

MOSHI SHIBIE YU JIQI XUEXI JISHU

模式识别

与机器学习技术

■ 牟少敏 时爱菊 著



冶金工业出版社

www.cnmp.com.cn

模式识别与机器学习技术

牟少敏 时爱菊 著



北京
冶金工业出版社
2019

内 容 提 要

模式识别与机器学习是计算机科学与技术的重要研究内容之一。

本书首先讲解了贝叶斯分类、支持向量机和人工神经网络等常用的机器学习算法，并对结构数据的核函数和增量支持向量机算法进行了全面综述，讲解了深度学习最新的模型和计算机视觉的基本知识。以农业为应用场景，结合作者的科研工作，详细介绍了基于卷积神经网络的树木识别和基于对抗生成网络的玉米病害图像生成的实际应用案例。最后介绍了模式识别与机器学习涉及的数学基础。书中配有模式识别与机器学习相应算法的 Python 源代码。

本书适合计算机科学与技术、数据科学与技术等相关专业的研究生和本科生使用，也可供从事农业大数据研究工作等相关人员参考。

图书在版编目(CIP)数据

模式识别与机器学习技术 / 牟少敏, 时爱菊著. —北京:
冶金工业出版社, 2019. 6

ISBN 978-7-5024-8130-8

I. ①模… II. ①牟… ②时… III. ①模式识别—研究
②机器学习—研究 IV. ①TP391.4 ②TP181

中国版本图书馆 CIP 数据核字(2019)第 122402 号

出 版 人 谭学余

地 址 北京市东城区嵩祝院北巷 39 号 邮编 100009 电话 (010)64027926

网 址 www.cnmp.com.cn 电子信箱 yjcs@cnmp.com.cn

责任编辑 卢 敏 美术编辑 吕欣童 版式设计 孙跃红

责任校对 卿文春 责任印制 李玉山

ISBN 978-7-5024-8130-8

冶金工业出版社出版发行; 各地新华书店经销; 三河市双峰印刷装订有限公司印刷
2019 年 6 月第 1 版, 2019 年 6 月第 1 次印刷

169mm×239mm; 9.75 印张; 185 千字; 141 页

46.00 元

冶金工业出版社 投稿电话 (010)64027932 投稿信箱 tougao@cnmp.com.cn

冶金工业出版社营销中心 电话 (010)64044283 传真 (010)64027893

冶金工业出版社天猫旗舰店 yjgycbs.tmall.com

(本书如有印装质量问题, 本社营销中心负责退换)

前 言

模式识别和机器学习是涉及计算机科学与技术、概率论与数理统计、图像处理技术、最优化和认知科学等多领域交叉的学科，是人工智能的重要组成部分，广泛应用于计算机视觉和语音识别等领域。

本书是作者在长期从事模式识别和机器学习研究及其在农业中应用的基础上编写的。全书共 11 章。第 1 章简要介绍了模式识别与机器学习的基础知识；第 2 章主要介绍了判别函数；第 3 章主要介绍了特征提取与特征选择；第 4 章、第 5 章和第 6 章深入浅出地介绍了机器学习和模式识别的基本算法；第 7 章主要介绍了人工神经网络和最新的几种深度神经网络；第 8 章结合作者的科研工作，以农业为背景，详细介绍了机器学习与模式识别的应用；第 9 章和第 10 章主要介绍了涉及模式识别和机器学习的数学基础知识；第 11 章介绍最优化理论与信息论。

时爱菊、王秀美、曹旨昊、苏婷婷和董萌萍参加了部分章节的编写和修改工作，在此表示衷心的感谢。

书中不妥之处敬请读者批评指正。

牟少敏

2019 年 3 月于山东农业大学

目 录

1 模式识别与机器学习基本概念	1
1.1 模式识别	1
1.1.1 基本概念	1
1.1.2 模式识别应用	4
1.1.3 模式识别系统	6
1.1.4 模式识别基本方法	7
1.1.5 模式识别基本问题	7
1.1.6 模式识别基本准则	8
1.2 机器学习	9
1.2.1 简介	9
1.2.2 机器学习的分类	9
1.2.3 深度学习	9
1.3 机器学习与模式识别算法评价指标	10
1.3.1 查全率与准确率	10
1.3.2 交叉验证	10
1.3.3 混淆矩阵	11
1.4 K 近邻算法	12
1.5 顶级会议和期刊	13
1.5.1 顶级会议	13
1.5.2 顶级期刊	14
1.5.3 国内重要期刊	17
2 判别函数	18
2.1 判别函数	18
2.1.1 线性判别函数	18
2.1.2 非线性判别函数	23
2.2 Fisher 线性判别函数	24
3 特征提取与选择	26
3.1 简介	26

3.2	基本概念	26
3.3	类别可分性判据	27
3.3.1	基于距离的可分性判据	27
3.3.2	基于概率分布的可分性判据	28
3.4	主成分分析	29
3.4.1	简介	29
3.4.2	基本原理	30
3.4.3	具体步骤	31
3.4.4	应用举例	32
3.4.5	核主成分分析法	33
3.5	图像特征	33
3.5.1	颜色特征	33
3.5.2	纹理特征	34
3.5.3	形状特征	35
4	贝叶斯分类	36
4.1	简介	36
4.1.1	相关统计概念	36
4.1.2	贝叶斯定理	37
4.2	贝叶斯分类	38
4.3	朴素贝叶斯分类	39
4.3.1	简介	39
4.3.2	基本原理	39
4.3.3	分类举例	39
4.4	贝叶斯网络	41
4.4.1	结构形式 1	41
4.4.2	结构形式 2	42
4.4.3	结构形式 3	42
4.4.4	举例	43
4.5	基于 Python 的朴素贝叶斯分类实现	44
5	聚类分析	46
5.1	聚类概念	46
5.2	聚类算法分类	46
5.3	相似性度量	47
5.4	聚类准则	49

5.5 C 均值聚类	49
5.5.1 简介	49
5.5.2 基本原理	49
5.5.3 C 均值算法的优缺点	50
5.6 模糊模式识别	50
5.6.1 简介	50
5.6.2 相关概念	50
5.6.3 模糊集合的运算	53
5.6.4 模糊 C 均值聚类	54
6 支持向量机	56
6.1 简介	56
6.2 VC 维与结构风险最小化	56
6.2.1 VC 维	56
6.2.2 经验风险最小化	56
6.3 基本原理	57
6.4 核函数	57
6.4.1 简介	57
6.4.2 结构数据的核函数	58
6.5 多类支持向量机	59
6.5.1 一对一支持向量机	60
6.5.2 一对多支持向量机	60
6.5.3 二叉树支持向量机	61
6.5.4 有向无环图支持向量机	62
6.6 单类支持向量机	62
6.6.1 基于密度的单类支持向量机	62
6.6.2 基于边界的单类支持向量机	63
6.7 基于增量学习的支持向量机	63
6.7.1 错误驱动法	63
6.7.2 固定划分法	63
6.7.3 过间隔法	63
6.7.4 错误驱动法+过间隔法	64
6.8 基于 Python 的支持向量机的实现	64
7 人工神经网络	67
7.1 简介	67

7.1.1	发展过程	67
7.1.2	人工神经元	68
7.1.3	人工神经网络特点	69
7.1.4	人工神经网络分类	69
7.1.5	激活函数	69
7.2	感知器	72
7.2.1	单层感知器	72
7.2.2	多层感知器	73
7.3	BP 神经网络	73
7.3.1	简介	73
7.3.2	基本原理	73
7.3.3	算法步骤	74
7.3.4	BP 网络参数选择	75
7.4	RBF 神经网络	75
7.5	Hopfield 网络	77
7.5.1	离散 Hopfield 神经网络	77
7.5.2	连续型 Hopfield 神经网络	78
7.6	最新的几种深度学习网络	78
7.6.1	受限玻尔兹曼机	79
7.6.2	卷积神经网络	79
7.6.3	深度信念网络	81
7.6.4	循环神经网络	81
7.6.5	生成对抗网络	82
8	机器学习与模式识别应用——以农业为例	83
8.1	简介	83
8.2	计算机视觉	83
8.2.1	定义	83
8.2.2	计算机视觉研究的主要内容	84
8.2.3	机器视觉系统	85
8.2.4	计算机视觉的开源库和编程工具	85
8.3	模式识别与机器学习在农业中应用研究	85
8.4	Python 的安装和搭建	86
8.5	基于深度学习的泰山树木分类识别	88
8.5.1	系统开发环境	88
8.5.2	树木图像采集	88

8.5.3	系统设计	89
8.5.4	源码实现	90
8.6	基于生成式对抗网络的小麦叶部病害图像生成	93
8.6.1	简介	93
8.6.2	图像数据	93
8.6.3	实验过程	94
9	线性代数	97
9.1	标量与向量	97
9.1.1	定义	97
9.1.2	常见的向量	97
9.2	向量运算	97
9.2.1	向量的加法	97
9.2.2	向量的减法	98
9.2.3	向量的标量乘法	98
9.3	线性方程组和行列式	98
9.4	矩阵	100
9.4.1	基本概念	100
9.4.2	特殊矩阵	100
9.4.3	矩阵运算	101
9.4.4	初等变换	104
9.4.5	矩阵的秩	104
9.4.6	矩阵与向量的关系	105
9.5	特征值和特征向量	106
9.5.1	定义	106
9.5.2	基本性质	106
9.6	矩阵的分解	107
9.6.1	三角分解	107
9.6.2	QR 分解	108
9.6.3	满秩分解	108
9.6.4	奇异值分解	108
9.7	张量	109
10	概率论与数理统计	110
10.1	随机事件	110
10.1.1	随机现象	110

10.1.2	随机事件	110
10.1.3	事件关系	110
10.1.4	事件运算	111
10.2	概率	111
10.2.1	定义	111
10.2.2	条件概率	112
10.2.3	先验概率和后验概率	113
10.2.4	事件独立性	113
10.3	随机变量及其分布	114
10.3.1	随机变量的定义	114
10.3.2	分布函数	114
10.4	随机变量的数字特征	115
10.4.1	数学期望	115
10.4.2	方差及标准差	116
10.4.3	分位数	116
10.4.4	协方差与协方差矩阵	116
10.4.5	变异系数	117
10.4.6	相关系数	117
10.5	常见的分布	118
10.5.1	二项分布	118
10.5.2	泊松分布	118
10.5.3	正态分布	119
10.5.4	均匀分布	120
10.5.5	指数分布	120
10.6	多元随机变量及其分布	120
10.6.1	二元随机变量及其分布函数	120
10.6.2	边缘分布	123
10.6.3	条件分布	124
10.7	数理统计	124
10.7.1	统计量及其分布	124
10.7.2	参数估计	126
10.7.3	单因素方差分析	128
10.7.4	回归分析与最小二乘法	130
11	最优化理论与信息论	132
11.1	最优化问题描述	132

11.2 对偶理论	133
11.2.1 等式约束的优化问题的求法	133
11.2.2 不等式约束的优化问题的求法	133
11.3 常用的优化算法	134
11.3.1 最速下降法	134
11.3.2 牛顿法	135
11.4 信息论基础	136
11.4.1 信息概念	136
11.4.2 自信息和信息熵	136
参考文献	138

1 模式识别与机器学习基本概念

1.1 模式识别

模式识别的概念最早于 20 世纪 20 年代提出，到 60 年代初逐步发展成为一门综合性学科。其研究内容涉及数学、机器学习、图像处理、计算机视觉和人工智能等多个领域。30 年代 Fisher 提出的统计分类理论，为统计模式识别打下理论基础；50 年代美籍华人傅京孙教授提出了结构模式识别；60 年代 L. A. Zadeh 提出模糊集合理论，为模糊模式识别奠定了理论基础；80 年代人工神经网络和 90 年代支持向量机，成为模式识别的主要方法。2016 年深度学习的提出又为模式识别的广泛应用奠定了良好的基础。

人类可以很容易地通过对事物的感知识别出图 1-1 场景中的对象——台历、电脑、水杯和它们的位置关系；那么人是如何实现这一功能的呢？其学习机制和学习方法又是什么呢？如何让机器模仿人类实现识别的功能呢？模式识别与机器学习的任务就是让计算机模仿人类的识别机理进行检测和识别的过程。



图 1-1 场景

1.1.1 基本概念

1.1.1.1 模式与模式类 (Pattern and Pattern class)

(1) 模式：时间和空间中客观存在的物体、行为和关系等称为模式。模式

具有可观察性、可度量性、可区分性和相似性等特点。

模式举例如下：

连续可枚举： $\{\text{Red, Green, Blue}\}$ ；

无限可列： $\{\text{所有的负整数}\}$ ；

有限连续：温度 $[10, 100]$ ；

无限不连续： $\{\text{所有的正整数}\}$ 。

(2) 模式类：模式类是一类事物的总称，这类事物是具有某些共同特性的模式的集合。

模式是一个具体的事物，模式类则是具体事物的抽象。

1.1.1.2 模式识别 (Pattern Recognition)

现实世界中，模式识别无处不在，无时不有。如：“物以类聚，人以群分”和“近朱者赤，近墨者黑”等都是模式识别的应用。模式识别指的是机器的模式识别，即让机器模仿人类识别的过程进行模式的识别。模式识别的目的是利用计算机对物体、行为和关系等模式进行分类，在错误概率最小或风险最低的情况下，使识别的结果尽量与客观现实相一致。简单地说，识别就是将某个具体的事物尽可能正确地归到某一类别。

1.1.1.3 特征 (Feature)

在模式识别中，被采集的每个对象称为样本。相关的每个因子或属性称为样本的特征，其量化的数据类型有数值型和非数值型两种。非数值型特征需要转化为相应的数值特征。样本的特征构成了样本特征空间，每个样本是特征空间中的一个点。

例如：对某种昆虫进行分类时，需要选择和提取它的各种特征并进行量化。例如：昆虫的颜色和大小等都可以作为识别的重要特征。

1.1.1.4 特征向量与特征空间 (Feature Vector and Feature Space)

如果对象 X 有 n 个特征量测量值，则可以把 X 看做一个 n 维列向量，表示为： $X = (x_1, x_2, \dots, x_i, \dots, x_n)^T$, ($i = 1, 2, \dots, n$)，即构成一个 n 维的特征向量。

各种不同取值的特征矢量的全体构成了多维特征空间。注：特征矢量就是特征空间中的一个点。一般用 X^n 或 R^n 表示特征空间。 n 是空间的维数，空间的维数就是特征的个数。特征向量举例说明如下：

学生 = {学号, 姓名, 性别, 籍贯, 年龄} 构成一个 5 维的学生的特征矢量。

每个特征向量的取值都是在一定范围内变化，例如：学号：20180001 至

201810000; 姓名: 1 至 8 个汉字组成; 性别: 男或女, 分别用 0 和 1 来表示; 籍贯: 中国某个城市或县城; 年龄: 是 16 至 40 间的正整数。5 个属性的取值范围构成了相应的特征空间。

1.1.1.5 风险决策 (Venture Decision)

对事物进行分类或决策时, 都有可能产生错误, 不同性质的错误会产生各种不同程度的损失, 即分类有风险, 决策须谨慎。衡量了决策后果的决策称为风险决策。例如: 进行股票交易就要冒风险; 金融投资、投资建设项目和企业的规划等都要冒风险。

1.1.1.6 模式空间与特征空间 (Pattern Space and Feature Space)

模式识别实质是从模式空间到特征空间, 再从特征空间到类别空间进行的映射的过程。对样本进行观测得到的数据的集合构成模式, 所有样本数据的集合构成模式空间。模式空间到特征空间通常需要适当的变换和选择, 即特征提取和选择。

利用某些知识和经验可以确定分类原则, 称为判别规则。根据适当的判别规则, 将特征空间里的样本区分成不同的类型, 从而将特征空间转换成了类别空间。类别空间中不同类别的分界面称为决策面。

1.1.1.7 归一化 (Normalization)

归一化是数据预处理常用的方法之一。归一化方法有两种形式, 一是把数据映射在 (0, 1) 之间; 二是把有量纲表达式经过变换, 化为无量纲的表达式, 便于不同单位或量级的指标进行比较。

例 1-1 试将 {2.5 3.5 0.5 1.5} 归一化后变为: {0.3125 0.4375 0.0625 0.1875}

解: 归一化的公式如式 (1-1) 所示:

$$x_i = \frac{x_i}{\sum_{j=1}^n x_j} \quad (1-1)$$

$2.5+3.5+0.5+1.5=8$, $2.5/8=0.3125$, $3.5/8=0.4375$, $0.5/8=0.0625$, $1.5/8=0.1875$ 。

如果归一化的公式如式 (1-2) 所示, 试计算归一化的结果的值是多少?

$$X = (X_0 - \text{Min}) / (\text{Max} - X_0) \quad (1-2)$$

式中 X_0 ——原始值;

X ——变换后的值。

假设学生有身高、体重和年龄三个特征变量，其度量单位分别为：厘米、千克和岁，假设其均值分别为：170 厘米，60 千克和 20 岁。由于量纲不同，如果数量值都为 120，对身高就太矮，对体重则严重超重，而对年龄则几乎不太可能。通常情况下，度量单位越小，数值越大，对结果的影响也越大。归一化目的是防止特征的某一维或几维对数据影响过大，加快程序运行。

1.1.1.8 正则化 (Regularity)

特征数量过多或训练样本过少有可能导致过拟合现象。通常采用降维技术或增加样本数量的方法解决过拟合现象。降维技术可以通过专家的经验进行特征的选择，也可以通过主成分分析 (Principal Component Analysis, PCA) 和因子分析法 (Factor Analysis, FA) 等降维算法来实现。在模式识别过程中，出现过拟合现象时，如果样本数量有限，也不期望减少特征数量，可以用正则化来一定程度控制或解决过拟合问题。

正则化是为了防止过度拟合现象 (过于复杂的模型)，在损失函数里增加每个特征的惩罚因子的过程，即正则化可以减少过度拟合问题。正则化保留所有的特征，但是减少参数的大小，实际上是通过参数约束，在某种程度上减少过拟合。

1.1.2 模式识别应用

模式识别的应用已经渗透到了各个应用领域。由于模式识别的主要任务是利用计算机对客观世界的事物进行判断和识别，因此模式识别应用场景有：生物识别技术、入侵检测系统、签名识别和医学影像等。生物识别技术是指虹膜识别 (Iris Recognition, IR)、人脸识别 (Face Recognition, FR)、指纹识别 (Fingerprint Recognition, FR) 和语音识别 (Automatic Speech Recognition, ASR) 等。

虹膜识别是利用人眼虹膜具有唯一性和稳定性等特点，进行身份认证和识别的过程。唯一性是指人的眼睛是没有完全相同虹膜结构，稳定性则是指虹膜纹理不变。只有活体才能进行虹膜识别。虹膜识别的过程分为 4 个步骤：虹膜图像获取、预处理、特征提取和匹配。虹膜图像采集是用特定的数码设备对人眼拍摄。预处理是对拍摄到的眼部图像进行图像平滑、边缘检测和图像分离等操作。特征提取是从分离出的虹膜图像中提取出特征并编码。特征提取和匹配则是根据特征编码与数据库中存储的虹膜图像特征编码进行比对、验证和识别。

人脸识别是通过人脸图像进行身份认证和识别的过程。为保证安全性，人脸识别技术常常采用活体检测技术，如：让人进行左转、右转、张嘴和眨眼等动作，指令配合错误则判断可能是伪造欺骗。活体检测技术已经成功应用于银行卡

的办理和宾馆入住等多个应用领域。

指纹识别是提取指纹图像的指纹特征，通过识别算法进行识别，来确定指纹所有人身份的生物特征的识别技术，主要包括：指纹图像采集、指纹图像预处理、特征提取、特征值比对与匹配等过程。指纹识别应用也非常广泛，如：智能手机常常采用指纹识别和人脸识别技术进行身份验证。

语音识别则是以语音为研究对象，利用语音信号处理和模式识别技术，让机器自动识别和理解人的语言的过程。即语音识别通过提取语音中的语言的文字信息，将人的语音中的词汇内容转换为计算机可读的数。语音识别技术广泛应用于语音输入法、语音拨号、语音导航和简单的听写数据录入等。常用的语音识别算法有：隐马尔可夫模型（Hidden Markov Model, HMM）、支持向量机和人工神经网络。目前深度学习应用于语音识别取得了非常好的效果。

文本分类是根据文本内容利用模式识别和机器学习技术进行文本分类的过程。许多模式识别与机器学习算法应用于文本分类。

入侵检测系统（Intrusion Detection System, IDS）是对计算机和网络资源的恶意使用行为进行识别和相应处理的系统，是为保证计算机系统的安全而设计与配置的能够及时发现并报告系统中未授权或异常现象的技术。目前基于机器学习与模式识别的入侵检测很多，如：基于遗传算法、基于 SVM 和基于神经网络的人侵检测等，其漏报率和误报率达到了实用要求。

车牌识别技术就是将汽车牌照从复杂背景中提取并进行识别的过程。简单地说就是对车牌的拍照进行识别。车牌识别技术可以解决传统人工登记费时费力的问题，实现自动抬杆、自动计费、自动验证用户身份、自动区分内外部车辆、自动计算车位数、自动报警等自动化和智能化的车辆管理。车牌识别技术主要包括：车牌提取、车牌预处理和车牌的识别。车牌的预处理主要包括车牌提取、车牌定位和字符分割。车牌定位由车牌定位和倾斜校正组成。车牌定位可采用颜色、边缘形状和纹理特征等进行检测定位。

模式识别与机器学习在医学影像分析和遥感图像分类中也取得了广泛的应用。如：医生可以根据心电图，判断病人是否患有心脏病，通过遥感图像分类可以获得土地利用和植被覆盖图，对环境保护和土地利用有辅助决策作用。

目前，虽然模式识别与机器学习技术应用广泛，但仍面临着许多挑战，存在许多问题需要深入研究。在人脸识别中，诸多因素影响着的识别的准确率。如：不同角度照射的光线；随着时间的流逝，特征的变化是绝对的，不变是相对的，如随着人的年龄变化，人脸也在发生变化；人脸角度变化的影响，人的仰视、低头等各种姿态变化，都会对识别的精度产生较大的影响。在语音识别中，存在着由于人的心理和生理变化而引起的语音变化；环境和信道等因素造成语音信号失真等问题需要深入研究。在大数据环境下模式识别算法改进等问题需要深入研究。

1.1.3 模式识别系统

典型的模式识别系统由数据获取、预处理、特征提取与选择、分类器设计和分类决策 5 部分组成，如图 1-2 所示。

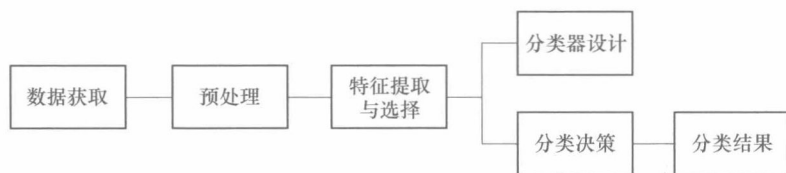


图 1-2 模式识别系统的组成

模式识别系统各部分的基本功能如下：

(1) 数据获取：数据的获取途径或来源有很多种方式。例如：可以通过传统的问卷调查等形式获取数据信息；可以通过网络爬虫软件，按照实际需求自动抓取互联网上的相关数据；可以通过各种传感器获取数据。数据可以是温度、湿度、文字、图像和声音等。

(2) 数据预处理：数据预处理可以有效地提高数据的质量，有利用提高模式识别和机器学习的性能。实际应用中，其方法多种多样，下面介绍几种常用的方法。

1) 数据归一化。把某个特征的所有样本取值限定在 0~1 范围内，如 $[-1, 1]$ 或 $[0, 1]$ 。

2) 数据二值化。把数据的特征取值根据阈值转为 0 或 1。

3) 数据缺损值处理。由于各种原因，导致获取的数据出现缺失的现象。对于缺损的特征数据，采取数据填补的方法解决。常用的填补方法有：均值、中位数和众数填补等。

4) 数据类型转换。如果数据的特征是非数值型的，则需要转换为数值型。

对由于信息获取装置或其他因素所造成的信息退化需要进行复原和去噪。对于离群点、不一致的值、重复数据及有特殊符号的值的也要进行相应的处理。

(3) 特征提取与选择：在获取了原始特征后，需要通过特征提取和选择获取生成有效特征。在保证识别精度的前提下，起到降维的作用，避免产生维数灾难。

(4) 分类器设计：假设样本集为 $D = \{x_1, x_2, \dots, x_n\}$ ，分别属于 c 个类别： $\omega_1, \omega_2, \dots, \omega_c$ 。分类器设计就是建立函数模型 $g(x)$ ，对未知类别的样本 x 进行判别分类的过程。其基本过程是采集样本构建训练集，建立判别函数，确定分类判别规则，确定分类函数机器相应的参数，利用判别函数对 x 进行分类。同