

模式识别与 人工智能

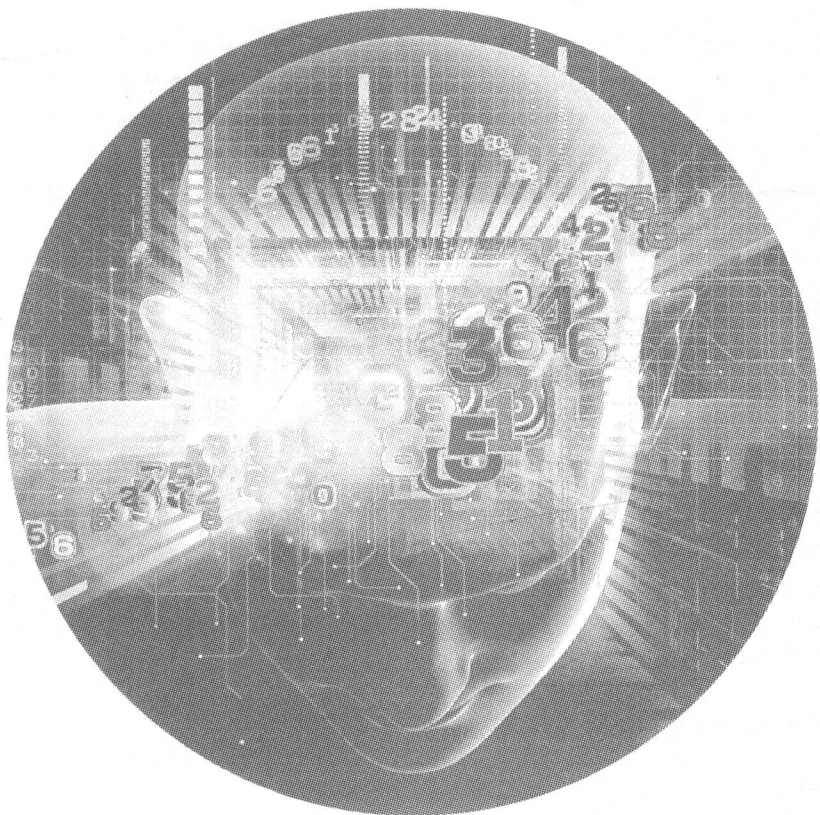
(基于MATLAB)

周润景◎编著

10余年模式识别领域教学与科研成果总结，理论知识系统全面，案例实践性强，所用数据为**酒瓶颜色的RGB**三基色原始数据，数据量大，信息丰富



清华大学出版社



模式识别与人工智能 (基于MATLAB)

周润景◎编著

清华大学出版社
北京

内 容 简 介

本书将模式识别与人工智能理论和实际应用相结合,以酒瓶颜色分类为例,介绍各种算法理论及相应的 MATLAB 实现程序。全书共分为 10 章,内容包括模式识别概述、贝叶斯分类器设计、判别函数分类器设计、聚类分析、模糊聚类分析、神经网络聚类设计、模拟退火算法聚类设计、遗传算法聚类设计、蚁群算法聚类设计、粒子群算法聚类设计,覆盖了各种常用的模式识别技术。

本书可作为高等院校自动化、计算机、电子和通信等专业研究生和高年级本科生的教材,也可作为计算机信息处理、自动控制等相关领域工程技术人员的参考用书。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

模式识别与人工智能:基于 MATLAB/周润景编著. —北京:清华大学出版社,2018
ISBN 978-7-302-48635-0

I. ①模… II. ①周… III. ①模式识别—MATLAB 软件 ②智能计算机—MATLAB 软件
IV. ①O235 ②TP387

中国版本图书馆 CIP 数据核字(2017)第 261766 号

责任编辑:袁金敏
封面设计:刘新新
责任校对:李建庄
责任印制:李红英

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社 总 机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课件下载: <http://www.tup.com.cn>, 010-62795954

印 装 者:清华大学印刷厂

经 销:全国新华书店

开 本:185mm×260mm 印 张:25

字 数:610 千字

版 次:2018 年 7 月第 1 版

印 次:2018 年 7 月第 1 次印刷

印 数:1~2500

定 价:89.00 元

产品编号:076802-01

前言

FOREWORD

随着模式识别技术的迅猛发展,目前该技术已经成为当代高科技研究的重要领域之一,不仅取得了丰富的理论成果,而且其应用范围越来越广泛,几乎遍及各个学科领域,如人工智能、机器人、系统控制、遥感数据分析、生物医学工程、军事目标识别等。由于其在国民经济、国防建设、社会发展的各个方面得到了广泛应用,因而越来越多的人认识到模式识别技术的重要性。

本书以实用性为宗旨,以对酒瓶颜色的分类设计为主,将理论与实践相结合,介绍了各种相关分类器设计。

第1章介绍模式识别的概念、模式识别的方法及其应用。

第2章讨论贝叶斯分类器的设计。首先介绍贝叶斯决策的概念,让读者对贝叶斯理论有所了解,然后介绍基于最小错误率和最小风险的贝叶斯分类器的设计,将理论应用到实践,让读者真正学会运用该算法解决实际问题。

第3章讨论判别函数分类器的设计。判别函数包括线性判别函数和非线性判别函数,本章首先介绍判别函数的相关概念,然后介绍线性判别函数 LMSE 和 Fisher 分类器的设计及非线性判别函数 SVM 分类器的设计。

第4章讨论聚类分析。聚类分析作为最基础的分类方法,涵盖了大量经典的聚类算法及衍生出来的改进算法。本章首先介绍相关理论知识,然后依次介绍 K 均值聚类、 K 均值改进算法、KNN 聚类、PAM 聚类、层次聚类及 ISODATA 分类器设计。

第5章讨论模糊聚类分析。首先介绍模糊逻辑的发展、模糊数学理论、模糊逻辑与模糊推理等一整套模糊控制理论,然后介绍模糊分类器、模糊 C 均值分类器、模糊 ISODATA 分类器及模糊神经网络分类器的设计。

第6章讨论神经网络聚类设计。首先介绍神经网络的概念及其模型等理论知识,然后介绍基于 BP 网络、Hopfield 网络、RBF 网络、GRNN、小波神经网络、自组织竞争网络、SOM 网络、LVQ 网络、PNN、CPN 的分类器设计。

第7章讨论模拟退火算法聚类设计。首先介绍模拟退火算法的基本原理、基本过程,然后介绍其分类器的设计。

第8章介绍遗传算法聚类设计,包括遗传算法原理及遗传算法分类器设计的

详细过程。

第9章介绍蚁群算法聚类设计,包括蚁群算法的基本原理、基于蚁群基本算法的分类器设计和改进的蚁群算法MMAS的分类器设计。

第10章介绍粒子群算法聚类设计,包括粒子群算法的运算过程、进化模型、原理及其模式分类的设计过程。

本书没有像大多数模式识别的书那样讲解烦琐的理论,而是简明扼要地介绍每一种算法的核心,并通过大量的实例介绍模式识别知识。书中针对每一种模式识别算法,按理论基础和实例操作两部分进行介绍。在读者掌握基础理论后,通过实例可以了解算法的实现思路和方法;进一步掌握核心代码编写,就可以很快掌握模式识别技术。

本书内容来自作者的科研与教学实践。读者在学会各种理论和方法后,可将书中的不同算法加以改造应用于自己的实际工作。

本书第1~3章由李楠编写,其余由周润景完成并统稿、定稿。参加本书编写的还有邵盟、南志贤、刘波、李艳、邵绪晨、冯震、崔婧、任自鑫、谢亚楠、祖晓玮、张赫、丁岩、井探亮、邢婧、陈萌。

在本书的编写过程中,作者力求完美,但由于水平有限,书中难免有不足之处,敬请指正。

作者

2018年3月

目录

CONTENTS

第 1 章 模式识别概述	1
1.1 模式识别的基本概念	1
1.1.1 模式的描述方法	1
1.1.2 模式识别系统	2
1.2 模式识别的基本方法	3
1.3 模式识别的应用	4
习题	6
第 2 章 贝叶斯分类器设计	7
2.1 贝叶斯决策及贝叶斯公式	7
2.1.1 贝叶斯决策简介	7
2.1.2 贝叶斯公式	7
2.2 基于最小错误率的贝叶斯决策	8
2.2.1 基于最小错误率的贝叶斯决策理论	8
2.2.2 最小错误率贝叶斯分类的计算过程	9
2.2.3 最小错误率贝叶斯分类的 MATLAB 实现	13
2.2.4 结论	20
2.3 最小风险贝叶斯决策	21
2.3.1 最小风险贝叶斯决策理论	21
2.3.2 最小错误率与最小风险的贝叶斯决策比较	22
2.3.3 贝叶斯算法的计算过程	23
2.3.4 最小风险贝叶斯分类的 MATLAB 实现	23
2.3.5 结论	32
习题	33
第 3 章 判别函数分类器设计	34
3.1 判别函数简介	34
3.2 线性判别函数	35

3.3	线性判别函数的实现	36
3.4	基于 LMSE 的分类器设计	36
3.4.1	LMSE 分类法简介	36
3.4.2	LMSE 算法原理	37
3.4.3	LMSE 算法步骤	37
3.4.4	LMSE 算法的 MATLAB 实现	38
3.4.5	结论	47
3.5	基于 Fisher 的分类器设计	48
3.5.1	Fisher 判别法简介	48
3.5.2	Fisher 判别法的原理	48
3.5.3	Fisher 分类器设计	48
3.5.4	Fisher 算法的 MATLAB 实现	51
3.5.5	识别待测样本类别	53
3.5.6	结论	71
3.6	基于支持向量机的分类法	71
3.6.1	支持向量机简介	71
3.6.2	支持向量机基本思想	71
3.6.3	支持向量机的几个主要优点	72
3.6.4	训练集为非线性情况	72
3.6.5	核函数	73
3.6.6	多类分类问题	73
3.6.7	基于 SVM 的 MATLAB 实现	74
3.6.8	结论	78
	习题	78
第 4 章	聚类分析	79
4.1	聚类分析	79
4.1.1	聚类的定义	79
4.1.2	聚类准则	80
4.1.3	基于试探法的聚类设计	80
4.2	数据聚类——K 均值聚类	81
4.2.1	K 均值聚类简介	81
4.2.2	K 均值聚类原理	81
4.2.3	K 均值算法的优缺点	83
4.2.4	K 均值聚类的 MATLAB 实现	84
4.2.5	待聚类样本的分类结果	87
4.2.6	结论	88
4.3	数据聚类——基于取样思想的改进 K 均值聚类	89
4.3.1	K 均值改进算法的思想	90

4.3.2	基于取样思想的改进 K 均值算法 MATLAB 实现	91
4.3.3	结论	96
4.4	数据聚类——K-近邻法聚类	96
4.4.1	K-近邻法简介	96
4.4.2	K-近邻法的算法研究	96
4.4.3	K-近邻法数据分类器的 MATLAB 实现	98
4.4.4	结论	103
4.5	数据聚类——PAM 聚类	103
4.5.1	PAM 算法简介	103
4.5.2	PAM 算法的主要流程	103
4.5.3	PAM 算法的 MATLAB 实现	105
4.5.4	PAM 算法的特点	112
4.5.5	K 均值算法和 PAM 算法分析比较	112
4.5.6	结论	112
4.6	数据聚类——层次聚类	114
4.6.1	层次聚类方法简介	114
4.6.2	凝聚的和分裂的层次聚类	114
4.6.3	簇间距离度量方法	115
4.6.4	层次聚类方法存在的不足	116
4.6.5	层次聚类的 MATLAB 实现	116
4.6.6	结论	121
4.7	数据聚类——ISODATA 算法概述	122
4.7.1	ISODATA 算法应用背景	122
4.7.2	ISODATA 算法的 MATLAB 实现	124
4.7.3	结论	136
	习题	136
第 5 章 模糊聚类分析		137
5.1	模糊逻辑的发展	137
5.2	模糊集合	138
5.2.1	由经典集合到模糊集合	138
5.2.2	模糊集合的基本概念	139
5.2.3	隶属度函数	142
5.3	模糊集合的运算	144
5.3.1	模糊集合的基本运算	144
5.3.2	模糊集合的基本运算规律	146
5.3.3	模糊集合与经典集合的联系	147
5.4	模糊关系与模糊关系的合成	149
5.4.1	模糊关系的基本概念	149

5.4.2	模糊关系的合成	152
5.4.3	模糊关系的性质	154
5.4.4	模糊变换	155
5.5	模糊逻辑及模糊推理	156
5.5.1	模糊逻辑技术	157
5.5.2	语言控制策略	158
5.5.3	模糊语言变量	159
5.5.4	模糊命题与模糊条件语句	159
5.5.5	判断与推理	161
5.5.6	模糊推理	162
5.6	数据聚类——模糊聚类	167
5.6.1	模糊聚类的应用背景	167
5.6.2	基于 MATLAB 的 GUI 工具的模糊算法构建——数据模糊化	168
5.6.3	基于 MATLAB 的 GUI 工具的模糊算法构建——FIS 实现	171
5.6.4	系统结果分析	174
5.6.5	结论	176
5.7	数据聚类——模糊 C 均值聚类	176
5.7.1	模糊 C 均值聚类的应用背景	176
5.7.2	模糊 C 均值算法	177
5.7.3	模糊 C 均值聚类的 MATLAB 实现	178
5.7.4	模糊 C 均值聚类结果分析	180
5.7.5	结论	182
5.8	数据聚类——模糊 ISODATA 聚类	182
5.8.1	模糊 ISODATA 聚类的应用背景	182
5.8.2	模糊 ISODATA 算法的基本原理	182
5.8.3	模糊 ISODATA 算法的基本步骤	183
5.8.4	模糊 ISODATA 算法的 MATLAB 程序实现	186
5.8.5	结论	199
5.9	模糊神经网络	199
5.9.1	模糊神经网络的应用背景	199
5.9.2	模糊神经网络算法的原理	200
5.9.3	模糊神经网络分类器的 MATLAB 实现	203
5.9.4	结论	214
	习题	215
第 6 章	神经网络聚类设计	216
6.1	什么是神经网络	216
6.1.1	神经网络的发展历程	216
6.1.2	生物神经系统的结构及冲动的传递过程	218

6.1.3	人工神经网络的定义	220
6.2	人工神经网络模型	221
6.2.1	人工神经元的基本模型	221
6.2.2	人工神经网络基本构架	222
6.2.3	人工神经网络的工作过程	224
6.2.4	人工神经网络的特点	225
6.3	前馈神经网络	225
6.3.1	感知器网络	227
6.3.2	BP 网络	229
6.3.3	BP 网络的建立及执行	231
6.3.4	BP 网络分类器的 MATLAB 实现	233
6.3.5	BP 网络的其他学习算法的应用	239
6.4	反馈神经网络	248
6.4.1	离散 Hopfield 网络的结构	248
6.4.2	离散 Hopfield 网络的工作方式	249
6.4.3	离散 Hopfield 网络的稳定性和吸引子	250
6.4.4	离散 Hopfield 网络的连接权设计	251
6.4.5	离散 Hopfield 网络分类器的 MATLAB 实现	252
6.4.6	结论	259
6.5	径向基函数	260
6.5.1	径向基函数的网络结构及工作方式	260
6.5.2	径向基函数网络的特点及作用	262
6.5.3	径向基函数网络参数选择	262
6.5.4	RBF 网络分类器的 MATLAB 实现	262
6.5.5	结论	269
6.6	广义回归神经网络	270
6.6.1	GRNN 的结构	270
6.6.2	GRNN 的理论基础	271
6.6.3	GRNN 的特点及作用	272
6.6.4	GRNN 分类器的 MATLAB 实现	272
6.6.5	结论	277
6.7	小波神经网络	277
6.7.1	小波神经网络的基本结构	277
6.7.2	小波神经网络的训练算法	279
6.7.3	小波神经网络结构设计	280
6.7.4	小波神经网络分类器的 MATLAB 实现	281
6.7.5	结论	292
6.8	其他形式的神经网络	292
6.8.1	竞争型人工神经网络——自组织竞争	292

6.8.2	竞争型人工神经网络——自组织特征映射神经网络	296
6.8.3	竞争型人工神经网络——学习向量量化神经网络	299
6.8.4	概率神经网络	305
6.8.5	CPN神经网络分类器的 MATLAB 实现	311
习题		318
第 7 章	模拟退火算法聚类设计	319
7.1	模拟退火算法简介	319
7.1.1	物理退火过程	319
7.1.2	Metropolis 准则	320
7.1.3	模拟退火算法的基本原理	320
7.1.4	模拟退火算法的组成	321
7.1.5	模拟退火算法新解的产生和接受	321
7.1.6	模拟退火算法的基本过程	322
7.1.7	模拟退火算法的参数控制问题	322
7.2	基于模拟退火思想的聚类算法	323
7.2.1	K 均值算法的局限性	323
7.2.2	基于模拟退火思想的改进 K 均值聚类算法	324
7.2.3	几个重要参数的选择	324
7.3	算法实现	325
7.3.1	实现步骤	325
7.3.2	模拟退火实现模式分类的 MATLAB 程序	326
7.4	结论	334
习题		335
第 8 章	遗传算法聚类设计	336
8.1	遗传算法简介	336
8.2	遗传算法原理	337
8.2.1	遗传算法的基本术语	337
8.2.2	遗传算法进行问题求解的过程	338
8.2.3	遗传算法的优缺点	338
8.2.4	遗传算法的基本要素	338
8.3	算法实现	341
8.3.1	种群初始化	341
8.3.2	适应度函数的设计	343
8.3.3	选择操作	345
8.3.4	交叉操作	346
8.3.5	变异操作	347
8.3.6	完整程序及仿真结果	348

8.4 结论	355
习题	355
第 9 章 蚁群算法聚类设计	356
9.1 蚁群算法简介	356
9.2 蚁群算法原理	357
9.2.1 基本蚁群算法原理	357
9.2.2 模型建立	359
9.2.3 蚁群算法的特点	362
9.3 基本蚁群算法的实现	363
9.4 算法改进	369
9.4.1 MMAS 算法简介	369
9.4.2 完整程序及仿真结果	371
9.5 结论	376
习题	376
第 10 章 粒子群算法聚类设计	377
10.1 粒子群算法简介	377
10.2 经典的粒子群算法的运算过程	377
10.3 两种基本的进化模型	378
10.4 改进的粒子群优化算法	379
10.4.1 粒子群优化算法原理	379
10.4.2 粒子群优化算法的基本流程	380
10.5 粒子群算法与其他算法的比较	381
10.6 粒子群算法分类器的 MATLAB 实现	382
10.6.1 设定参数	382
10.6.2 初始化	382
10.6.3 完整程序及仿真结果	382
10.7 结论	387
习题	387
参考文献	388

第1章

模式识别概述

1.1 模式识别的基本概念

模式识别(pattern recognition)也称机器识别,就是通过计算机用数学技术方法来进行模式的自动处理和判读。该学科的主要任务是利用计算机进行模拟人的识别能力,提出识别具体客体的基本理论与实用技术。其最基本的方法是计算,即计算要识别的事物与已知的标准事物的相似程度,从而让机器能判别事物。因此,找到度量不同事物差异的有效方法是研究的关键。随着计算机技术的发展,人类有可能研究复杂的信息处理过程。我们把环境与客体统称为“模式”。信息处理过程的一个重要形式是生命体对环境及客体的识别。对人类来说,特别重要的是对光学信息(通过视觉器官来获得)和声学信息(通过听觉器官来获得)的识别。这是模式识别的两个重要方面。模式识别是确定一个样本的类别属性的过程,即把某一样本归属于多个类型中的某个类型。例如,数据分类,结果就是将待分类数据按属性分类;指纹识别,就是使用指纹的总体特征如纹形、三角点等来进行分类,再用局部特征如位置和方向等来进行识别用户身份;还有语音识别、生物认证、字符识别等。

1.1.1 模式的描述方法

对于模式的描述方法有两种:定量描述和结构性描述。其中,定量描述通过对事物的属性进行度量,用一组数据来描述模式;结构性描述是对事物所包含的成分进行分析,用一组基元来描述模式。

针对定量描述方法,一个具体的研究对象称为样品。对于一个样品来说,必须确定一些与识别有关的因素作为研究的根据,每一个因素称为一个特征。模式用样品的一组数据来描述。模式的特征集一般可以用特征向量表示。特征向量中的每个元素称作特征。假设一个样品 \mathbf{X} 有 n 个特征,若用小写字母 x 来表示特征,则可以把 \mathbf{X} 看作一个 n 维列向量,该向量 \mathbf{X} 称为特征向量,记作:

$$\mathbf{X} = (x_1, x_2, \dots, x_n)^T \quad (1-1)$$

模式识别问题就是根据 \mathbf{X} 的 n 个特征来判别模式 \mathbf{X} 属于 $\omega_1, \omega_2, \dots, \omega_M$ 类中的哪一类。其目的是在特征空间和解释空间之间建立一种特殊的对应关系。其中,特征空间由特征向量构成,包括模式的度量、属性等,通过对具体对象进行观测得到;解释空间由所属模式类别的集合构成。

1.1.2 模式识别系统

模式识别的关键是如何利用计算机进行模式识别,并对样本进行分类。执行模式识别的基于计算机的系统(可以是各种有计算能力的处理器系统)称为模式识别系统。该系统具有两种工作方式,训练方式和识别方式。

图 1-1 是一个典型的模式识别系统,由数据获取、预处理、特征提取、分类决策及分类器设计等部分组成。该系统各个组成部分的功能概括如下。

位置类别模式的分类

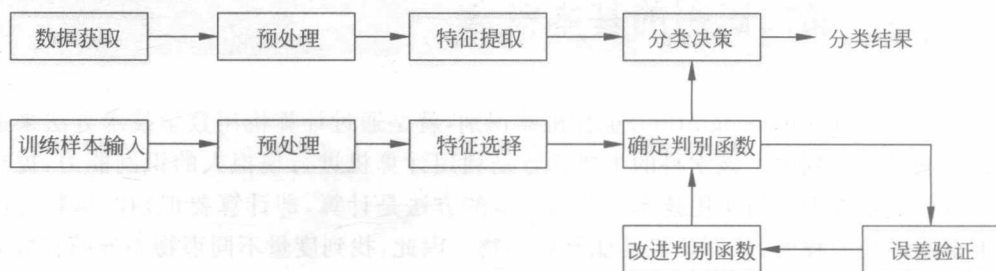


图 1-1 模式识别系统

(1) 数据获取：一般情况下,获取的信息类型有以下几种。

- 一维波形：心电图、脑电波、声波、震动波形等。
- 二维图像：文字、地图、照片等。
- 物理参量：体温、化验数据、温度、压力、电流、电压等。

(2) 预处理：对由于信息获取装置或其他因素所造成的信息退化现象进行复原、去噪,加强有用信息。

(3) 特征提取：由信息获取部分获得的原始信息,其数据量一般相当大。为了有效地实现分类识别,应对经过预处理的信息进行选择或变换,得到最能反映分类本质的特征,构成特征向量。其目的是将维数较高的模式空间转换为维数较低的特征空间。

(4) 分类决策：在特征空间中用模式识别方法(由分类器设计确定的分类判别规则)对待识别模式进行分类判别,将其归为某一类别,输出分类结果。这一过程对应于特征空间向类别空间的转换。

(5) 分类器设计：为了把待识别模式分配到各自的模式类中,必须设计出一套分类判别规则。基本做法是收集一定数量的样本作为训练集,在此基础上确定判别函数、改进判别函数和误差检验。

其中分类器设计在训练方式下完成,利用样本进行训练,确定分类器的具体参数,而分类决策在识别方式下起作用,对待识别的样本进行分类决策。

1.2 模式识别的基本方法

模式识别作为人工智能的一个重要应用领域,目前得到了飞速发展。针对不同的对象和不同的目的,可以使用不同的模式识别理论或方法。目前,基本的技术方法有如下四种。

1. 统计模式识别

统计模式识别是首先根据待识别对象所包含的原始数据信息,从中提取出若干能够反映该类对象某方面性质的相应特征参数,并根据识别的实际需要从中选择一些参数的组合作为一个特征向量,根据某种相似性测度,设计一个能够对该向量组表示的模式进行区分的分类器,就可把特征向量相似的对象分为一类。

2. 结构模式识别

当需要对待识别对象的各部分之间的联系进行精确识别时,就需要使用结构模式识别方法。结构模式识别是根据识别对象的结构特征,将复杂的模式结构先通过分解划分为多个相对更简单且更容易区分的子模式,若得到的子模式仍有识别难度,则继续对其进行分解,直到最终得到的子模式具有容易表示且容易识别的结构为止,通过这些子模式就可以复原原先比较复杂的模式结构。

3. 模糊模式识别

模糊集理论认为,模糊集合中的一个元素,可以不是百分之百地确定属于该集合,而是可以以一定的比例属于该集合,不像传统集合理论中某元素要么属于要么不属于该集合的定义方式,更符合现实当中许多模糊的实际问题,描述起来更加简单合理。在用机器模拟人类智能时模糊数学能更好地描述现实当中具有模糊性的问题,进而更好地进行处理。模糊模式识别就以模糊集理论为基础,根据一定的判定要求建立合适的隶属度函数来对识别对象进行分类。

正是因为模糊模式识别能够很好地解决现实当中许多具有模糊性的概念,使其成为一种重要的模式识别方法。在进行模糊识别时,也需要建立一个识别系统,需要对实际的识别对象的特征参数按照一定的比例进行分类,这些比例往往是根据人为的经验作为参考值,只要符合认可的经验认识就行,之后建立能够处理模糊性问题的分类器对不同类别的特征向量进行判别。

4. 神经网络

神经网络分为四种类型,即向前型、反馈型、随机型和自组织竞争。神经网络作为模式识别技术当中最重要的方法之一,相对于传统的模式识别方法,它具有如下优势。

(1) 神经网络属于学习及自适应能力很强的方法。

(2) 对于任意给定的函数,神经网络都能够无限逼近,这是因为在分类的整个过程中,神经网络通过调整权值不断地明确分类所依据的精确关系。

(3) 神经网络属于非线性模型,这使得它能够灵活地模拟现实世界中的数据之间的复杂关系模型。

1.3 模式识别的应用

模式识别是一种智能的活动,它包括分析和判断两个过程。随着计算机性能的提高、互联网技术的迅速发展、理论研究的深入以及和其他研究领域(如机器学习、数学、统计学、生物学等)的促进、融合,现在模式识别技术已经被普遍地应用于生物学(自动细胞学、染色体特性研究、遗传研究)、天文学(天文望远镜图像分析、自动光谱学)、经济学(股票交易预测、企业行为分析)、医学(心电图分析、脑电图分析、医学图像分析)、工程(产品缺陷检测、特征识别、语音识别、自动导航系统、污染分析)、军事(航空摄像分析、雷达和声呐信号检测和分类、自动目标识别)、安全(指纹识别、人脸识别、监视和报警系统)等领域。

本书基于 MATLAB 实现不同模式识别方法的应用,以酒瓶颜色的分类为主。由不同材料制成的不同颜色和项目的玻璃必须被分类,以获得高质量的可回收原料。在玻璃回收厂,玻璃瓶被分类放置到容器中,然后一起熔化处理产生新的玻璃。在这个过程中,玻璃瓶被分选到不同容器这一步是很重要的,因为生产玻璃的不同客户需要不同的颜色,并且回收的瓶子颜色混杂,对瓶子的分类没有预先设定。在这种情况下,多数操作员根据经验手工分选回收的瓶子,以使生产的玻璃达到期望的颜色。这既费时又费力,而通过模式识别的方法分类,则既解放了生产力,又提高了效率。表 1-1 为 59 组三元色数据,其中前 29 组作为训练数据,后 30 组作为测试数据。前 29 组数据的类别已经给出。

表 1-1 三元色数据

序号	A	B	C	所属类别
1	1739.94	1675.15	2395.96	3
2	373.3	3087.05	2429.47	4
3	1756.77	1652	1514.98	3
4	864.45	1647.31	2665.9	1
5	222.85	3059.54	2002.33	4
6	877.88	2031.66	3071.18	1
7	1803.58	1583.12	2163.05	3
8	2352.12	2557.04	1411.53	2
9	401.3	3259.94	2150.98	4
10	363.34	3477.95	2462.86	4
11	1571.17	1731.04	1735.33	3
12	104.8	3389.83	2421.83	4
13	499.85	3305.75	2196.22	4
14	2297.28	3340.14	535.62	2
15	2092.62	3177.21	584.32	2

续表

序 号	A	B	C	所属类别
16	1418.79	1775.89	2772.9	1
17	1845.59	1918.81	2226.49	3
18	2205.36	3243.74	1202.69	2
19	2949.16	3244.44	662.42	2
20	1692.62	1867.5	2108.97	3
21	1680.67	1575.78	1725.1	3
22	2802.88	3017.11	1984.98	2
23	172.78	3084.49	2328.65	4
24	2063.54	3199.76	1257.21	2
25	1449.58	1641.58	3405.12	1
26	1651.52	1713.28	1570.38	3
27	341.59	3076.62	2438.63	4
28	291.02	3095.68	2088.95	4
29	237.63	3077.78	2251.96	4
30	1702.8	1639.79	2068.74	—
31	1877.93	1860.96	1975.3	—
32	867.81	2334.68	2535.1	—
33	1831.49	1713.11	1604.68	—
34	460.69	3274.77	2172.99	—
35	2374.98	3346.98	975.31	—
36	2271.89	3482.97	946.7	—
37	1783.64	1597.99	2261.31	—
38	198.83	3250.45	2445.08	—
39	1494.63	2072.59	2550.51	—
40	1597.03	1921.52	2126.76	—
41	1598.93	1921.08	1623.33	—
42	1243.13	1814.07	3441.07	—
43	2336.31	2640.26	1599.63	—
44	354	3300.12	2373.61	—
45	2144.47	2501.62	591.51	—
46	426.31	3105.29	2057.8	—
47	1507.13	1556.89	1954.51	—
48	343.07	3271.72	2036.94	—
49	2201.94	3196.22	935.53	—
50	2232.43	3077.87	1298.87	—
51	1580.1	1752.07	2463.04	—
52	1962.4	1594.97	1835.95	—
53	1495.18	1957.44	3498.02	—
54	1125.17	1594.39	2937.73	—
55	24.22	3447.31	2145.01	—
56	1269.07	1910.72	2701.97	—
57	1802.07	1725.81	1966.35	—
58	1817.36	1927.4	2328.79	—
59	1860.45	1782.88	1875.13	—