	76

第8章 句法模式识别

§ 8.1 引言	77
§ 8.2 描述模式的各种文法和语言	78
§ 8.2.1 PDL 文法	78
§ 8.2.2 树状文法	79
§ 8.2.3 高维文法	81
§ 8.2.4 属性文法	82
§ 8.3 模糊文法	84
§ 8.3.1 模糊语言	84
§ 8.3.2 模糊文法	84
§ 8.3.3 模糊自动机	85
§ 8.4 形式语言的误差校正剖析	90
§ 8.4.1 两个句子模式之间的距离	90
§ 8.4.2 一个句子与一个语言之间的距 离	92
§ 8.4.3 误差校正剖析	92
§ 8.4.4 树状自动机的误差校正	95
参考文献	99

第3篇 模板设计

第9章 隶属函数的确定

§ 9.1 推理法	101
§ 9.1.1 三角形的模板	102
§ 9.1.2 相似度隶属函数	102
§ 9.2 模糊统计	103
§ 9.3 正态分布的利用	104

§ 9.4 参数估值	106
参考文献	110

第10章 线性鉴别函数的学习

§ 10.1 目标函数	111
§ 10.1.1 感知准则	112
§ 10.1.2 费舍尔准则	112
§ 10.1.3 最小平方误差准则	114
§ 10.2 学习算法	114
§ 10.2.1 梯度下降法	114
§ 10.2.2 感知准则函数法	115
§ 10.2.3 松弛法	117
§ 10.2.4 费舍尔算法	117
§ 10.2.5 最小平方误差法	118
§ 10.2.6 何-卡氏 (Ho-Kashvap) 法	119
§ 10.3 普遍性与收敛性	120
§ 10.3.1 普遍性	120
§ 10.3.2 收敛性	120
参考文献	122

第11章 文法推断

§ 11.1 一些基本概念	123
§ 11.2 有限状态文法的推断	124
§ 11.3 二维文法推断	126
参考文献	127

第1篇 概 论

第1篇包括三大部分内容。第1章是引言,内容包括模式识别的发展简史、模式识别的重要应用领域和模式识别的基本概念等。第2章是预备知识,介绍模式识别中分类的基本准则、相似度以及模糊集合的基本概念,这些是模式识别研究中的基础理论。第3章是模式识别方法论,在这一章中概要地将以后各章所用的基本分析方法加以统一的描述,以便读者对本书的全貌及其特点有一个概括的了解,为阅读以后各章节打一个基础。

第1章 什么是模式识别

本章概括介绍模式识别技术,使读者对于什么是模式识别有一个大致的了解。全章分三个部分。在§1.1中,我们扼要谈谈模式识别技术的发展历史。§1.2介绍模式识别技术的一些主要应用领域,这里我们尽可能罗列各种各样的用途。由于现今模式识别的用途已经十分广泛,对于模式识别应用的深入介绍,需要专门的著作,我们就不准备就这一问题在本节以及本书其他章节中作深入全面的讨论。§1.3介绍模式识别技术的基本概念和基本方法,为本书以后各章节提供一个轮廓。

§1.1 模式识别技术发展简史

模式识别技术是随着计算机的发展而发展起来的。从本世纪60年代开始,人们已着手研究如何应用计算机进行文字、图形、数据等模式的识别和分类。

如果追根溯源,模式识别技术早在电子计算机发明以前就已经存在了。例如,1929年美国的陶舍克曾在一份专利中描述过这样一台机器,它能识别用铅字印成的数字。陶舍克的机器用10块镂空的模板分别对应10个数字。在模板上投射数字的正象,如果模板与被投影数字一样,则被投影数字的黑色笔划就覆盖着模板的所有透明部分,只有很少的光线能透过模板,但被投影数字的黑色笔划却覆盖不了其他模板的所有透明部分。为了识别文件上的一个数字,陶舍克的机器就把这个数字的像依次投射到10块模板上,用光电管测量透过每一块模板的总光量,这个数字就被识别为与透过的总光量最小的那个模板对应的数字。换句话说,把一个数字识别为与它匹配的那块模板的数字。这类机器可以说是模式识别最早的应用。

然而模式识别技术的真正的发展时期仍然是在电子计算机出现和推广应用以后。在大约30年的时间里,模式识别技术得到了迅速的发展。现在已广泛应用在许多领域中。

如前面提到的,最早的模式识别方法是利用“模板匹配”的概念。这种概念十分简单明了,以致并没有引起多少专门的理论研究。人们希望找到更深入地描述模式识别的理论。60年代至70年代初,研究的注意力集中在用统计决策理论描述模式识别问题,即统计模式识

别^[2]。这是不难理解的,因为识别首先就是一种决策过程,而到60年代时已经颇为成熟的统计决策、检测和估值理论为统计模式识别理论的建立铺平了道路。

在成功地解决了许多模式识别问题的同时,统计方法也遇到了一些难以克服的困难。首先,统计模式识别要求对识别对象有正确的测量值特征,要知道其统计特性,但实际上这往往做不到。其次,当模式很复杂时,特征数量变得非常庞大,分析处理工作将非常复杂。这一困难把许多研究者引导到一个与统计模式识别不同的方法上,其基本思想就是,一个复杂的模式可以由一些简单模式以递归形式描述。例如,数字可以由笔划描述。这种递归描述与文章中的语句可以通过单词来描述十分相似。由于这一原因,许多形式语言中的方法都可以被引用,这就是句法模式识别^[3]。句法模式识别也称为结构模式识别,这是因为它对模式的描述是以结构信息的利用为基础的。句法模式识别的研究自70年代以来得到很大的发展。

由于大多数实际模式具有模糊不确定性,用模糊集合描述与分析模式识别问题是合理的发展方向。模糊集合理论是60年代中期由札德(L.A.Zadeh)提出的。随着模糊集合理论的发展,出现了一个新的研究模式识别的方法——模糊模式识别^{[4], [5]}。实际上,模式识别的研究是引起模糊集合理论的出现和推动模糊集合理论发展的一个重要因素。

本章末尾参考文献[6]—[26]中给出了21本有关模式识别的专著。从这些书中可以了解模式识别的主要内容和历史。应该指出,模式识别尚处于发展的幼年时期,许多内容尚不很成熟,不很系统,甚至它包含的范围也不是很明确和很确定的,其中模糊模式识别尤其如此。随着科学技术的发展,这一学科的内容也必定会日新月异地变化而更加丰富起来。

§ 1.2 模式识别技术的重要应用

经过20多年的迅速发展,模式识别技术已经相当广泛地应用于各个领域,这些领域不但包括工农业和自然科学、也包括第三产业和社会科学。例如,文字识别、心电图与脑电图分析、语音识别、指纹辨认、医学诊断、地震波分析、目标检测、天气预报、机械零件探伤、加工过程监视、照片与景物分析等,都是模式识别的一些重要应用领域。在不远的将来,模式识别的最重要的用途也许是机器人与第5代、第6代计算机。

下面更全面地列举模式识别的各个应用领域,以备参考^[23]。

1. 科学应用

- (1) 天文:望远镜分辨率的改进和大气损失的去除。
- (2) 地球和行星探测:火山口计数,颜色分析,地形测量,大气测量和分析,危险点估计,天体地质特征分析与综合,机器人探险。
- (3) 地质:大地测量与绘图,照片嵌拼,表面模型拟合,地图绘制。
- (4) 泡沫室:粒子轨迹与电子显微结晶学。
- (5) 特殊图象的产生。
- (6) 卫星数据分析。
- (7) 遥远星球生命探测与数据分析。

2. 生命与行为科学

- (1) 人类学。
- (2) 考古学。

- (3) 昆虫学。
- (4) 生物学与植物学：微生物学，生态学，动物学。
- (5) 心理学：社会心理学，犯罪心理学。
- (6) 控制论。
- (7) 信息管理系统。
- (8) 教育。
- (9) 通信交往。

3. 工业应用

- (1) 文字识别。
- (2) 用图象控制机械(过程控制)。
- (3) 签名分析。
- (4) 语音分析。
- (5) 照片识别。
- (6) 探矿(地下分析)。
- (7) 内变形探测(X光的与超声的)。
- (8) 商业照片增强。
- (9) 电影片模拟。
- (10) 电子玩具设计。
- (11) 自动化细胞学。

4. 医学应用

- (1) 显微观察与生物医学数据：血细胞计数与验血，癌细胞辨认与检验，神经测量，染色体核分类，骨骼成分分析，自动聚焦与定位，脑组织研究。
- (2) 放射性同位素检查。
- (3) X射线检查与断面照像术，血管厚薄测量，心脏形状测量，脑癌检测，脑血栓检测，牙的绘图与分析，骨骼结构分析，脑病诊断，骨架结构分析。
- (4) 心电图与矢量心电图分析。
- (5) 脑电图描绘与神经生物信号处理。
- (6) 药物作用。
- (7) 基因染色体研究。

5. 农业应用

- (1) 收成分析。
- (2) 土壤分析。
- (3) 过程控制。
- (4) 地球资源摄影。

6. 行政应用

- (1) 天气预报：云跟踪与水温测量。
- (2) 专用系统：交通分析与控制，城市增长决定，烟雾检测与测量，空中交通雷达数据处理。
- (3) 地球资源数据与遥感。

7. 某些军事用途

- (1) 空中摄影与遥感。
- (2) 声纳信号检测与分类。
- (3) 目标辨识。

§ 1.3 模式识别的基本概念

模式(Pattern)这个词的意思是指对事物的描写,或者是指作模仿用的完美的模型。识别(Recognition)是基于对事物性质的认识而能够对事物正确分类与描述,它被认为是人类及其他生物的基本属性。我们每时每刻都在进行和完成识别的动作。人类之所以是一种很高等的信息系统,部分原因就是人类具有高超的模式识别能力。

模式识别不仅指感官对物体的感觉,它也是人们的一种基本的思维活动。根据被识别模式的性质,我们可以把识别动作分为两大类:具体事物的识别与抽象事物的识别。例如,对文字、照片、音乐、语言等周围事物的识别,就属于具体事物的识别,可以叫做感性认识,这一类识别涉及对于空间时间中具体事物的辨认和分类;另一方面,我们能识别一个已知的论点,或一个问题的解,这并不靠眼睛或耳朵,这种过程涉及对抽象事物的识别,可以叫做概念性的或理性的认识。

人类对模式识别过程的机理,目前仍不很清楚。对于具体事物的认识主要是心理现象,对于抽象事物的认识则主要是思维现象。人们对于具体事物的认识,涉及人与客观事物在人类感官中所引起的刺激之间的关系。当一个人感受到一个模式时,他把此感觉与他从自己过去的经验中得来的一般概念或线索结合起来,并做出归纳性的推理判断。人类识别实际上就是估计输入数据与一组已知的具有不确定性的模板相匹配时的差异。这些模板依赖于人们过去的经验并形成识别的线索和先验信息。这样,模式识别问题可视为通过搜索各模板特征,对输入数据与各模板之间的相似性进行鉴别。因此,模式识别过程可以在逻辑上分为两个主要内容:

- (1) 对于人类和其他生物的模式识别能力的研究。
- (2) 为设计具有实现特定用途的识别能力的设备的理论研究与技术开发。

第1个问题涉及有关诸如心理学、生理学与生物学等学科,第2个问题主要涉及工程、计算机与信息科学等领域。我们主要讨论后面一类问题。

设计模式识别系统一般包括以下几个主要内容。首先是对被识别事物测量而得的输入数据的表示。每个测量值描述物体(模式)的一个特性。举例说,假设所讨论的模式为字符,这时可用图 1.1 所示的网格表示模式,图中每个小方格称为一个像素(元素)。假设网格共有 d 个像素,则字符可用模式矢量表示为

$$\mathbf{x} = [x_i] = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix} = [x_1 \ x_2 \ \cdots \ x_d]^T \quad (1.3-1)$$

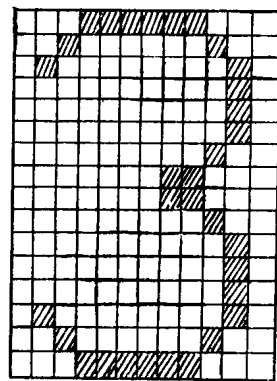


图 1.1 字符的网格表示

或用集合表示为

$$X = \{x_i\} = \{x_1, x_2, \dots, x_d\} \quad (1.3-2)$$

其中 i 为按一定扫描次序排列的由 1 到 d 的下标。当第 i 个像素属于字符的一部分时, $x_i = 1$, 否则 $x_i = 0$ 。模式矢量中包含模式的一切有用信息, 对于某类模式的物体的测量, 可视为给每个模式特性(即每个像素)指定集合 $\{x_i\}$ 中的一个符号。

当测量结果为一实数时(通常是如此), 可以把模式矢量想象为 d 维欧氏空间里的一点。属于同一类的各模式对应于散布在测量空间的某一区域中的点的总体。图 1.2 是一个简单的例子, 其中有 两类模式, 用 ω_1 和 ω_2 代表。在此例中, 假设 ω_1 和 ω_2 分别为一群男人和一群女人。每个模式由两个测量的特性, 身高和体重表示出来。因此, 模式矢量的形式为 $\mathbf{x} = [x_1 \ x_2]^T$, 其中 x_1 代表身高, x_2 代表体重。每个模式可看作是两维空间里的一点。为了表示清楚简单, 图中的两类是不相交的。当然, 在实际情况下, 不能保证测量的结果是不相交的集合。

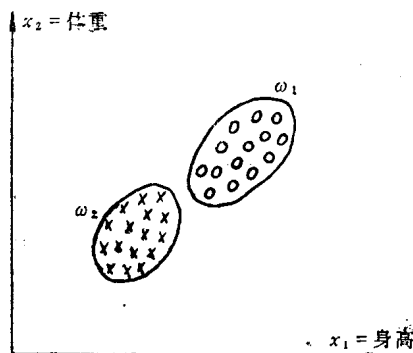


图 1.2 两类不相交模式

设计模式识别系统的第 2 个问题是 从输入数据中提取特征(或属性), 并压缩模式矢量的维数。这常常称为预处理和特征提取。例如, 在语音识别中, 常用的特征是声音持续时间, 不同频率范围的能量比、语峰位置和峰值随时间的移动等。

一个模式类中一般包括若干个模式。特征是针对模式类而言的。一个模式类的特征代表该类中所有模式的共性, 通常称为类内特征。代表不同模式类的差别的特征则称为类间特征。那些对所讨论的所有类别共有的类内特征不包含可鉴别的信息, 可以忽略。特征提取被认为是模式识别中的重要问题。如果能确定出每类模式的鉴别特征的完备集, 那么模式的识别与分类就不会有多大困难。此时, 自动识别就简化为简单的匹配过程, 或查表过程。然而, 在实际出现的大多数模式识别问题中, 确定鉴别特征完备集, 如果不是不可能, 那也是一个十分困难的问题, 幸而我们常常能从观测的数据中找出某些鉴别特征, 利用这些特征可以简化模式识别问题。

设计模式识别系统的第 3 个问题是确定最终决策过程。在被识别模式的观测数据已经被表示为模式空间的模式点以后, 我们希望识别机决定这些数据属于哪一类模式。假定此机器设计为能识别用 $\omega_1, \omega_2, \dots, \omega_c$ 代表的 c 个不同类别, 则可认为模式空间包含 c 个区域, 其中

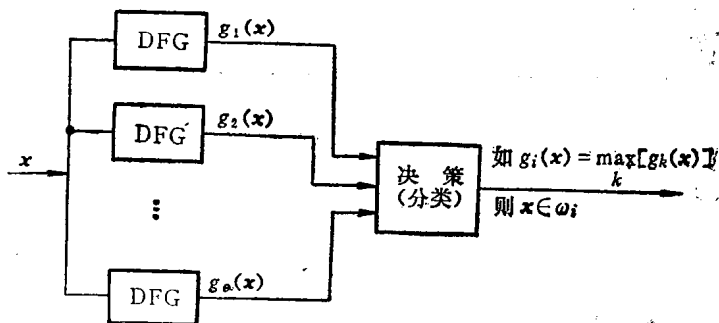


图 1.3 模式分类器框图

每一个包含某一类的模式点。这样，模式识别问题可看作是产生决策边界的问题，这些边界根据观测到的矢量将 c 类模式分割开。让我们定义决定这些边界的决策函数是 $g_1(\mathbf{x}), g_2(\mathbf{x}), \dots, g_c(\mathbf{x})$ ，这些函数也称为鉴别函数，它们是模式 $\mathbf{x}$ 的标量单值函数。如果 $g_i(\mathbf{x}) > g_j(\mathbf{x})$ ， $i, j = 1, 2, \dots, c, i \neq j$ ，则模式 $\mathbf{x}$ 属于模式类 ω_i 。换句话说，如果决策函数 $g_i(\mathbf{x})$ 对模式 $\mathbf{x}$ 具有最大值，则 $\mathbf{x} \in \omega_i$ 。图 1.3 为使用决策过程的自动分类方案的框图，其中 DFG 表示决策函数产生器。

DFG 如何构成，图 1.3 未表示出来。实际上，决策函数可有不同的实现方案。当被识别模式的先验知识完全知道时，决策函数产生器可根据这些知识准确地予以设计。换句话说，在此情况下可以事前精确设计出模式的模板，如图 1.4 所示。该图中还将预处理与特征提取

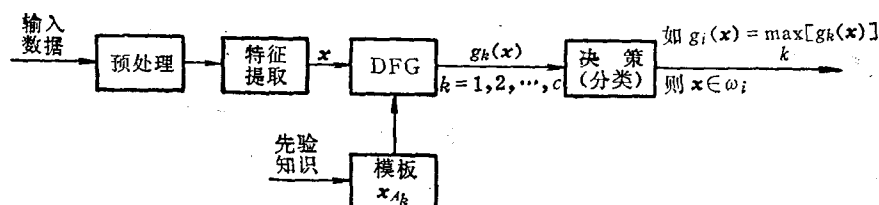


图 1.4 有完整先验知识情况下的模式识别系统框图

两部分单独画出。如前所述，特征提取是模式识别所必需而又高度面向具体问题的。预处理是为特征提取做准备的，因而也随具体问题而有不同的内容。但是除此以外，预处理有共同性的任务，即对输入数据进行滤波，除去信号中的噪声和干扰；以及对信号进行数字化处理，以便于计算机接受。此外，我们把图 1.3 中的多个 DFG 合并在一起，用一个 DFG 来表示。此处 DFG 对输入模式矢量 $\mathbf{x}$ 与先验模式模板 $\mathbf{A}$ 的矢量 $\mathbf{x}_{A_k}$ ($k = 1, 2, \dots, c$) 进行比较。 $\mathbf{x}$ 与某个模板 $\mathbf{x}_{A_k}$ 符合度越高，则由 DFG 产生的 $g_i(\mathbf{x})$ 也就越大。

当被识别模式的先验知识不完全时，为了设计模式的模板，须利用训练(或学习)的方法。一开始可以根据不完全的先验知识给出一模板，其中包括一些未知参数。这些参数可以利用一些已知子样(输入模式)，用统计的方法加以估值。当训练子样的数目足够大时，就可以很精确地估计出参数值。这些训练子样同样要经过预处理和特征提取。实际上，这里的预处理和特征提取部件的构造可以与识别支路中的完全相同，或者共用一套。图 1.5 是这种情况下模式识别系统的框图。

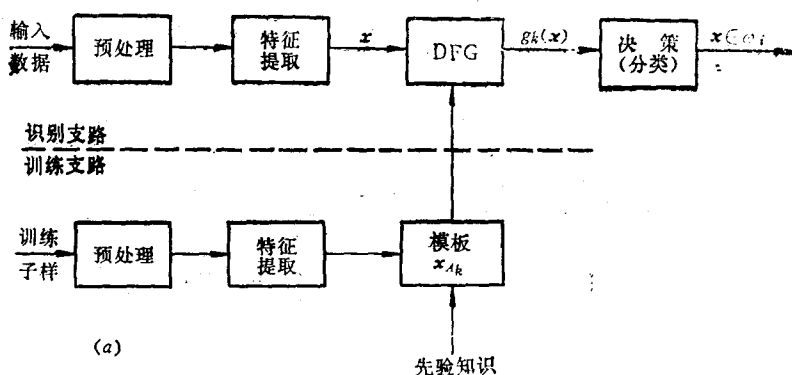


图 1.5 先验知识不完整情况下的模式识别系统框图

参 考 文 献

- [1] G.Tauschek, Reading Machines, U.S.Pat.2,026,329, Dec.1935.(Appl.May 1929)
- [2] C.K.Chow, An Optimum Character Recognition System Using Decision Functions, IRE Trans.EC-6, 217, 1957.
- [3] R.Narasimhan, A Linguistic Approach to Pattern Recognition, Rept.21, Digital Computer Lab., Univ.of Illinois, 1962.
- [4] R.E.Bellman, R.Kalaba, L.A.Zadeh, Abstraction and Pattern Classification, J. Math.Anal.& Appl., Vol.13, 1-7, 1966.
- [5] E.Ruspini, A New Approach to Clustering, Inf.Control, Vol. 15, 22-32, 1969.
- [6] M.J.Nilsson, Learning Machines-Foundations of Trainable Pattern-Classifying Systems, McGraw-Hill, 1965.
- [7] М.М.Бонгард, Проблема Узнавания, Наука, 1967.(有英译本)
- [8] P.W.Becker, Recognition of Patterns, Polyteknisk Forlag, 1968.
- [9] E.A.Patrick, Fundamentals of Pattern Recognition, Prentice-Hall, 1972.
- [10] K.Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, 1972. (有中译本)
- [11] Н.Г.Зароруйко, Методы Распознавания ч Их Применения, Изд-во Советское Радио, 1972.
- [12] R.O.Duda, P.E.Hart, Pattern Classification, and Scene Analysis, John Wiley & Sons, 1973.
- [13] C.H.Chen, Statistical Pattern Recognition, Hayden Book Co., 1973.
- [14] J.R.Ullman, Pattern Recognition Techniques, Butterworth Co., 1973.(有中译本)
- [15] J.T.Tou, R.C.Gonzalez, Pattern Recognition Principles, Addison-Wesley.1974.
- [16] K.S.Fu, Syntactic Methods in Pattern Recognition, Academic Press, 1974.
- [17] K.S.Fu(ed.), Digital Pattern Recognition, Springer-Verlag, 1976.
- [18] A.Rosenfeld(ed.), Digital Pattern Analysis, Springer-Verlag, 1976
- [19] В.А.Ковалевский, Методы Оптимальных Решений ч распознаваний Изображений, Наука, 1977.(有英译本)
- [20] R.C.Gonzalez, M.G.Thomason, Syntactic Pattern Recognition-An Introduction, Addison-Wesley, 1978. (有中译本)
- [21] T.Pavlidis, Structural Pattern Recognition, Springer-Verlag, 1978.(有中译本)
- [22] J.C.Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press 1981.
- [23] A.Kandel, Fuzzy Techniques in Pattern Recognition, John Wiley & Sons, 1982.
- [24] 傅京孙:《模式识别及其应用》, 科学出版社, 1984.
- [25] 李月景:《图象识别技术及其应用》, 机械工业出版社, 1985.
- [26] 蔡元龙:《模式识别》, 西北电讯工程学院出版社, 1986.

第2章 相似测度与模糊集合

在对一些待识别模式进行分类时,通常是按照它们之间的距离或某种相似性进行的。两个模式之间的距离愈小,则它们属于同一类别的可能性愈大;反之,距离愈大则属于同一类别的可能性愈小。或者用相似性来描述:两个模式之间的相似度愈大,则它们属于同一类别的可能性愈大;反之,相似度愈小则属于同一类别的可能性愈小。显然,相似性与距离是相反的概念。相似性愈大则距离愈小,反之亦然。

距离与相似度的概念,在模式识别技术发展初期就被利用来作为统计模式识别及聚类分析的分类准则。在§2.1与§2.2中,我们分别给出距离与相似度的一般定义以及各种不同的具体定义。在§2.2中我们还研究了相似度与距离之间的关系。

在模糊识别中,人们使用隶属度作为分类的准则。某个待识别的模式对于某一模式类的隶属度愈大,则该模式属于该类的可能性愈大。或者换一种描述,让我们研究待识别模式对于不同类别的隶属度,这些隶属度的数值一般是各不相同的,我们认为待识别模式属于隶属度最大的那一类。

隶属度是模糊集合的基本概念与核心问题。目前,模糊集合理论已日益为人们所熟悉,有不少著作可以参考^{[1]~[8]}。我们准备全面介绍模糊集合的内容,而只讨论一些与模式识别密切有关的问题,并介绍一些新的见解。

在§2.3中,我们讨论模糊集合的基本概念和模糊集合理论的创始人扎德关于模糊集合的定义。与此同时,我们还提出了一种模糊测度的概念和以模糊测度为基础的新的模糊集合的定义,并讨论相似度与隶属度之间的关系。事实上,无论隶属度、距离还是相似度,都是某种测度,我们把它们看作模糊测度,并且证明,模糊集合中的隶属度与模式识别中的相似度属于同一概念。这样,我们就有可能把经典的模式识别与模糊模式识别的理论统一起来。本节还讨论了条件隶属度的概念,并在§2.5中将其用于复合事件隶属度的描述。

§2.4介绍模糊集合的运算。在这里象§2.3中一样,除了扎德定义的运算外,我们着重讨论了模糊集合的测度运算。我们看到,测度运算是一种比扎德运算更具普遍性的运算。

在§2.6中讨论模糊关系与模糊矩阵的运算。这是模式识别中,尤其是模糊聚类中经常用到的模糊集合的内容。

§ 2.1 距 离

下面将研究模式识别中的最基本的概念:距离与相似度。让我们先讨论距离的概念,它的一般性定义以及模式识别中常用到的几种距离。

§ 2.1.1 距离的一般定义

我们首先介绍拓扑学中关于距离的一般定义。

设有空间(例如模式的特征空间) Ω 及空间中的任意两点 x 与 y 。下面的映射 $d(x, y)$ 称为两点间的距离^{[4]~[7]};

$$d(x, y): \Omega \times \Omega \rightarrow R^+ \quad (2.1-1)$$

如果映射满足下面 3 个条件:

$$d(x, x) = 0, \quad \forall x \in \Omega \quad (2.1-2)$$

$$d(x, y) = d(y, x) \quad \forall x, y \in \Omega \quad (2.1-3)$$

$$d(x, y) \leq d(x, z) + d(z, y), \quad \forall x, y, z \in \Omega \quad (2.1-4)$$

如果(2.1-1)式的值域为 $[0, 1]$, 即

$$d(x, y): \Omega \times \Omega \rightarrow [0, 1] \quad (2.1-5)$$

则 $d(x, y)$ 称为 x 与 y 间的归一化距离, 并用 $d_0(x, y)$ 表示。

§ 2.1.2 常用的几种距离^{[1], [8]}

具体的距离有许多不同的定义, 其中常用的有以下几种:

(1) 欧氏(Euclidean)距离。欧氏距离的定义如下:

$$\begin{aligned} d_e(x, y) &= \|x - y\|^{\frac{1}{2}} \\ &= \left[\sum_{i=1}^d (x_i - y_i)^2 \right]^{\frac{1}{2}} \end{aligned} \quad (2.1-6)$$

其中 d 为矢量 x, y 的维数。

(2) 加权欧氏距离。加权欧氏距离为欧氏距离的推广, 它的定义如下:

$$d_{w_e}(x, y) = \left[\sum_{i=1}^d (w_i (x_i - y_i)^2) \right]^{\frac{1}{2}} \quad (2.1-7)$$

(3) 汉氏(Hamming)距离。汉氏距离的定义如下:

$$d_h(x, y) = \sum_{i=1}^d |x_i - y_i| \quad (2.1-8)$$

(4) 马氏(Mahalanobis)距离。马氏距离的定义如下:

$$\begin{aligned} d_{m_o}(x, y) &= (x - y)^T C^{-1} (x - y) \\ &= \sum_{i=1}^d \sum_{j=1}^d w_{ij} (x_i - y_i) (x_j - y_j) \end{aligned} \quad (2.1-9)$$

式中 C 为一正定矩阵, w_{ij} 为矩阵 C^{-1} 的元素, 马氏距离是加权欧氏距离的推广。当 C 为对角矩阵时, 即化为加权欧氏距离。

(5) 敏氏(Minkowsky)距离。敏氏距离的定义如下:

$$d_{m_i}(x, y) = \left[\sum_{i=1}^d |x_i - y_i|^q \right]^{\frac{1}{q}} \quad (2.1-10)$$

其中 q 为一正数。显然, 当 $q=1$ 时敏氏距离化为汉氏距离, 而当 $q=2$ 时化为欧氏距离。

§ 2.2 相 似 度

我们首先根据传递性的概念独立地给出相似度的定义。然后, 根据相似度与距离是彼此意义相反而又密切相关的两个概念, 再从距离的定义诱导出相似度的定义。

由独立给出的相似度的定义能够导出它与距离的函数关系;反之,由距离给出的相似度的定义也必然会导出一种传递关系。这就是距离与相似度之间的内在联系。

§ 2.2.1 相似度的定义

相似度可独立地定义^{[5], [8], [9]}如下。设 x, y 为特征空间 Ω 中的两个点。下面的映射 $s(x, y)$ 称为 x 与 y 间的相似度:

$$s(x, y): \Omega \times \Omega \rightarrow [0, 1] \quad (2.2-1)$$

如果映射满足下面 3 个条件:

$$s(x, x) = 1, \quad \forall x \in \Omega \quad (2.2-2)$$

$$s(x, y) = s(y, x), \quad \forall x, y \in \Omega \quad (2.2-3)$$

$$s(x, y) \geq s(x, z)s(z, y), \quad \forall x, y, z \in \Omega \quad (2.2-4)$$

(2.2-2)式代表自反性。(2.2-3)式代表对称性。而(2.2-4)式表示相似关系的传递性,它的意义是,两点间的联系比经过任何第 3 点的间接联系更为密切,至少相等。也可以这样描述:在相似性条件下,相联系的环节愈少则联系愈密切。

满足上述定义的相似度也是多种多样的。下面给出几个例子。

(1) 数量积

$$s(x, y) = \begin{cases} 1, & \text{当 } i = j \\ \sum_{i=1}^d x_i y_i / M, & \text{当 } i \neq j \end{cases} \quad (2.2-5)$$

其中 M 为一适当选择的正数,满足 $M \geq \sum_{i=1}^d x_i y_i \quad \forall x, y \in \Omega$ 。

(2) 绝对值减数

$$s(x, y) = \begin{cases} 1, & \text{当 } i = j \\ 1 - C \sum_{i=1}^d |x_i - y_i|, & \text{当 } i \neq j \end{cases} \quad S_{ij} = \begin{cases} 1 & i=j \\ 1 - C \sum_{k=1}^d |x_{ik} - y_{ik}| & i \neq j \end{cases} \quad (2.2-6)$$

其中 C 适当选取,使 $\sum_{i=1}^d |x_i - y_i| \leq 1, \quad \forall x, y \in \Omega$ 。

(3) 最小最大比

$$s(x, y) = \frac{\sum_{i=1}^d \min(x_i, y_i)}{\sum_{i=1}^d \max(x_i, y_i)} \quad (2.2-7)$$

如前所述,相似度也可以从距离的概念引导出来。下面我们用归一化距离 $d_0(x, y)$ 来表示相似度 $s(x, y)$ 。 $s(x, y)$ 可用下列映射由 $d_0(x, y)$ 来定义:

$$f: d_0(x, y) \rightarrow s(x, y) \in [0, 1] \quad (2.2-8)$$

或者表示为

$$s(x, y) = f[d_0(x, y)] \quad (2.2-8a)$$

其中 f 须满足下列条件:

$$\text{若 } d_0(x_1, y_1) > d_0(x_2, y_2), \quad \text{则 } s(x_1, y_1) \leq s(x_2, y_2) \quad (2.2-9)$$