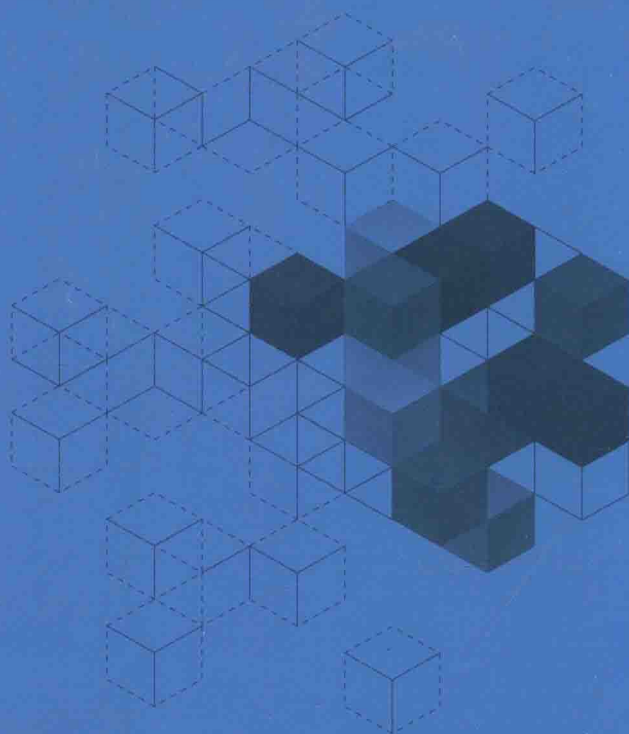


西北工业大学第十八期专著出版基金资助

模式识别 之特征选择

THE FEATURE SELECTION OF
PATTERN RECOGNITION

杨宏晖 申 昇◎著



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

模式识别之特征选择

杨宏晖 申 昇 著

電子工業出版社

Publishing House of Electronics Industry

北京 · BEIJING

内 容 简 介

本书系统深入地论述用于模式识别的特征选择的理论与方法。在内容上注重系统、全面地论述特征选择理论和方法,并力求反映国内、外特征选择领域的最新发展趋势。

全书重点论述各种有监督和无监督特征选择理论及方法;全面涵盖特征选择原理、特征选择方法分类、各类特征选择方法特点及特征选择算法性能评价方法等知识;论述各种特征评价方法;论述各种滤波式、封装式特征选择方法的原理、工作流程及算法步骤;论述特征选择集成方法、特征选择方法与样本选择及分类器集成的融合方法;还包括用于无监督模式识别的基于图谱理论的特征选择方法,用于优化深度学习网络的特征选择方法等的最新研究成果。

本书提供用于大型数据集和高维数据分析的特征选择算法及生物信息学应用的最新资料,涵盖了水声目标识别、字符识别、基于脑功能磁共振成像和单核苷酸多态性等生物信息分析的疾病诊断的多种应用实例。

本书可作为高等院校研究生和高年级本科生的教材外,也可供计算机信息处理、水声信号处理、生物信息处理等领域中从事人工智能、模式识别、机器学习、数据挖掘工作的广大科技人员和高校师生参考。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究。

图书在版编目(CIP)数据

模式识别之特征选择 / 杨宏晖, 申昇著. —北京: 电子工业出版社, 2016.6

ISBN 978-7-121-28268-3

I. ①模… II. ①杨… ②申… III. ①模式识别 IV. ①TP391.4

中国版本图书馆 CIP 数据核字(2016)第 045117 号

策划编辑: 王晓庆

责任编辑: 王晓庆

印 刷: 北京天宇星印刷厂

装 订: 北京天宇星印刷厂

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编: 100036

开 本: 787×1 092 1/16 印张: 11.75 字数: 301 千字

版 次: 2016 年 6 月第 1 版

印 次: 2016 年 6 月第 1 次印刷

定 价: 48.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010)88254888, 88258888。

质量投诉请发邮件至 zltz@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式: wangxq@phei.com.cn。

前言

“偃师所造倡者，趣步俯仰，颔其颐则歌合律，捧其手则舞应节，千变万化，惟意所适。”据史料记载，西周时期，中国的能工巧匠偃师研制出了能歌善舞的伶人，这是中国最早记载的机器人。从古至今，如何使机器做到人可以做的智能工作，是人们一直追求的目标。时至今日，世界各国越来越重视对人工智能的科学研究。

人工智能、模式识别、机器学习是高度相互依赖的多学科领域。数据挖掘技术则融合了这些多学科技术，利用各个应用领域产生的数据来解决各个领域相关的问题。数据挖掘技术炼数成金，从海量数据中提炼宝贵的知识，其在应用领域的应用得到了巨大收益。特征选择是数据挖掘技术的重要组成部分，可以为人们从海量数据中挖掘特定信息、抽取知识提供技术途径。

本书全面阐述特征选择的原理和方法及最新研究成果，论述特征选择在水声目标识别、字符识别及脑功能磁共振成像和单核苷酸多态性生物信息分析与处理等领域的应用原理和方法。本书凝练了作者十余年在特征选择原理及应用的研究成果，获得了国家重大基础研究项目和国家重点实验室基金等项目的支持。本书共 9 章。第 1 章是绪论部分。第 2 章针对特征选择基础知识和原理做了详细系统的归纳和说明。第 3 章论述了特征选择的核心技术特征评价方法。第 4 章和第 5 章致力于滤波式和封装式特征选择算法的构建。第 6 章论述了特征选择、样本选择与分类器集成融合方法。第 7 章重点论述了无监督特征选择方法。第 8 章阐述了特征选择算法的集成。第 9 章论述了特征选择优化深度学习的方法。

在这里要特别感谢西北工业大学的孙进才教授，他对水声目标特征选择研究的贡献也体现在这本书中，感谢他对本书的大力协助和支持，感谢他多年来对本书作者的关心和爱护。

还要感谢美国新墨西哥大学及 The Mind Research Network 的 Executive Science Officer Vince D. Calhoun 教授，感谢他在脑功能磁共振成像和单核苷酸多态性生物信息分析与处理研究工作中的帮助。

本人指导的博士生和硕士生在资料整理、实验、计算等方面为完善本书内容做了大量工作，他们是陈兆基、戴健、杜方键、袁帅、王芸、陈含露、甘安琴、姚晓辉、沈飞、韩振，在此表示诚挚谢意。

本书的出版获得“西北工业大学第十八期专著出版基金”的资助，在此表示感谢。

鉴于用于模式识别的特征选择是一项复杂的学科，涉及的知识面广，本书是作者十余年科研成果的凝练，虽竭尽全力，错误和缺点仍在所难免，诚恳读者指正，并对进一步修改和完善提出宝贵意见。

杨宏晖

2016 年 6 月

目 录

第 1 章 绪论	1
1.1 概述	1
1.2 本书章节安排	3
参考文献	4
第 2 章 特征选择原理	5
2.1 定义	5
2.1.1 特征相关性定义	5
2.1.2 特征选择定义	5
2.2 特征选择过程	6
2.2.1 特征子集的搜索	6
2.2.2 特征子集的评价	8
2.2.3 特征子集搜索的终止	9
2.2.4 选择结果确认	9
2.3 特征选择方法分类	9
2.3.1 滤波式 (Filters)	9
2.3.2 封装式 (Wrappers)	10
2.3.3 混合式 (Hybrid Algorithm)	11
2.3.4 嵌入式 (Embedded Solutions)	12
2.4 特征选择算法的评价	12
2.4.1 特征选择算法的两个重要参数	12
2.4.2 特征选择算法的稳定性评价	12
2.5 本章小结	14
参考文献	15
第 3 章 特征评价方法	17
3.1 概述	17
3.2 基于距离的评价准则	17
3.2.1 质心距离准则	17
3.2.2 类类间最小距离准则	17
3.2.3 类内间距离准则	17
3.2.4 混合距离准则	18
3.3 基于互信息理论的特征评价准则	18
3.3.1 特征选择问题中的信息熵	18
3.3.2 特征选择问题中的互信息	21
3.3.3 互信息特征评价准则	23

3.3.4	连续特征的离散化	24
3.4	基于支持向量机的特征评价准则	26
3.4.1	SVM 的工作原理	26
3.4.2	SVM 分类器分类性能估计及设计	33
3.4.3	基于 SVM 分类性能估计的特征评价准则	37
3.5	本章小结	38
	参考文献	38
第 4 章	滤波式特征选择算法	40
4.1	顺序搜索特征选择算法	40
4.1.1	顺序向前特征选择算法	40
4.1.2	顺序向后特征选择算法	41
4.1.3	顺序双向特征选择算法	41
4.2	Relief 特征选择算法	41
4.3	最近邻收缩规则	42
4.4	Simba 算法	43
4.5	实验及结果	44
4.5.1	实验数据及实验设置	44
4.5.2	实验及讨论	44
4.6	本章小结	47
	参考文献	47
第 5 章	基于人工免疫系统的特征选择算法	49
5.1	人工免疫系统的研究历史和现状	49
5.2	基于免疫机理的特征选择算法	50
5.2.1	算法原理	50
5.2.2	编码与初始种群的生成	50
5.2.3	适应度函数	50
5.2.4	免疫操作	51
5.3	基于免疫克隆机理的特征选择方法	52
5.3.1	算法原理	52
5.3.2	问题编码	53
5.3.3	亲合度函数	53
5.3.4	抗体克隆数	53
5.3.5	免疫克隆算子	53
5.4	遗传算法	55
5.4.1	遗传算法基本概念	55
5.4.2	遗传算法用于特征选择原理	55
5.5	实验与讨论	57
5.5.1	实验目的	57
5.5.2	实验数据	57

5.5.3 实验结果及讨论	57
5.6 本章小结	64
参考文献	65
第 6 章 特征选择、样本选择与分类器集成融合	67
6.1 概述	67
6.2 样本选择原理	67
6.2.1 样本选择的定义	67
6.2.2 样本选择的必要性	67
6.2.3 样本选择方法分类	68
6.2.4 加权最近邻收缩样本选择算法	69
6.2.5 加权免疫克隆样本选择算法	70
6.3 分类器集成	72
6.3.1 分类器集成概述	72
6.3.2 经典分类器集成方法	73
6.3.3 分类器选择性集成算法	76
6.4 样本选择与特征选择融合方法	79
6.4.1 WRNN-FSA 融合算法	79
6.4.2 WICIS-FSA 融合算法	80
6.4.3 实验与讨论	81
6.5 样本、特征与分类器集成的融合方法	85
6.5.1 样本选择、特征选择与分类器集成融合原理	85
6.5.2 MISFFS 算法	86
6.5.3 WRNNIF-SVME 算法	87
6.5.4 WICIF-SVME 算法	87
6.5.5 实验与讨论	88
6.6 本章小结	96
参考文献	96
第 7 章 无监督特征选择算法	101
7.1 基于特征相关性的无监督特征选择	101
7.1.1 基于互信息的无监督特征选择	101
7.1.2 基于非参数互信息的无监督特征选择	105
7.1.3 基于特征相似性的无监督特征选择	107
7.2 基于谱图理论的无监督特征选择	110
7.2.1 基于谱的无监督特征选择	111
7.2.2 谱聚类无监督特征选择	114
7.3 本章小结	118
参考文献	118
第 8 章 特征选择算法集成	121
8.1 概述	121

8.2	特征选择算法集成	121
8.2.1	基本概念	121
8.2.2	个体特征选择器的生成方法	122
8.2.3	特征选择算法集成方法	124
8.3	基于特征排序表的特征选择算法集成方法	125
8.3.1	特征排序表集成概念	125
8.3.2	特征排序表集成方法	126
8.4	特征选择算法集成实验及讨论	127
8.4.1	实验数据介绍	127
8.4.2	基于 Bagging 的特征选择算法集成实验	127
8.4.3	IFSA 算法的集成实验结果及讨论	128
8.4.4	顺序向前特征选择算法集成的实验结果及讨论	130
8.4.5	实验结论	131
8.5	特征选择算法集成的应用实例——基于 fMRI 图像及基因的疾病诊断	131
8.5.1	应用背景	131
8.5.2	利用特征选择与 SVM 集成融合 fMRI 和 SNP 数据进行疾病诊断	134
8.6	本章小结	142
	参考文献	142
第 9 章	特征选择用于深度学习	146
9.1	深度置信网络基础模型——受限玻尔兹曼机	146
9.1.1	受限玻尔兹曼机模型	147
9.1.2	高斯-伯努利受限玻尔兹曼机	148
9.1.3	受限玻尔兹曼机的训练	148
9.2	深度置信网络及其特征学习	150
9.2.1	深度置信网络的提出	150
9.2.2	深度置信网络的训练	150
9.2.3	用于分类识别的深度置信网络	153
9.2.4	特征学习结果分析	153
9.3	基于互信息的深度特征优化选择	155
9.3.1	平均互信息无监督深度特征优化选择	156
9.3.2	非参数互信息无监督深度特征优化选择	161
9.3.3	信息增益有监督深度特征优化选择	165
9.3.4	联合互信息有监督深度特征优化选择	170
9.4	基于特征选择优化的 DBN 模型	173
9.4.1	基于特征选择优化的 DBN 结构	173
9.4.2	基于平均互信息特征选择优化的 DBN 模型	174
9.4.3	基于非参数互信息特征选择优化的 DBN 模型	176
9.5	本章小结	178
	参考文献	178

第1章 绪 论

1.1 概 述

随着计算机技术、信息技术、互联网技术的飞速发展，数据存储成本显著下降，人们面对前所未有的数据爆炸性增长的局面。从海量数据中挖掘特定信息、抽取知识的理论和应用变得越来越重要，世界进入大数据智慧时代。

数据挖掘技术^[1,2]融合了机器学习、统计学、数据库等多学科技术，数据挖掘技术炼数成金，从海量数据中提炼宝贵的知识。特征选择是数据挖掘技术的重要组成部分^[3,4]。特征选择可以减少特征数目，剔除不相关、冗余或噪声数据，加快数据挖掘算法的速度，提高数据挖掘算法性能，如提高模式识别算法的正确识别率。从 20 世纪 70 年代开始，特征选择就成为机器学习、统计模式识别领域的重要研究内容，并广泛地应用到文本识别、语音处理、图像处理、生物信息分析等领域。本书将重点论述特征选择用于机器学习、模式识别的理论和方法。

特征选择应用于统计模式识别系统工作原理的框图如图 1.1 所示，识别系统包括学习过程和测试过程。

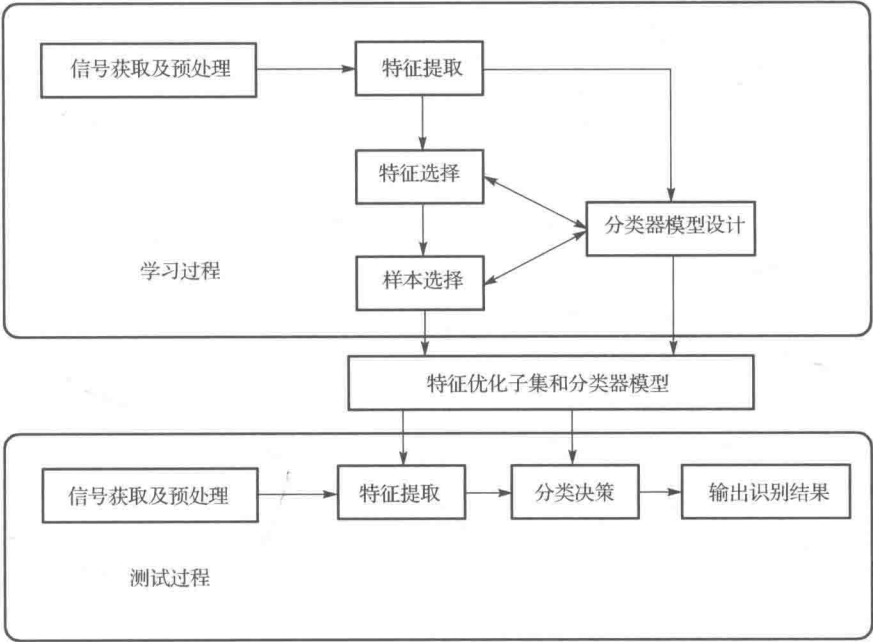


图 1.1 模式识别系统工作原理框图

学习过程主要包括：信号获取及预处理、特征提取、特征选择、样本选择及分类器模型设计；测试过程主要包括：信号获取及预处理、特征提取、分类决策和输出识别结果。

特征提取就是对目标的某些物理性质进行数学描述，具体地讲就是对原始数据进行变换，得到最能反映目标分类的本质特征。一般把原始数据构成的空间称为测量空间，把分类识别赖以进行的空间称为特征空间。通过映射可以把维数较高的测量空间中表示的目标，变为维数较低的特征空间中表示的目标。特征提取是目标分类的关键环节，特征提取得好，即所得的特征向量中包含的类别信息足够多，而干扰信息足够少，则分类效果就好；反之，再好的分类器也无法成功地分类。

由于新技术的不断出现，在各个应用领域，描述目标的新的特征越来越多。特征的增多，一方面带来了更多潜在的分类信息，另一方面，分类器的分类性能并不是总随着特征数目的增加而增加的。通常，分类器的分类性能先随着特征数目的增加而增加，增加到一定的值后会下降，这个现象被称为 Hughes 现象^[5]。Hughes 在 1968 年给出了一个广义上的数据复杂度、训练样本数目、正确分类率三者之间的关系，如图 1.2 所示。

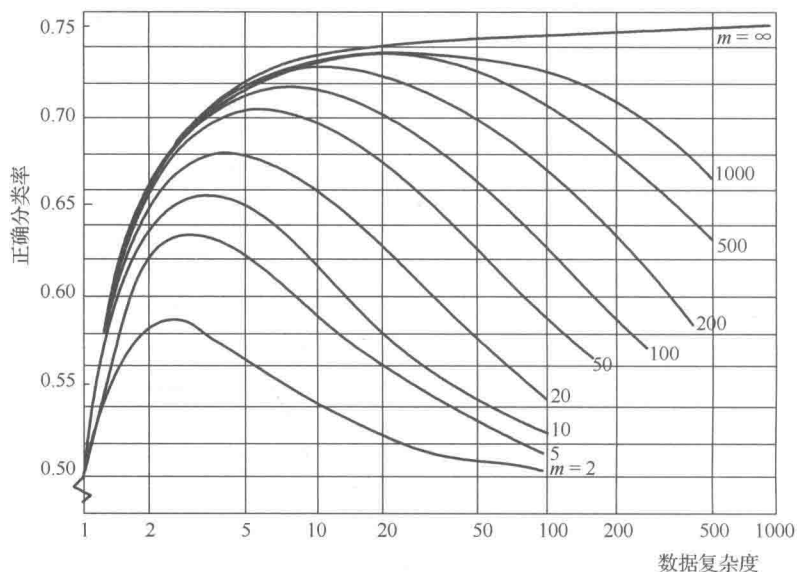


图 1.2 数据复杂度、训练样本数目、正确分类率三者之间的关系

在图 1.2 中，纵轴是所有可能的分类器对两类样本的平均正确分类率，横轴是样本数据的复杂度，图中曲线是不同训练样本数目条件下，正确分类率随训练样本数据复杂度的增加而变化的关系曲线。在水下目标识别中，训练样本数据的复杂度和水下目标样本的特征数目有关。从图中可以看到，如果训练样本数目固定，当特征数目开始增加时，平均正确分类率随特征数目的增加而增高，但是当特征数目增加到一定的数值时，平均正确分类率开始下降。这种现象一方面是由于有限的训练样本造成的，另一方面在众多的特征中，也可能存在冗余特征和与分类任务不相关的特征，从而造成分类器分类性能下降。因此，在有限样本数目的情况下，从大量的目标特征中选择最优的特征子集具有重要的实际意义。

特征选择、样本选择及决策系统的设计是识别系统中很关键的三个方面。特征选择和样本选择是为了从训练样本集中获得更小的特征子集和样本子集，且所训练的分类器的性能优于或等于原训练样本集训练的分类器的性能。其目的就是优化学习算法的分类性能。特征选择可以解决样本数目有限而特征数目不断增加、维数多且复杂、冗余度高，而导致分类系统

分类性能下降的小样本识别问题；样本选择可以解决训练样本集含有噪声样本、冗余样本及无关样本，导致分类系统分类性能下降、不稳定的问题。

分类决策就是在特征空间中用统计的方法把被识别对象归为某一类型。基本方法是在样本训练集基础上确定某个判决规则，按这种判决规则对被识别对象进行分类所造成的错误识别率最小或引起的损失最小。

本书重点讨论特征选择的原理及方法。

1.2 本书章节安排

第1章给出了基于统计模式识别理论的识别系统的组成，给出识别系统各个组成部分的主要功能。

第2章详细论述了特征选择基本概念，特征选择算法框架，特征选择算法各组成部分的功能及特点，特征选择算法的分类及不同算法的原理和特点，最后给出了特征选择算法的评价方法。

第3章论述了特征选择方法中的关键技术——特征评价方法，论述了基于距离、互信息理论及支持向量机的特征评价方法，分析了各个方法的原理、实现方法及特点。

第4章论述了滤波式特征选择方法，分析了滤波式特征选择方法的原理，论述了几种经典特征选择方法的核心思想、算法步骤及各自的特点。

第5章论述了封装式特征选择方法，重点介绍了基于人工免疫机理的特征选择方法。论述了这类算法在特征搜索、特征评价等关键技术方面的特点与优势。

第6章讲述了将特征选择方法用于数据融合的方法，论述了样本层、特征层及决策层的融合技术。本章首先给出了样本选择的定义和方法，论述了分类器集成设计方法；接着论述了特征选择与样本选择融合方法；最后论述了将三者融合在一个框架下的理论和方法。

第7章论述了无监督特征选择方法的原理、实现方法及应用。重点论述了基于互信息的无监督特征选择方法、基于非参数互信息的无监督特征选择方法及基于谱图理论的无监督特征选择方法。

第8章介绍了用于大规模数据处理的特征选择集成算法。论述了特征选择集成方法的原理、组成部分、分类及各种方法的特点。详细论述了特征选择算法集成在融合脑功能磁共振成像(fMRI)和单核苷酸多态性(Single Nucleotide Polymorphisms, SNPs)这两种高维生物信息数据进行疾病诊断的应用实例。

第9章论述了最新模式识别技术——深度学习理论及方法，论述了深度学习网络初始化方法、训练方法、学习过程监测方法及学习过程调整方法。论述了基于特征选择方法的深度学习网络优化方法。

本书中的知识点都是通过一个一个实际的应用实例来讲解的，读者可以轻松地掌握实现需求的技术细节，特别是在各章中融入了水声目标识别、字符识别、基于脑功能磁共振成像和单核苷酸多态性生物信息分析的疾病诊断的多种应用实例，可以为这些领域的研究人员提供参考。

参 考 文 献

- [1] R. Agrawal, T. Imielinski, and A. Swami. Database Mining: A Performance Perspective, IEEE Trans. Knowledge and Data Eng., vol. 5, no. 6, pp. 914-925, 1993.
- [2] J. Han and M. Kamber. Data Mining: Concepts and Techniques. Morgan Kaufman, 2001.
- [3] A.L. Blum and P. Langley. Selection of Relevant Features and Examples in Machine Learning, Artificial Intelligence, vol. 97, pp. 245-271, 1997.
- [4] Huan Liu and Lei Yu. Toward Integrating Feature Selection Algorithms for Classification and Clustering. IEEE Transactions on Knowledge and Data Engineering, vol. 17, No. 4, pp.491-502, 2005.
- [5] Hughes G F. On the Mean Accuracy of Statistical Pattern Recognizers. IEEE Trans. Information Theory, IT-14(1):55-63, 1968.

第2章 特征选择原理

2.1 定 义

2.1.1 特征相关性定义

在模式识别任务中,确定特征是否和分类任务相关是非常重要的。Kohavi 和 John 关于特征相关性、冗余性的定义得到广泛的认可^[1]。根据特征对贝叶斯分类器(对某一给定问题的最优分类器)分类性能的影响,特征相关性、冗余性的定义可归纳如下。

定义 2.1: 如果只去掉特征 V_i ,就使得最优的贝叶斯分类器的分类性能恶化,则称 V_i 是强相关特征。

定义 2.2: 如果一个特征 V_i 不是强相关的,且存在特征子集 S ,使得贝叶斯分类器在特征子集 S 上的分类性能比在 $S \cup V_i$ 上的分类性能差,则称特征 V_i 是弱相关的。

定义 2.3: 既不是强相关又不是弱相关的特征就是不相关特征。

定义 2.4: 如果对于一个特征 V_i ,存在特征子集 S ,使得贝叶斯分类器在特征子集 $S \cup V_i$ 上的分类性能比在 S 上的分类性能差或没有改变,则称该特征为冗余特征。

如图 2.1 所示,单独使用特征 x 或特征 y 得到的数据分类结果完全相同,因此特征 x 与特征 y 中存在冗余特征。如图 2.2 所示,由于单独使用特征 y 无法区分两类数据,即其对识别任务没有贡献,因此特征 y 是与识别任务不相关的特征。

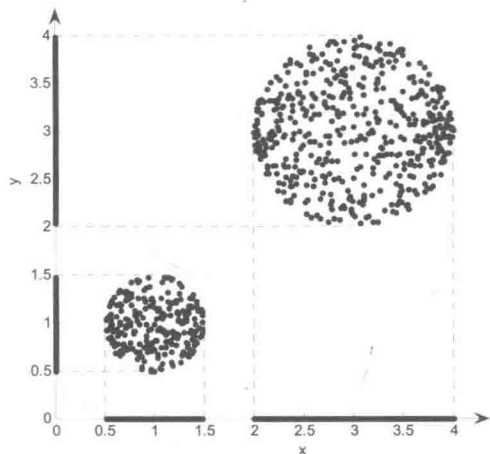


图 2.1 冗余特征示意图

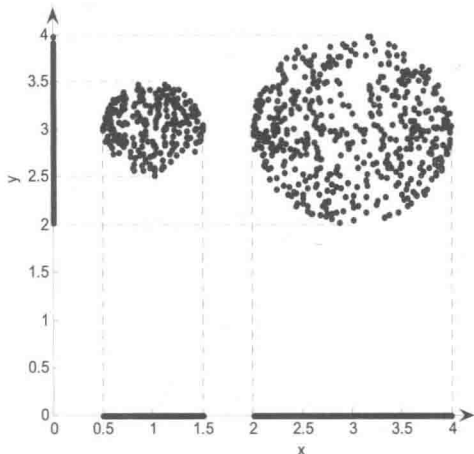


图 2.2 不相关特征示意图

2.1.2 特征选择定义

在模式识别中,特征选择的目的是优化学习算法的分类性能。设 X 为训练样本集

$$X = \{(x_i, y_i) | x_i \in \mathbb{R}^d, y_i \in Y, i = 1, 2, \dots, n\} \quad (2.1)$$

式中, d 是样本的特征数目, Y 是有限的类标集合。

特征选择可以定义为^[2,3]:

- (1) 给定 $m \leq d$, 从 d 个原始特征中选出 m 个特征, 使得分类器的分类错误率最低;
- (2) 给定错误分类率, 从 d 个原始特征中选择出满足预先设定的最小错误分类率的最少的 m 个特征。

2.2 特征选择过程

在有 d 个特征的情况下, 特征选择过程是根据某一评价函数对原始特征集合的所有 2^d 个特征子集进行搜索, 选择一个使评价函数值最优的特征子集的过程。但在实际中, 当特征数目较多时, 对所有的特征子集进行穷举搜索是不现实的。因此, 对特征选择方法进行研究的目的是解决这一问题, 使得用户在不需要对特征空间的所有子集进行穷举搜索的情况下, 仍能找到原始特征集合的一个最优或次优的特征子集。

如图 2.3 所示, 特征选择过程是一个循环搜索最优特征子集的过程, 主要包含 4 个部分^[2]: 特征子集的生成, 特征子集的评价, 搜索终止条件, 选择结果确认。特征子集的生成是一个搜索过程, 其生成候选特征子集后, 特征评价函数对每个候选特征子集进行评价并与前一个循环中得到的最优特征子集相比较; 如果新子集更优, 则替代原最优特征子集。特征子集的生成和评价过程循环往复直到搜索终止条件被满足。最后选择的最优特征子集利用先验知识进行评价, 或者利用实测数据进行评价。

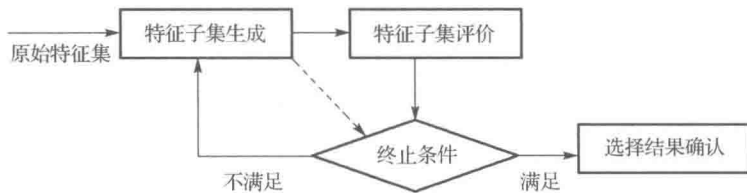


图 2.3 特征选择过程

2.2.1 特征子集的搜索

特征子集的生成是最优特征子集搜索过程中很重要的一步。在每个循环中, 产生有可能是最优特征子集的特征子集。特征子集的产生方法确定了搜索最优特征子集的搜索起点及后续的搜索方式。

1. 搜索起点的确定

搜索起点 (也就是初始特征子集) 的产生有三种方式: 空集、特征全集和随机产生的特征子集。最优特征子集从空集开始并逐步向里面加入特征, 称为向前特征选择算法; 最优特征子集从全集开始并逐步从里面删除特征, 称为向后特征选择算法; 也可以同时从空集和全集开始, 并同时从里面加入和删除特征, 这个方法称为双向特征选择算法^[4]。特征子集的搜索也可以从一个随机产生的特征子集开始来避免得到局部最优解。

2. 搜索策略的确定

搜索起点确定之后, 特征空间的搜索方向也就确定了。下一步需要确定的是如何组织搜

索算法对整个搜索空间进行搜索,即确定搜索策略。在有 d 个特征的情况下,有 2^d 个特征子集。在现有的特征选择算法中,常用的搜索策略有三种:完全搜索、顺序搜索和随机搜索。

(1) 完全搜索

在给定评价函数后,完全搜索保证在整个特征空间搜索出最优特征子集。穷举搜索算法就是一种完全搜索算法。启发式搜索算法可以在减少搜索空间的情况下做到完全搜索,保证所有可能的特征子集被考虑到。Narendra 提出的分支定界算法(Branch and Bound, B&B)^[5]是最常见的完全搜索算法。

在分支定界算法中,要求所使用的评价函数具有单调性。对于集合 $S_0 \subset S_1 \subset \dots \subset S_n$,给定某一评价函数 J ,若评价函数 J 为单调函数,则应满足 $J(S_0) \leq J(S_1) \leq \dots \leq J(S_n)$ 。许多距离和信息评价函数如 Mahalanobis 距离、Bhattacharyya 距离,以及 Fisher 判别函数等都具有单调性。Narendra 在提出分支定界算法时,使用 Mahalanobis 距离作为评价函数。当某一特征集合的评价函数值小于阈值 B 时,该特征集合的所有子集的评价函数值都小于阈值 B ,因此,该特征集合的所有子集也不可能是最优的特征子集。评价函数的单调性使得分支定界算法无须对整个搜索空间进行穷举搜索,只需对少数特征集合进行评价即可找到全局最优解。因此分支定界算法是能最好地替换穷举搜索方式的特征选择方法。

分支定界算法的搜索过程可以使用搜索树进行说明。图 2.4 所示为从 6 个特征中选两个特征的搜索树,每个节点上标的数字表示去掉的特征序号,变量 $Z_i (i=1,2,3,4)$ 表示在第 i 级上去掉的特征。给定要求选择的特征数目 M ,总特征数为 d ,则搜索树的级数为 $\bar{M} = d - M$ 级。对于 6 个特征选两个特征的问题,搜索树共有 4 级。假设评价函数 J 满足单调性,则 $J(Z_1) \geq J(Z_1, Z_2) \geq \dots \geq J(Z_1, Z_2, \dots, Z_{\bar{M}})$,其中 $J(Z_i)$ 表示去掉特征 Z_i 后剩余特征的评价函数值。在搜索树的每一级上,各特征集合的子集数目不同,评价函数值小的特征集合的子集数目多于评价函数值大的特征集合。例如,在第一级上,去掉特征 3 后的特征集合的评价函数值是这一级集合的评价函数中最大的,因此该集合的子集数目比同级的其他集合的子集数目要少;在第二级上,由第一级中数字标号为 1 的节点产生的子集合有 3 个,分别为数字标号为 2、3、4 的节点对应的集合,在这 3 个集合中,位于最右边的集合(数字标号为 4 的节点对应的集合)的评价函数值是最大的,该集合的子集数目是 3 个集合中最少的。分支定界算法由特征全集开始,对搜索空间进行后向搜索。分支定界算法在搜索最优特征集合时,在每一级上都选择未搜索过的节点中最右边的节点开始搜索(评价函数最大的节点)。分支定界算法中评价函数的阈值 B 在初始时刻设为 0,在搜索树最右边的分支上,每个节点的 J 值都大于初始时刻的阈值。在搜索过程中,当搜索到树的叶节点时,就用相应的 J 值代替原来的阈值 B ,搜索向上回溯。在搜索未到达搜索树的叶节点,而某一节点的 J 值小于阈值 B 时,则将该节点以下的子树从搜索树中删除。从上述内容可以看到分支定界算法的计算过程复杂、计算量大,当特征数目较大时,该算法受到很大限制。

(2) 顺序搜索

为了避免完全搜索带来的巨大计算量,而采取的次优搜索算法,如顺序向前搜索算法(Sequential Forward Selection, SFS)、顺序向后搜索算法(Sequential Backward Selection, SBS)及双向搜索算法(Bidirectional Selection, BS)^[4]。顺序搜索算法无法全面检验特征间的相互关系,当特征数目较高时,常会漏掉最优特征组合方式^[6]。后面章节将详细介绍这几类算法的工作原理。

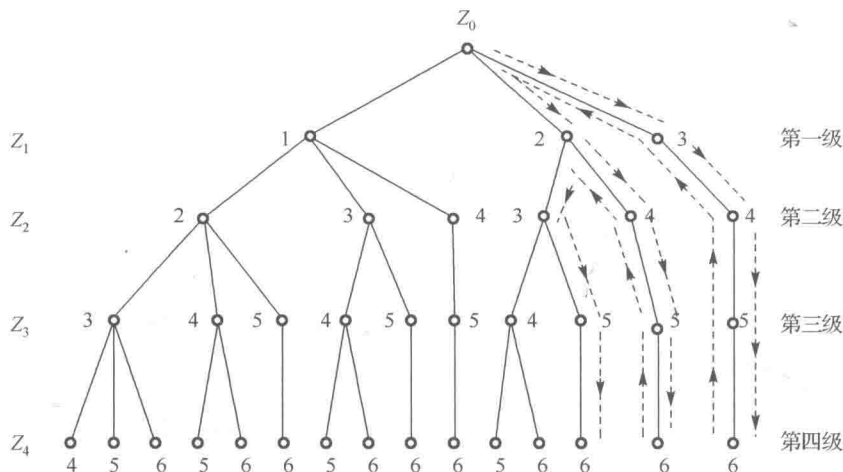


图 2.4 分支定界算法的搜索树和搜索过程

(3) 随机搜索

从一个随机选择的特征子集开始，并在下一个特征子集的生成过程中加入一定的随机性。随机搜索可以避免搜索过程停止在局部极小点。遗传算法 (Genetic Algorithm, GA) [7~13]、随机爬山法 (Random-start Hill-climbing, RH) [14] 等是随机搜索算法。后面章节将详细介绍用于特征选择的遗传算法的工作原理。

(4) 混合搜索策略

混合搜索策略将先验知识与随机搜索相结合，根据先验知识，引导算法从可能出现最优解的空间进行搜索，这样可以利用先验知识，并且避免随机搜索过于发散。基于人工免疫机理的特征选择算法就是利用混合搜索策略 [15~18] 的。

2.2.2 特征子集的评价

在模式识别中，特征子集的评价就是根据给定评价函数评价特征子集对分类的重要性；在循环搜索中，如果新的特征子集优于原来的特征子集，则将其替代。这里要注意的是，应用不同的评价函数选择出的最优特征子集是不同的。根据评价函数和学习算法的关系，可以将评价函数分为独立评价函数和非独立评价函数。

1. 独立评价函数

独立评价函数：与学习算法无关，仅根据训练样本固有的特性评价特征子集的分类性能。常用的独立评价函数有距离评价函数、信息评价函数、相关性评价函数、一致性评价函数 [19~21]。

(1) 距离评价函数也称为可分性准则。对一个两类问题，如果特征 X 的存在使两类的条件概率之间的差异大于特征 Y ，则选择特征 X ；如果两个特征的存在使两类的条件概率没有差异，则特征 X 和 Y 两者之间没有差别。

(2) 信息评价函数用来衡量一个特征的存在所提供的信息增益的大小。特征 X 的信息增益定义为使用特征 X 前后的不确定性之间的差值。如果特征 X 的信息增益大于特征 Y 的信息增益，则优先选择特征 X 。

(3) 相关性评价函数用来度量由一个特征的值预测另一个特征的值的的能力大小。相关系数是一种经典的相关性度量标准，使用相关系数可以确定某一特征和某个类别之间的相关性

大小。如果特征 X 与类别 C 之间的相关性大于特征 Y 与类别 C 之间的相关性，则优先选择特征 X 。相关系数还可以用来确定一个特征对其他特征的依赖性，相关系数值的大小表示某一特征的冗余程度。所有的相关性评价函数又可以分为距离评价函数和信息评价函数，变量与变量之间的相关性可以使用距离和信息增益来度量。一致性评价函数是一类比较新的评价函数。

(4) 一致性评价函数与训练数据高度相关。使用一致性评价函数可以选择满足不一致率要求的最小特征子集，通常对数据的不一致率要求由用户给定。

2. 非独立评价函数

非独立评价函数：用学习算法的分类性能来评价特征子集对分类的重要性。正确分类率是最常用的非独立评价函数。

第3章将详细介绍各种特征评价函数的原理及应用。

2.2.3 特征子集搜索的终止

搜索终止条件是特征子集产生及评价停止的条件。特征子集的产生过程和评价函数都对终止条件的产生有影响。

基于特征子集的产生过程的终止条件有：(1) 预先设定所要选择的特征集合的特征数目作为终止条件；(2) 预先设定特征子集产生过程的循环次数。

基于评价函数的终止条件可以是：(1) 根据增加或删除某一特征是否产生一个评价函数值更好的特征子集来终止特征选择过程；(2) 根据某一评价函数是否得到了一个评价函数值最优的特征子集来终止特征选择过程。

2.2.4 选择结果确认

选择结果确认就是确认选择结果是否有效。如果事先知道特征集中哪些特征是冗余的、不相关的，那么可以直接对特征选择结果进行判断。但对于复杂的实测数据提取的高维特征，则无法预先知道哪些特征是冗余的、不相关的，因此可以比较特征选择前、后分类器的正确分类率，来判断选择结果是否有效。

2.3 特征选择方法分类

根据评价函数和学习算法的关系，可以将现有的特征选择方法分成4类：滤波式、封装式、混合式和嵌入式。

2.3.1 滤波式 (Filters)

滤波式特征选择算法完全独立于分类算法，与分类算法的分类性能及其他参数无关^[22]。滤波式特征选择算法可以视为数据预处理过程。表2.1所示为滤波式特征选择算法的工作原理。改变评价函数、搜索方式可以得到各种滤波式特征选择算法。滤波式特征选择算法常常使用独立评价函数。典型的滤波式特征选择算法将在后续章节中详细论述。滤波式特征选择算法的通用性使得其可用于解决各种特征选择问题，但是由于其与学习算法无关，因此选择的特征子集分类性能常常低于其他算法。