

Broadview
WWW.BROADVIEW.COM.CN

数据仓库与数据挖掘
技术应用丛书

更通俗的机器学习方法介绍
更广阔的数据挖掘应用视角
更丰富的机器学习与数据挖掘的解决方法

Machine Learning and Data Mining: Methods and Applications

机器学习与数据挖掘： 方法和应用

[美] Ryszard S. Michalski Ivan Bratko Miroslav Kubat 等著

朱明 等译

 **WILEY**



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

<http://www.phei.com.cn>

数据仓库与数据挖掘技术应用丛书

机器学习与数据挖掘：方法和应用

Machine Learning and Data Mining: Methods and Applications

[美]Ryszard S. Michalski Ivan Bratko Miroslav Kubat 等著

朱明 等译

電子工業出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书分为5个部分,共18章,较为全面地介绍了机器学习的基本概念,并讨论了数据挖掘和知识发现中的有关问题及多策略学习方法,具体地阐述了机器学习与数据挖掘在工程设计,文本、图像和音乐,网页分析、计算机病毒和计算机控制,医疗诊断、生物医疗信号分析和水质分析中的生物信号处理等方面的应用情况。

本书收集众多不同领域中数据挖掘的实际案例,以此来说明数据挖掘的具体解决方法,以期为广大读者提供一个更为广阔的数据挖掘应用视角。

本书的读者,可以是任何对机器学习与数据挖掘感兴趣的工程技术人员、业务管理人员,或是从事具体技术工作的其他人员。本书也可作为大专院校相关课程的重要辅导教材。

Copyright©1998 by John Wiley & Sons Ltd.

All rights reserved. Authorized translation from the English language edition published by John Wiley & Sons, Inc.

本书简体中文专有翻译出版权由John Wiley & Sons Inc. 授予电子工业出版社。未经许可,不得以任何方式复制或抄袭本书的任何部分。

版权贸易合同登记号: 图字: 01-2002-6435

图书在版编目(CIP)数据

机器学习与数据挖掘: 方法和应用/ (美) 米哈尔斯基 (Michalski, R.S.) 等著; 朱明等译. —北京: 电子工业出版社, 2004.1

(数据仓库与数据挖掘技术应用丛书)

书名原文: Machine Learning and Data Mining: Methods and Applications

ISBN 7-5053-9224-7

I. 机… II. ①米… ②朱… III. ①机器学习 ②数据采集 IV. ①TP18 ②TP274

中国版本图书馆 CIP 数据核字 (2003) 第 090444 号

责任编辑: 孙学瑛

印 刷: 北京增富印刷有限公司

出版发行: 电子工业出版社 <http://www.phei.com.cn>

北京市海淀区万寿路 173 信箱 邮编 100036

经 销: 各地新华书店

开 本: 787×980 1/16 印张: 28 字数: 625 千字

版 次: 2004 年 1 月第 1 版 2004 年 1 月第 1 次印刷

印 数: 4 000 册 定价: 58.00 元

凡购买电子工业出版社的图书,如有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系。联系电话:(010) 68279077。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。

出版说明

如果没有对海量数据进行科学分析的能力，沃尔玛的老板再精明，也绝对想不到“啤酒与尿布”这两个风马牛不相及的东西之间还有着千丝万缕的联系，而将它们放在一起，竟然增加了啤酒销量，可见数据分析的巨大威力。

信息系统数年中收集了海量数据，且数据还正以指数级增长，企业迫切地需要高效、精确、科学地分析数据，以找出其背后的寓意，进而了解企业的经营状况和外部环境，做出科学的决断，在现代激烈的竞争中胜出。所以，如何将数据点石成金，更是摆在我们面前很现实也很诱人的一个问题。

现在，很多人已经意识到数据中潜在的大量商机，并踏踏实实地进行着从数据中沙里淘金的工作。特别是在信息化的大潮中，上至政府，下到企业，从银行到电信，再到网站、超市，人们都希望用数据分析这根魔杖赢得先机。与此同时，人们也在企盼着相关书籍，以便工作中学习参考。在广泛征询专家和用户的基础上，秉着选题全面、内容经典、译者严谨的原则，我们适时地推出了这套《数据仓库与数据挖掘技术应用丛书》，以飨读者。本丛书有如下几本：

- 数据仓库基础
- OLAP 解决方案：多维信息系统的构建技术
- 数据仓库工具箱：维度建模的完全指南（第二版）
- 数据仓库生命周期工具箱：设计、开发和部署数据仓库的专家方法
- 数据仓库及其在电信领域中的应用
- 疑难数据仓库专家解决方案
- IBM 数据仓库和商业智能工具
- 可视化数据挖掘：数据可视化和挖掘的技术和工具
- 点击流数据仓库
- Web 数据挖掘：将客户数据转化为客户价值
- 企业信息工厂
- 机器学习与数据挖掘：方法和应用

本丛书既包括商业智能（BI）的基础——数据仓库（DW），也包括数据仓库上的两类不同目的的数据增值操作——联机分析处理（OLAP）和数据挖掘（DM）；既覆盖基础理论，如数据仓库基础，又提供不同领域的解决方案，如数据仓库在电信、银行、保险等领域的应用。

本丛书来自国外数据库领域一些著名作者的畅销书，以及国内第一线实施者的精心总结。如一直位居 AMAZON 畅销书榜的数据仓库领域的畅销书作家 Ralph Kimball 的《数据仓库工具箱：维度建模的完全指南（第二版）》、《数据仓库生命周期工具箱：设计、开发和部署数据仓库的专家方法》，数据仓库之父 William H.Inmon 的《企业信息工厂》等。

丛书的译者均来自工作在该领域一线的人员，既有该领域的理论和实践经验，又具备中英文翻译的功底。且多位译者先前均已读过原著，所以，自感翻译的过程不再是枯燥，而是情趣盎然，乐在其中。

出版高品位、高品质的图书是博文视点的努力目标。希望您对我们的工作多提宝贵意见。您的意见是我们创造精品的动力源泉。

如果您希望将您的工作经验感悟等总结成书，我们将为您提供一流的服务，共创精品图书。

我们的联系方式如下：

地址：北京复兴路 47 号天行建商务大厦 604

邮编：100036

电话：010-51922832，68216158

传真：010-51922823

E-mail: jsj@phei.com.cn; zsh@phei.com.cn

博文视点资讯有限公司

2003 年 10 月

博文视点资讯有限公司 (BROADVIEW Information Co.,Ltd.) 是信息产业部直属的中央一级科技与教育出版社——电子工业出版社 (PHEI) 与国内最大的 IT 技术网站 CSDN.NET 和最具专业水准的 IT 杂志社《程序员》合资成立的以 IT 图书出版为主业、开展相关信息和知识增值服务的资讯公司。

我们的理念是：创新专业出版体制；培养职业出版队伍；打造精品出版品牌；完善全面出版服务。

秉承博文视点的理念，博文视点的产品线为面向 IT 专业人员的出版物和相关服务。博文视点将重点做好以下工作：

- (1) 在技术领域开发专业作（译）者群体和高质量的原创图书
- (2) 在图书领域建立专业的选题策划和审读机制
- (3) 在市场领域开创有效的宣传手段和营销渠道

博文视点有效地综合了电子工业出版社、《程序员》杂志社和 CSDN.NET 的资源和人才，建立全新专业的立体出版机制，确立独特的出版特色和优势，将打造 IT 出版领域的著名品牌，并力争成为中国最具影响力的专业 IT 出版和服务提供商。

作为合资公司，博文视点的团队融合了各方面的精英力量：原电子工业出版社 IT 图书专业出版实力的代表部门——计算机图书事业部的团队；《程序员》杂志社和 CSDN 网站的主创人员；著名 IT 专业图书策划人周筠女士及其创作群。这是一个整合专业技术人员和专业出版人员的团队；这是一个充满创新意识和创作激情的团队；这是一个不断进取、追求卓越的团队。

电子工业出版社与《程序员》杂志和 CSDN 网站的合作以最有效率的方式形成了出版资源、媒体资源、网络资源的整合和互动，成为 2003 年 IT 出版界备受瞩目的事件。

“技术凝聚实力，专业创新出版”，BROADVIEW 与您携手共迎信息时代的机遇与挑战！



博文视点

地址：北京市复兴路 47 号天行建商务大厦 604 室

邮 编：100036

总 机：010-51922832 传 真：010-51922823

本版编辑部：010-51922839 外版编辑部：010-51922825

<http://www.broadview.com.cn> 投稿及读者反馈：editor@broadview.com.cn

译者序

随着数据库技术的应用越来越普及，人们正逐步陷入“数据丰富，知识贫乏”的尴尬境地。在此背景下，数据挖掘技术，又称数据库知识发现，于20世纪90年代开始迅速兴起。这一信息领域是基于机器学习、统计分析等多种学科的计算机技术的，它能够有效帮助人们将巨大的数据资源转换为有用的知识与信息资源，进而可以帮助人们科学地做出各种决策。

机器学习是一个有关对学习过程中的计算方法的研究，以及如何应用基于计算机的学习系统解决实际问题的学科领域。机器学习中一个重要的研究内容就是对于从样本中获取相应概念描述方法的研究。因此许多机器学习方法可直接用来解决数据挖掘问题。而数据挖掘问题就是从大规模数据库中搜索出有趣模式和重要规律的问题。

本书主要是为广大机器学习或数据挖掘的非专业人员，却对机器学习或数据挖掘应用及其入门知识感兴趣的读者而编写的。因此本书收集众多不同领域中数据挖掘的实际案例，以此来说明数据挖掘的具体解决方法，期望能为广大读者提供一个更为广阔的数据挖掘应用视角。而这一点正是本书与其他数据挖掘书籍的最大区别。

本书分为5个部分，共18章，较为全面地介绍数据挖掘在众多领域和实际问题求解中的应用情况。本书第1部分，共4章，分别介绍了机器学习的基本概念，并讨论了数据挖掘和知识发现中的有关问题及多策略学习方法。此外第1部分还介绍了机器学习与归纳逻辑编程的一些实际应用案例。

本书第2部分，共4章，主要介绍机器学习与数据挖掘在工程设计方面的应用情况。具体内容包括：有限元设计问题求解中的机器学习方法；工业机器故障检测中的基于事例推理方法；设备部件装配规划求解的多策略方法；防摩擦轴承系统中的归纳学习方法。

本书第3部分，共3章，主要介绍机器学习与数据挖掘在文本、图像和音乐方面的应用情况。具体内容包括：文本之间相互关系的挖掘方法；行李X光片中引爆雷管图像的模式识别方法；音乐表达规律的数据挖掘方法。

本书第 4 部分, 共 4 章, 主要介绍机器学习与数据挖掘在网页分析、计算机病毒和计算机控制方面的应用情况。具体内容包括: 网页搜索中网页分类识别的机器学习方法; 计算机病毒自动分析识别中的机器学习方法; 控制技术中行为复制的机器学习方法; 空中交通控制中的数据挖掘方法。

本书第 5 部分, 共 3 章, 主要介绍机器学习与数据挖掘在医疗诊断、生物医学信号分析和水质分析中生物信号处理方面的应用情况。具体内容包括: 机器学习在医学诊断中的应用; 生物信号分析中的机器学习方法; 机器学习在河流水质生物分类中的应用。

本书内容从问题描述出发, 着重介绍各个问题的实质, 如何利用机器学习与数据挖掘来解决相关问题, 以及所介绍的机器学习与数据挖掘方法的基本特点等, 力求深入浅出, 以便能够为广大非机器学习与数据挖掘专业的读者, 提供一个了解机器学习与数据挖掘的应用视角。本书立足应用, 介绍相关的机器学习与数据挖掘方法, 因此更加通俗易懂。此外本书所涵盖的大量机器学习与数据挖掘实际案例, 也会为读者联系自己的实际问题, 提出有效的基于机器学习与数据挖掘的解决方法, 奠定了较好的基础。

本书的读者, 可以是任何对机器学习与数据挖掘感兴趣的工程技术人员、业务管理人员, 或者是从事具体技术工作的其他人员。本书也可作为大专院校相关课程的重要辅导教材。

本书的翻译工作得到了吴炜、徐骞、王庆伟、顾智宇、钟捷飞、李靖、黄永刚、黄科、朱磊、殷俊、黄振等许多同志的积极协助, 他们帮助完成了本书部分内容的翻译与校对工作, 在此向他们表示感谢。此外, 我妻子王晓岚帮助完成了本书部分内容的录入工作, 特此致谢。

由于翻译时间仓卒, 加上本书涉及诸多实际应用领域, 专业用语和词汇较多, 本书翻译内容可能会存在一些问题, 敬请读者谅解, 并批评指正。

译 者

2003.08.08 于合肥

前 言

本书的读者对象不一定是机器学习或数据挖掘方面的专家，而是对机器学习或数据挖掘的多种应用及其入门知识感兴趣的读者。机器学习是一个致力于有关学习过程中计算方法的研究，以及应用计算机学习系统解决实际问题的领域。机器学习中一个重要的研究内容就是从样本中获取相应的概念描述的方法。样本表示可以采用多种形式，尤其是可以采用二维关系数据表形式来表示，因此许多机器学习方法可被直接应用于解决数据挖掘问题，并起着重要的作用。数据挖掘问题就是从大规模数据库中搜索出有趣模式和重要规律。而在数据挖掘问题中，具有挑战性且目前存在许多尚待解决问题的领域，则是有关从文本、图像或声音序列（音乐）之中抽取出模式或规则的问题。本书有一部分专门讨论这类问题的有关应用情况。

本书第 1 部分介绍了机器学习的基本概念，并讨论了数据挖掘和知识发现中的有关问题及多策略方法。这部分还介绍了机器学习与归纳逻辑编程的若干应用，其中后者是机器学习的当前一个分支领域。

随后各个章节是由若干应用主题组成的。这些主题包括：设计与工程，文本、图像和音乐中模式或规则的发现，计算机与控制系统及医药与生物。本书中所介绍的应用中有一些是已投入使用的，而另一些则是研究性应用。本书收集应用实例的指导性标准就是：机器学习方法必须应用于一个困难而有意义的现实世界问题中，并产生满意的或至少是真正有前景的结果。

本书是由一组国际科学家实际编写而成的。每个章节均有单独的作者，他们是机器学习或相关领域的主要专家，分别代表了来自 11 个国家的主要研究小组，这些国家分别是：澳大利亚、奥地利、比利时、加拿大、法国、以色列、韩国、波兰、斯洛文尼亚、英国和美国。然而编者有必要指出：由于现在有如此多的优秀研究人员在机器学习和数据挖掘领域工作，因此不可能挑选出特别“具有代表性”的作者。许多被邀请为本书撰稿的人士目前或以前就是本书编者的同事或合作者。然而被选中的作者都是杰出的科学家，他们的工作涵盖了本书主题的很大范围，因此编者能够将他们所贡献的内容编入到本书中而感到非常荣幸。

编者借此机会希望感谢所有为本书工作的人们。特别要感谢 George Mason 大学机器学习和推理实验室的研究人员,他们在各种技术和文稿方面提供了无价的帮助,特别是 Zoran Duric, Ken Kaufman, Seokwon Lee 和 Qi Zhang。我们还要感谢 John Wiley & Sons 公司的 Roslyn Meredith 和 Gaynor Redvers-Mutton, 他们在本项目中也提供了密切的合作。

最后,我们希望读者们将会从本书中得到有益的帮助和指导,并能够将它作为有关机器学习与数据挖掘及其无数实际应用的有价值的信息源。

Ryszard S. Michalski

Ivan Bratko

Miroslav Kubat

目 录

第 1 部分 基本概念

第 1 章 机器学习方法概述	(2)
1.1 导论	(2)
1.2 机器学习任务	(4)
1.2.1 认知观点	(5)
1.2.2 表示问题	(7)
1.3 泛化空间的搜索	(11)
1.3.1 学习的归纳本质	(11)
1.3.2 穷尽搜索	(13)
1.3.3 启发式搜索	(14)
1.4 学习经典任务	(16)
1.4.1 分而治之学习法	(16)
1.4.2 主动覆盖: AQ 学习	(24)
1.4.3 学习算法评估	(27)
1.5 如何利用谓词逻辑	(29)
1.5.1 从关系中学习 Horn 子句	(30)
1.5.2 反转归并	(34)
1.5.3 理论修正	(36)
1.5.4 构造归纳	(38)
1.6 人工发现	(40)
1.6.1 概念形成	(41)
1.6.2 寻找自然定律	(46)
1.6.3 动态系统的发现	(49)
1.7 如何处理搜索空间过大	(50)
1.7.1 类比提供搜索启发	(50)
1.7.2 基于示例学习	(51)
1.8 机器学习的近邻	(53)

1.8.1	神经网络	(53)
1.8.2	遗传算法	(55)
1.9	混合系统与多策略学习	(57)
1.9.1	熵网络	(58)
1.9.2	基于知识的神经网络	(59)
1.9.3	AQ 泛化中的遗传搜索	(60)
1.9.4	GA 与神经网络的结合	(61)
1.10	展望	(61)
	参考文献	(62)
第 2 章	数据挖掘与知识发现: 对问题和多策略方法的回顾	(65)
2.1	前言	(65)
2.2	机器学习与多策略数据分析	(67)
2.2.1	从具体实例中抽取通用规则	(68)
2.2.2	概念聚类	(72)
2.2.3	构造性归纳	(73)
2.2.4	选择最有代表性的样本	(74)
2.2.5	定性与定量结合的发现	(75)
2.2.6	定性预测	(75)
2.2.7	基于机器学习方法的总结	(77)
2.3	数据分析任务中的分类	(78)
2.4	INLEN 中各操作的集成	(81)
2.5	聚类和学习操作的说明	(84)
2.6	数据与规则的可视化	(86)
2.7	结构属性的规则学习	(89)
2.8	从决策规则中学习决策结构	(91)
2.9	表示空间的自动改善	(93)
2.9.1	确定最相关的属性	(93)
2.9.2	新属性的产生	(94)
2.10	应用展示: 经济与人口统计数据中的发现	(94)
2.10.1	背景	(94)
2.10.2	实验 1: 多操作的集成	(95)
2.10.3	实验 2: 子群中的异常识别	(96)
2.10.4	实验 3: 利用结构属性	(97)
2.10.5	实验 4: 利用构造性归纳运算操作	(99)
2.11	总结	(100)

参考文献	(101)
第3章 机器学习在多个领域的应用	(102)
3.1 前言	(102)
3.2 规则归纳在多个领域中的应用	(103)
3.2.1 提高化工过程控制中的产量	(103)
3.2.2 信用评估决策	(104)
3.2.3 机械设备故障诊断	(105)
3.2.4 天体对象的自动分类	(106)
3.2.5 监测旋转乳液的质量	(107)
3.2.6 减少照排印刷时的条纹现象	(107)
3.2.7 改善油气分离质量	(108)
3.2.8 预防电力变压器故障	(109)
3.2.9 规则归纳在其他领域的应用	(110)
3.3 规则归纳的其他应用研究	(110)
3.3.1 填表工作的自动化	(111)
3.3.2 支持知识库维护	(111)
3.3.3 航天飞机引擎的测试	(112)
3.3.4 严重暴风雨的预报	(112)
3.3.5 直升机叶片的修理	(112)
3.3.6 预测蛋白质结构	(113)
3.3.7 钢厂调度自动化	(113)
3.3.8 更多应用及其相关方法	(113)
3.4 若干策略和经验	(114)
3.4.1 问题的明确描述	(114)
3.4.2 确定表示方法	(115)
3.4.3 训练数据的收集	(115)
3.4.4 评估学习获得的知识	(116)
3.4.5 知识库的具体应用	(116)
3.4.6 机器学习应用的效能来源	(117)
参考文献	(118)
第4章 归纳逻辑编程的应用	(120)
4.1 前言	(120)
4.2 ILP方法与其他机器学习方法的比较	(122)
4.3 预测化合物的诱变性	(123)
4.4 放电机器中的技能重建	(125)

4.4.1 表示方法的设计	(125)
4.4.2 学习结果和专家评估	(126)
4.5 ILP 的一些其他应用	(128)
4.6 总结	(130)
参考文献	(131)

第 2 部分 设计与工程

第 5 章 机器学习在有限元计算中的应用	(134)
5.1 简介	(134)
5.2 向 FEM 产生器添加一个专家系统	(136)
5.3 学习问题、实例和背景知识	(137)
5.3.1 问题的关系特性	(137)
5.3.2 实例来源	(137)
5.3.3 正面实例	(138)
5.3.4 反面实例	(139)
5.3.5 背景知识	(139)
5.3.6 学习集概要	(141)
5.4 以前的实验	(142)
5.4.1 GOLEM 的实验	(142)
5.4.2 FOIL 的实验	(143)
5.4.3 mFOIL 的实验	(144)
5.4.4 CLAUDIEN 的实验	(144)
5.4.5 MILP 的实验	(144)
5.4.6 FOSSIL 的实验	(145)
5.4.7 属性值算法的实验	(145)
5.5 选择一个合适的学习算法	(145)
5.6 根据 CLAUDIEN 学习	(147)
5.7 归纳的规则的后处理	(150)
5.8 结果	(152)
5.8.1 知识库与 ES Shell	(152)
5.8.2 对专家系统的评价	(153)
5.9 总结	(156)
参考文献	(157)

第 6 章 归纳学习和基于事例的推理在工业机器故障检测方面的应用	(159)
6.1 简介	(159)
6.2 归纳学习与基于事例的推理	(160)
6.3 更好地利用经验	(162)
6.4 应用	(162)
6.4.1 CFM 56-3 引擎的故障检测	(163)
6.4.2 机器人轴心的故障检测	(165)
参考文献	(168)
第 7 章 经验装配序列规划: 多策略构造学习方法	(170)
7.1 前言	(170)
7.2 NOMAD 中的表示与规划	(172)
7.3 多策略构造学习	(176)
7.4 NOMAD 的学习场景	(177)
7.5 与先前研究进行比较	(181)
7.6 结束语	(183)
参考文献	(184)
第 8 章 归纳学习设计入门: 关于防摩擦轴承系统的设计方法和实例研究	(186)
8.1 导论	(186)
8.2 一种学习设计规则的方法	(187)
8.2.1 概述	(187)
8.2.2 学习规则集的经验性错误	(188)
8.2.3 应用已学习到的规则处理新例子	(189)
8.3 一个示范问题的描述	(189)
8.4 归纳方法的应用	(191)
8.4.1 变量的定性值	(192)
8.5 训练与事件测试	(193)
8.5.1 设计知识源	(193)
8.5.2 样本数据库	(194)
8.6 结果分析	(194)
8.6.1 从训练样本中学习规则	(195)
8.6.2 以递增学习方法评估预备样本	(196)
8.6.3 得到结果的可信度	(197)
8.7 总结	(199)
参考文献	(200)

第 3 部分 文本、图像和音乐模式的测定

第 9 章 找出文本之间的关联	(202)
9.1 介绍	(202)
9.2 FACT 系统结构	(204)
9.3 关联	(207)
9.4 查询语言	(208)
9.5 查询操作	(210)
9.6 关系表达式	(213)
9.7 对新闻数据运用 FACT 系统	(213)
9.8 总结	(216)
参考文献	(217)
第 10 章 学习图像中的模式	(220)
10.1 导论	(220)
10.2 计算机视觉中机器学习的研究工作	(221)
10.3 室外场景彩色图像的语义解释	(224)
10.3.1 MIST 方法	(224)
10.3.2 实现和实验结果	(226)
10.4 检查行李 X 光图像中的引爆雷管	(229)
10.4.1 预备知识	(229)
10.4.2 问题描述	(231)
10.4.3 方法和实验结果	(232)
10.5 视频图像序列中的动作识别	(234)
10.5.1 来自动作的功能	(234)
10.5.2 运动的计算	(236)
10.5.3 实验	(238)
10.6 结论与未来的研究	(242)
10.6.1 室外场景彩色图像的语义解释	(242)
10.6.2 行李 X 管图像中的引爆雷管检测	(242)
10.6.3 识别视频图像序列中的动作	(242)
10.6.4 在视觉系统中结合学习的优点	(243)
参考文献	(244)
第 11 章 机器学习在音乐研究领域的应用: 深入音乐 表达现象的经验调查	(246)
11.1 介绍	(246)

11.2	学习对象：富有表现力的音乐演奏	(248)
11.3	背景知识的特性和价值	(248)
11.4	方法一：在音乐符号的层次上学习	(250)
11.4.1	目标概念	(251)
11.4.2	定性的领域理论	(251)
11.4.3	IBL-SMART 学习算法	(254)
11.4.4	实验	(255)
11.5	方法二：在结构层次上学习	(258)
11.5.1	实验	(260)
11.6	对真实艺术的演奏的一次机器学习分析	(263)
11.7	实验结果的讨论	(266)
11.7.1	定量的分析	(267)
11.7.2	对于音乐理论有用的定性结果	(269)
11.8	总结	(269)
	参考文献	(270)

第 4 部分 计算机系统和控制系统

第 12 章	网页哨兵：万维网页学习者	(274)
12.1	概述	(274)
12.2	网页哨兵	(274)
12.3	学习	(280)
12.3.1	该学些什么	(280)
12.3.2	怎样描述 Pages, Links 和 Goals	(280)
12.3.3	应该用什么样的学习方法	(282)
12.4	实验结果	(283)
12.4.1	UserChoice? 能学习到多精确的程度	(283)
12.4.2	牺牲覆盖率能改进准确率吗	(285)
12.5	总结	(286)
	参考文献	(287)
第 13 章	计算机病毒的生物启发式防御	(288)
13.1	介绍	(288)
13.2	背景	(289)
13.2.1	计算机病毒	(289)
13.2.2	病毒的检测、清除和分析	(290)