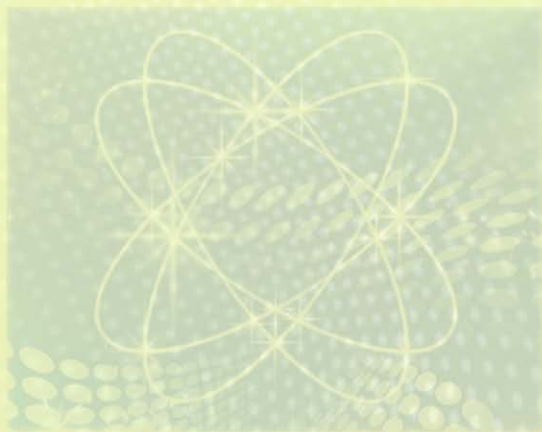


特征选择与提取研究与应用

陈争光 杨冬风 著



黑龙江教育出版社

图书在版编目(CIP)数据

特征选择与提取研究与应用 / 陈争光 杨冬风著
-- 哈尔滨: 黑龙江教育出版社, 2012. 7
ISBN 978 - 7 - 5316 - 6514 - 4

I. ①特… II. ①陈…②杨… III. ①特征抽取—研究 IV. ①TP391.4

中国版本图书馆 CIP 数据核字(2012)第 154039 号

特征选择与提取研究与应用

Tezheng Xuanze Yu Tiqu Yanjiu Yu Yingyong

陈争光 杨冬风 著

责任编辑 徐永进
封面设计 张 慰
责任校对 程 佳
出版发行 黑龙江教育出版社
(哈尔滨市南岗区花园街 158 号)
印 刷 黑龙江远东联达教育文化传媒有限公司
开 本 880 毫米×1230 毫米 1/32
印 张 11.25
字 数 250 千
版 次 2012 年 7 月第 1 版
印 次 2012 年 7 月第 1 次印刷

书 号 ISBN 978 - 7 - 5316 - 6514 - 4 定 价 30.00 元

黑龙江教育出版社网址: www.hljep.com.cn

如需订购图书,请与我社发行中心联系。联系电话:0451 - 82529593 82534665

如有印装质量问题,请与我社联系调换。联系电话:0451 - 82529347

如发现盗版图书,请向我社举报。举报电话:0451 - 82560814



前 言

模式识别诞生于 20 世纪 20 年代,其目的是用机器去完成人类智能中通过视觉、听觉、触觉等感官去识别外部环境的工作。在 20 世纪 60 年代模式识别迅速发展成为一门科学,它所研究的理论和方法在许多科学和技术领域中受到了广泛的重视,推动了人工智能的发展。

特征选择与提取是模式识别过程的关键环节,在很多实际问题中常常不容易找到那些对于识别最重要的特征,或受条件限制不能对它们进行测量,这就使特征选择和提取的任务复杂化而成为构造模式识别系统最困难的任务之一,特征选择与提取也成为许多科研工作者研究的重要课题。

本书对特征选择与提取相关论著中的重点内容进行了概述和系统、科学的分类,从而较为全面地介绍了各种特征选择与提取的方法和重要结论。在图像模式识别中,大量的应用是关于颜色、形状和纹理的特征选择与提取,本书着重对这三个方面的算法进行分析、对比和总结,对于在模式识别领域中从事科学研究的学者来说,具有重要的指导意义。在本书的后半



特征选择与提取研究与应用

部分,著者从应用的角度来论述在具体的科研实践中如何选择特征选择与提取方法。这部分包含了马铃薯外部特征选择与提取、种蛋品质检测中的特征选择与提取、钼靶 X 线图像乳癌检测中的特征选择与提取以及大豆生长阶段预测中特征选择与提取。

全书共分 12 章。第 1 章绪论,主要说明特征选择和提取研究的目的、意义、研究的现状以及本书的主要内容和章节安排。第 2 章特征选择和提取的内涵及类的可分性判据,主要讲述特征的一些性质、特征选择和提取的定义以及类的可分性判据。第 3 章特征提取算法的分类,主要对大量的特征提取算法进行系统的分类,并简要介绍各种特征提取算法。第 4 章特征选择算法研究,主要介绍一般特征选择算法的步骤,对特征选择算法进行系统的分类,并着重分析和研究两种特征选择算法——遗传算法和人工鱼群算法。第 5 章颜色特征提取算法研究,首先对颜色的基本特性,各种颜色空间进行介绍,然后对颜色矩、颜色集、颜色直方图、颜色相关图等颜色特征提取算法进行分析和研究。第 6 章形状特征提取算法研究,首先对形状识别进行概述,然后着重对基于傅里叶描述子的形状识别方法、基于主分量分析的形状特征提取、基于不变矩理论的特征提取方法、Hough 变换形状特征提取以及链码技术进行分析和研究。第 7 章纹理特征提取算法研究,首先介绍纹理的定义和纹理分析研究的内容,对现有的纹理特征提取算法进行系统、科学的分类,然后对统计方法、结构化方法、基于模型的方法以及基于信号处理的纹理特征提取算法进行了分析和研究。第



8 章马铃薯外部形态特征选择与提取研究,以马铃薯为研究对象,提出马铃薯的外部形态特征(重量、大小、创伤缺陷面积、绿皮缺陷)提取方法。第 9 章种蛋品质检测中特征选择与提取研究,本研究通过建立种蛋筛选的机器视觉检测系统,将机器视觉和特征选择与提取技术结合起来进行种蛋筛选。实现未上孵无精蛋剔除、种蛋外部品质特征提取以及无损伤种蛋性别鉴别。第 10 章钼靶 X 线图像乳癌检测中特征选择与提取研究。本研究中,不是单独使用神经网络,粗糙集,模糊图像处理当中的一种智能技术,而是把三种技术集成起来,降低各自的弱点,增强各自的能力。第 11 章大豆生长阶段预测中特征选择与提取研究。本研究中,根据统计数据,提取出大豆生长发育中具有标志性意义的特征参数,采用神经网络构建简单有效的预测模型,用来指导大豆种植者对大豆生产进行有效管理。第 12 章神经网络与模式识别,特征选择与提取是模式识别的关键环节,神经网络是模式识别非常重要的分类工具,因此,把神经网络在模式识别中应用的相关知识专门提取出来,希望能为相关人员起到指引的作用。

本书由黑龙江八一农垦大学教师陈争光;黑龙江八一农垦大学教师杨冬风合著。第 1,2,5,6,7,8 章由陈争光撰写,共 12 万字;3,4,9,10,11,12 章由杨冬风撰写,共 12 万字。在撰写本书的过程中,参考和引用了一些论文和资料,在此对援引资料的作者表示衷心的感谢。

本书是在以下项目的支持下完成的:(1) 黑龙江省教育厅项目:马铃薯外部形态特征提取研究(项目编号:11541258);



特征选择与提取研究与应用

(2) 农业部 948 项目: 高油、高蛋白质、高异黄酮大豆种质资源及先进技术引进与创新(项目编号: 2008 - Z24) ; (3) 黑龙江省教育厅项目: 基于嵌入式的未上孵种蛋性别鉴别(项目编号: 11551317) 。

由于模式识别在各种领域中正在迅速的发展之中,各方面的研究层出不穷,加之作者水平有限,本书的内容难免存在疏漏和不足之处,恳请广大读者批评指正。



目 录

1 绪论	(1)
1.1 研究的目的和意义	(1)
1.2 研究现状	(4)
1.3 本书的主要内容和章节安排	(8)
2 特征选择与提取的内涵与类的可分性判据	(11)
2.1 特征的特点	(11)
2.2 特征的类别	(12)
2.3 特征的形成	(13)
2.4 特征选择与提取内涵	(14)
2.5 特征选择和提取的作用	(15)
2.6 类的可分离性判据	(15)
2.6.1 基于距离的可分性判据	(16)
2.6.2 基于概率密度函数的可分性判据	(19)
2.6.3 基于熵函数的可分性判据	(20)
3 特征提取算法分类	(22)
3.1 特征提取的分类	(22)



3.2	低层特征提取	(24)
3.3	高层次特征提取	(27)
4	特征选择算法研究	(31)
4.1	特征选择问题的构成	(32)
4.1.1	搜索起点和方向	(33)
4.1.2	搜索策略	(33)
4.1.3	特征评估函数	(34)
4.1.4	停止准则	(35)
4.2	特征选择方法分类	(35)
4.2.1	按照搜索策略分类	(36)
4.2.2	按特征集合评估准则划分特征选择算法	(42)
4.3	遗传算法	(45)
4.3.1	运算过程	(46)
4.3.2	基本遗传算子	(46)
4.4	人工鱼群算法与改进的人工鱼群算法	(49)
4.4.1	人工鱼群算法原理	(49)
4.4.2	改进的人工鱼群算法	(50)
4.4.3	仿真实验及对比分析	(53)
5	颜色特征提取方法研究	(58)
5.1	颜色的基本特性	(59)
5.2	颜色模型	(60)
5.3	常用的颜色特征提取方法	(64)
5.3.1	基于颜色空间的特征提取	(64)



5.3.2	基于颜色直方图的特征提取	(66)
5.3.3	颜色矩	(67)
5.3.4	颜色集	(68)
5.3.5	颜色相关图	(68)
6	形状特征提取方法研究	(70)
6.1	形状识别概述	(71)
6.2	形状识别系统	(73)
6.3	常用形状识别方法	(74)
6.4	基于傅里叶描述子的形状识别方法	(76)
6.5	基于主分量的形状识别	(78)
6.6	基于不变矩理论的形状特征提取	(79)
6.7	Hough 变换目标形状特征检测	(82)
6.7.1	Hough 变换检测直线	(82)
6.7.2	Hough 变换检测圆	(86)
6.7.3	Hough 变换检测椭圆	(89)
6.7.4	广义 Hough 变换	(91)
6.8	链码技术	(94)
6.8.1	Freeman 链码	(94)
6.8.2	新的链码编码方式	(95)
6.8.3	Freeman 直线链码识别	(96)
6.8.4	平均链码	(100)
6.8.5	Freeman 链码对圆及椭圆的识别	(101)
6.8.6	链码和	(102)
6.8.7	链码差	(104)



7 纹理特征提取方法研究	(105)
7.1 纹理的定义及纹理分析	(105)
7.1.1 纹理的定义和描述	(106)
7.1.2 纹理分析的研究内容	(108)
7.2 纹理特征提取算法分类	(109)
7.3 统计方法	(111)
7.3.1 灰度直方图法	(111)
7.3.2 自相关函数法	(113)
7.3.3 He 和 Wang 的纹理谱	(114)
7.3.4 灰度游程长度法	(115)
7.3.5 灰度共生矩阵纹理分析法	(117)
7.4 结构化方法	(120)
7.5 模型方法	(121)
7.5.1 分形模型	(121)
7.5.2 马尔可夫随机场(MRF) 模型	(126)
7.6 信号处理的方法	(127)
7.6.1 基于信号分解的纹理分析方法	(127)
7.6.2 优化的信号纹理分析方法	(128)
8 马铃薯外部形态特征提取与选择研究	(134)
8.1 研究的目的是和意义	(134)
8.2 国内外研究现状	(138)
8.3 研究的主要内容	(139)
8.4 马铃薯图像快速获取	(140)
8.4.1 Directshow 核心技术	(141)



8.4.2	马铃薯外部检测图像提取过程	(143)
8.4.3	程序架构	(144)
8.4.4	使用 DirectShow 枚举设备	(144)
8.5	马铃薯图像分割算法研究	(148)
8.5.1	阈值法的基本原理	(148)
8.5.2	三次样条在多阈值分割中的应用	(153)
8.5.3	图像分割阈值的确定方法	(156)
8.5.4	应用实例及实验效果	(159)
8.6	马铃薯大小特征提取	(161)
8.6.1	面积特征提取	(162)
8.6.2	周长特征提取	(163)
8.6.3	体积特征提取	(169)
8.6.4	检测结果对比分析	(174)
8.7	马铃薯表面缺陷特征提取	(175)
8.7.1	曲面面积计算的理论基础	(176)
8.7.2	实验过程	(178)
8.7.3	结论与讨论	(180)
8.8	马铃薯绿皮检测方法研究	(181)
8.8.1	颜色特征提取	(182)
8.8.2	识别方法	(186)
8.8.3	结果与分析	(191)
9	种蛋品质检测中特征选择与提取研究	(193)
9.1	研究的目的是和意义	(193)
9.2	国内外研究现状	(196)



特征选择与提取研究与应用

9.3	研究的主要内容	(200)
9.4	系统的硬件构成	(201)
9.5	种蛋图像预处理	(202)
9.5.1	图像灰度变换	(202)
9.5.2	图像平滑处理	(203)
9.5.3	种蛋图像分割	(204)
9.5.4	种蛋图像边缘检测	(207)
9.6	种蛋外部尺寸特征提取	(209)
9.7	种蛋重量特征提取	(213)
9.7.1	重量 - 面积线性回归模型	(214)
9.7.2	重量 - 周长线性回归模型	(216)
9.7.3	多元线性回归模型	(218)
9.8	种蛋表面缺陷特征提取	(222)
9.8.1	脏斑血斑特征提取	(223)
9.8.2	蛋壳裂纹特征提取	(225)
9.9	种蛋性别鉴别	(234)
9.9.1	种蛋蛋形指数计算	(235)
9.9.2	离散小波变换种蛋边缘特征提取	(237)
9.9.3	BP 神经网络设计	(239)
9.9.4	网络训练及网络测试	(240)
9.10	无精蛋特征提取研究	(241)
9.10.1	种蛋蛋边缘轮廓近似椭圆拟合方法	(242)
9.10.2	特征计算数学模型	(244)
9.10.3	网络训练及网络测试	(245)



10 钼靶 X 光图像乳癌检测中特征选择与提取研究	(247)
10.1 研究的目的及意义	(247)
10.2 研究现状	(249)
10.3 研究的主要内容	(252)
10.4 相关智能技术	(255)
10.4.1 图像模糊处理	(255)
10.4.2 粗糙集	(262)
10.5 自适应的模糊图像增强算法	(270)
10.5.1 Russo 的基本方法	(271)
10.5.2 方法改进	(273)
10.5.3 实验结果	(275)
10.6 FCM 聚类算法进行图像分割	(276)
10.6.1 FCM 聚类算法	(277)
10.6.2 基于 FCM 的信息表构造和区域划分	(279)
10.6.3 基于差异度的初始等价关系及相似域划分	(280)
10.6.4 图像分割实验	(281)
10.7 乳腺特征提取	(286)
10.8 粗糙神经分类器	(288)
10.9 结果与讨论	(290)
10.9.1 评价标准	(290)



特征选择与提取研究与应用

10.9.2	增强效果讨论	(293)
10.9.3	分割结果分析	(295)
10.9.4	产生规则的数目和整体精度	(297)
11	大豆生长阶段预测中特征选择与提取研究	(299)
11.1	实验数据来源及记录方法	(301)
11.2	特征参数的选择	(302)
11.3	模型的构建	(303)
11.3.1	插值法	(304)
11.3.2	神经网络模型构建	(305)
11.4	结论	(308)
12	神经网络与模式识别	(309)
12.1	神经网络与模式识别	(309)
12.2	神经网络模式识别方法与传统模式识别方法 的关系	(310)
12.3	神经网络模式识别的特点	(310)
12.4	用于模式识别的神经网络模型	(312)
12.5	神经网络模式识别系统	(313)
12.6	神经网络模式识别的典型方法	(315)
12.6.1	多输出型	(316)
12.6.2	单输出型	(316)
参考文献		(319)

1 绪 论

1.1 研究的目的是和意义

模式识别(Pattern Recognition) 是借助计算机,自动模拟人类对外部世界某一特定环境中的客体、过程和现象的识别功能(包括视觉、听觉、触觉、判断等) 的科学技术。

模式识别技术在社会生活和科学研究的许多方面有着巨大的现实意义,已经在许多领域得到了广泛应用。随着计算机技术和人工智能、思维科学研究的迅速发展,模式识别技术正在向更高、更深的层次发展。

一个模式识别系统由数据获取、预处理、特征选择和提取以及识别四大部分组成,如图 1 - 1 所示。

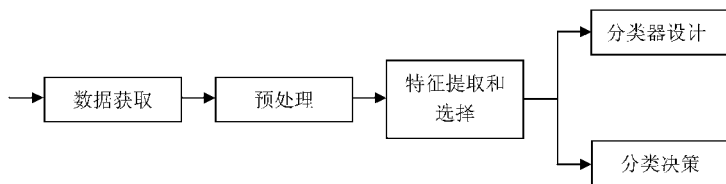


图 1 - 1 模式识别系统框图

特征选择是指从一组特征中挑选出对分类最有利的特征,达到降低特征空间维数的目的。特征提取是指通过映射(变换) 的方法获取最有效的特征,实现特征空间维数从高维向低



特征选择与提取研究与应用

维的变换。从定义可以看出,实现特征选择和提取的前提是确定特征是否有效的标准,在此标准下寻找最有效的特征子集。

从分类的角度看,模式识别是把具体事物归到具体某一类别的过程。也就是先用一定数量的样本,根据它们之间的相似性进行分类器设计,而后用所设计的分类器对待识别的样本进行分类决策。识别过程既可以在原始数据空间(预处理之后的数据)中进行,也可以对原始数据进行变换,将数据映像到最能反映分类本质的特征空间中进行。很明显,经过特征选择与提取后分类器的设计会更为容易。因为通过更为稳定的特征表示,分类器的性能得到了提高,并且由于删去多余或不相关的信息,使得研究对象之间的本质联系更容易被发现。可以说,特征选择与提取是模式识别系统设计成败的关键。但是,由于在很多实际问题当中,那些重要的特征常常掩盖在纷繁芜杂的表象之下,不能够直接获得,或者受条件限制不能对它们进行测量,使得特征选择和提取的任务复杂化,成为构造模式识别系统、提高决策精度的最困难的任务之一。

目前特征选择与提取广泛应用在社会生活和科学研究的各个领域。

(1) 农产品品质检测

农产品在其生产过程中由于受到人为和自然等复杂因素的影响,产品品质差异很大,如形状、大小、色泽变化等。使用特征选择与提取技术对农产品的外部品质特征进行提取进而实现农产品的自动分级是目前许多科学工作者研究的重要课题。目前国内外利用特征选择与提取技术进行农产品品质自动识别研究的对象极其广泛,如谷粒的表面裂纹检测和农作物种子的分级、根据梅、黄瓜、玉米、竹笋、西红柿、辣椒、苹果等的大小、形状、颜色和表面特征等进行分级等。



(2) 医学图像分析

自从发现伦琴射线以来,医学成像技术的不断进步,相继出现了 CT、MRI 成像技术,给医生带来越来越丰富的图像信息,也加深了人类对自身的认识和理解。随着数字图像处理和分析在理论和技术上的逐渐完善,以及数字医学影像技术的发展,对数字医学图像的处理和分析成为国内外研究的热点。研究者使用特征选择与提取技术降低数字医学图像特征维数,提取出感兴趣区域特征,辅助医生判读医学图像,提高诊断的准确性和效率。研究的对象包括钼靶 X 线图像、CT 图像、MRI 造影等。

(3) 工业探伤

金属材料在锻造、焊接过程中,会形成各种内部缺陷,它们的存在和扩大会导致设备使用寿命缩短、安全可靠降低。使用超声检测探伤时,定性(分类)比较困难,判断准确率受人为影响较大。目前一些研究者使用 LabView 建立能对缺陷进行识别的智能系统,利用特征选择与提取技术对探伤缺陷进行分类识别。桥梁裂纹损伤识别是及时发现桥梁裂纹保证桥梁安全作业的重要依据,应用特征选择与提取技术识别桥梁裂纹是目前许多学者研究的重要课题。

(4) 图像数据库检索

近年来,随着信息技术的发展,数字图像的数量正在飞速增长。如何在海量的数字图像中找到我们所需要的图像数据越来越被人们所关注。基于内容的图像检索成为当前热门的研究课题。它是一门有关图像特征相似性匹配的新技术,需要深入研究图像特征描述方法、特征提取与索引方法,以及相似性度量与快速检索算法等问题,便于以图像中包含的各类典型特征(即图像内容本身)为客观检索对象,避免以文本为主观标识所造