

 现代计算机科学与技术教材系列

数据挖掘与数学建模

SHUJU WAJUE YU SHUXUE JIANMO



廖 芹 郝志峰 陈志宏 编著



国防工业出版社
National Defense Industry Press

现代计算机科学与技术教材系列

数据挖掘与数学建模

廖芹 郝志峰 陈志宏 编著

国防工业出版社

·北京·

内 容 简 介

本书以国际数据挖掘标准流程(GRISP-DM)为依据,以企业管理面临的现实问题为应用案例,由浅入深介绍数据挖掘方法及其解决问题过程的数据理解、数据处理、数据分析、数学建模、模型结果评估等内容,并引入应用广泛的数据挖掘 Clementine 软件辅助问题案例的解决,使读者不仅可以集中地学习数据挖掘的主要理论方法,而且可以了解基于数据挖掘的数学建模过程,可以学习应用软件辅助解决问题的操作方法。

本书把理论、案例、建模、软件辅助结合一体统一叙述,简述理论,突出应用,详细分析,展示过程,既考虑高校学生的学习需要,分本科生与研究生学习层次,又考虑企业管理者的应用与实践需要。

本书可作为数据挖掘理论与技术的教学、实践、应用和提高的教科书或参考书。适合高等学校本科高年级学生、研究生以及学习数据挖掘、数学模型课程的学生使用,也适合相关的企业管理与决策支持技术人员使用。

图书在版编目(CIP)数据

数据挖掘与数学建模/廖芹,郝志峰,陈志宏编著. —北京:国防工业出版社,2010.2
(现代计算机科学与技术教材系列)
ISBN 978-7-118-06671-5

I. ①数... II. ①廖... ②郝... ③陈... III. ①数据采集-数学模型-高等学校-教材 IV. ①TP274

中国版本图书馆 CIP 数据核字(2009)第 242949 号

※

国防工业出版社 出版发行

(北京市海淀区紫竹院南路 23 号 邮政编码 100048)

北京奥鑫印刷厂印刷

新华书店经售

*

开本 710×960 1/16 印张 23¼ 字数 483 千字

2010 年 2 月第 1 版第 1 次印刷 印数 1—4000 册 定价 39.00 元

(本书如有印装错误,我社负责调换)

国防书店:(010)68428422

发行邮购:(010)68414474

发行传真:(010)68411535

发行业务:(010)68472764

《现代计算机科学与技术教材系列》

编审委员会

主	编	刘椿年	许满武		
执行副主编		陈国良	赵致琢		
编	委	王志英	康立山	刘大有	刘椿年
		许满武	宋方敏	林作铨	陈向群
		李师贤	陈国良	洪先龙	应明生
		陆汝占	傅育熙	赵致琢	唐常杰
		张明义	彭群生	朱 洪	周龙骧
		李立武	承恒达	李西宁	陈 溧
		阎瑞琪			

总 序

民众多好饮酒,中外概莫能外。酒馆和酿酒坊伴随饮酒客而起,人类对酒的喜爱造就了酒文化和一个庞大的产业。好酒能卖好价钱,能使文人诗兴大发,催生佳作,还能解人间百难。于是,酿天下名酒自然成为不少人的毕生追求。

怎样才能酿出好酒呢?国人的看法不尽相同。崇信洋酒的人主张引进国外的生产工艺,学习洋人的生产和经营理念,而喜欢国酒的人则主张走自己的路,但不排除借鉴国外先进的科学技术和管理经验。这样的争论或许永远不会终结,但外国人重视科学酿酒,这一点是值得我们学习和借鉴的。

计算机科学教育,如同酿酒工业的生产一样,科学办学迄今还只是部分学者的一种理想。与国内一样,国外的计算机科学教育并没有像他们的科学酿酒业一样,实现科学办学。也许科学办学要远比科学酿酒困难得多。譬如,怎么实现科学办学?甚至怎么推出一套科学的系列教材都是一篇大文章。

这套教材的创作始于教育部面向 21 世纪教育与教学改革 13-22 项目的研究。2000 年,在 13-22 项目研究工作即将完成之际,一些学者开始认识到面对计算机科学与技术的高速发展,我们亟需一套体现科学办学思想、反映内涵发展要求、服务教育与教学改革、参与构建学科人才培养科学体系的系列教材。强调系列教材是因为那时已经意识到计算机科学教育本质上是一项科学活动,但长期以来教师向学生传授科学技术知识的方式方法科学性不强。由于高等教育几百年来一直沿袭经验方式而非科学方式办学,大学教学的方式方法仍然还停留在古代作坊式的阶段,只不过今天使用的教学技术手段先进而已。在经验办学方式下,无论是研究型大学还是教学型大学,由于种种原因,教学活动的全过程存在着太多的漏洞和质量上的隐患。科学办学是对高等教育界传统的一个挑战,尽管在认识上,人们不难理解,科学办学是经验办学的最高形式,而经验办学应该成为科学办学的有益补充。

13-22 项目组积极探索,率先倡导科学办学理念,初步构建了一个体现科学办学思想,反映内涵发展要求的计算机科学一级学科人才培养科学理论体系,为学科专业教育探索新天地,走向科学办学和发展学科系列教材提供了一个认知基础。

长期以来,学术界一直在探索计算机专业教育的规律。ACM 和 IEEE/CS 的

专家小组在走访了全美 400 多位著名计算机科学家的基础上,以学科方法论作为切入点开展教学改革理论研究,于 1989 年发表了具有开创性意义的成果,尽管他们并未意识到自己的工作是以学科方法论的研究作为切入点,探讨内涵发展的道路。1990 年前后,在迷宫中探索行走的专家小组经大师和精英群的指点实际上已经摸到了走出迷宫的大门,却没有打开并进入一个崭新的天地。这一点从他们在 2000 年网上公布的 CC2001 报告最先删除了 CC1991 报告中有关学科方法论的内容便不难看出(注:后经中国人的提醒又补充写入)。

与此同时,中外教材建设也一直没有停止探索,国内外出版社先后出版了种类繁多的计算机类专业教材。这些教材中不乏精品和上乘之作,但难觅具有鲜明特色、真正一体化设计并且符合科学办学要求的系列教材。多数丛书和系列教材基本上还只是出版社对出自作者个人创作的教材通过冠名“丛书”或“系列”的方法结集出版以求强势效应。尽管如此,不少优秀作者和学者理所当然地进入了编审委员会的视线。西方发达国家在计算机科学学科的领先优势曾使许多人不自觉地将目光转向海外,试图从世界名牌大学使用的教材中去寻找蓝本。遗憾的是与国内一样,经验办学并没有使西方大学在教材建设方面摆脱“各自为政,各行其是”的阴影。此时,我们如梦初醒,毕竟科学办学是前无古人的一项创举。随着学科的不断深入,在迈向深蓝知识海洋的今天,外国人未必比中国人在科学办学方面占有更多天时地利的优势。不经意中的发现使我们惊喜和激动,同时深感责任重大且平添担忧:即使能够写出系列教材的一体化设计,我们是否真能确认这项改革的正确性?真能推出科学的系列教材?可是,除了实践和试验,我们别无捷径可循。

从 20 世纪 50 年代起,我国几代学者苦苦追赶了西方发达国家半个世纪,依靠引进、学习、跟踪、改进、创新的高新技术发展思维定式曾使我们付出了高昂的学费和沉重的代价。固然,在高新技术领域,依靠“引进”和“泊来”,取得了长足的发展和进步,填补了不少国内的“空白”,但在水准上始终与发达国家保持着一段距离,一种在行业内部看来时长时短、难于逾越的差距。这种差距主要表现在对高精尖学科的发展我们缺乏思想、概念、理论、方法、技术、制度、规范和设计的原始创新和发展模式的全面创新,研究工作总是跟在别人后面亦步亦趋。我们缺乏在发展中另辟蹊径,走自己道路的机制和氛围,迷信洋人,盲目追随西方学术发展道路的习惯思维方式几乎导致国人失去了创新的机能,这是一个国家和民族发展高新技术学科和产业致命的硬伤。

高新技术领域竞争的成败,关键取决于人才与文化。现代科学技术的创新,已不单纯是一个学术问题,还是一个与文化、人文密切相关的问题。科学教育求真求是,技

术教育求实求精,人文教育求灵求善,艺术教育求美求新。没有科学技术知识,人的认识和生活难免停留在原始社会,而没有人文精神和艺术的陶冶,科学技术的创新必然失去力量的源泉。可见,走自己的道路,发展中国的科技创新体系,在某种程度和意义上,成败的关键在于大学能否真正培养一大批高素质的人才。高等学校要实现培养大量高素质计算机专业人才的目標,需要在前进中不断地进行系统的、科学的总结和深刻的反省,需要对遇到的问题进行科学的分析和判断,作出正确的决策。

工欲善其事,必先利其器。倘若教师不能在思想上摆脱陈旧的思维定式,用先进的理念武装头脑,勇于探索前人没有走过的发展道路,那么,即使采用了世界一流大学的全套教材,恐怕也难于培养出世界一流的人才。中西文化、人文传统之间的差异之大,中外教育思想之间的差异之大使得中国教育的现代化决不是一个引进和模仿就可以轻易解决的问题。教师的职业不是贩卖知识。授业、传道、启蒙、解惑技能的高低,不仅取决于教师知识的广博和深厚,更重要的在于远见、卓识、探索、创新、敬业、求真的本领和身先垂范。

身处 21 世纪,面对国家的期望,处在科学技术发展浪潮之巅的计算机科学教师,任重道远。我们就像茫茫林海中的探险者,环顾苍翠的群山,犹如身陷迷宫一样。计算机科学教育,敢问路在何方?其实,我们的出路或许只有一条,那就是系统总结前人的经验和教训,设法努力登上山峰,居高眺望,探寻走出林海的希望之路。

2001 年初,《现代计算机科学与技术教材系列》编审委员会正式成立,一年后,计算机科学一级学科系列教材一体化设计研究报告的第一部分在《计算机科学》杂志第 6 期长篇发表。编审委员会为系列教材的出版制定了严格、详细的操作程序,并通过多种渠道,选择了第一批教材的作者。在体制创新方面本次教材出版设立了学术编审人,跟踪编审教材的创作内容,力求教材的尽善尽美。可以预期,《现代计算机科学与技术教材系列》将是一套基于计算机科学一级学科人才培养科学理论体系,体现内涵发展和科学办学要求,反映一体化设计的系列教材建设的一个尝试。然而,就像任何新生事物一样,她难免存在缺点和不足,我们诚恳地希望关心和使用本套系列教材的师生、读者,在使用中将批评或建议留下来,帮助改进教材建设工作,修正存在的不足之处。

一些学者对于一级学科人才培养科学理论体系的能行性表示怀疑。带着这个问题,在贵州大学和国内部分高等学校和一大批知名学者的支持下,从 1999 年夏天起在贵州大学连续举办了“计算机科学与技术高级研讨班”,向(博士)研究生和中青年教师陆续开设了研究生核心学位课程“高等计算机体系结构”、“并行算法设计基础”、“分布式算法设计基础”、“高等逻辑基础”和“形式语义学基础”,后来又进一步开设

了本科生重点课程“算法设计与分析”、“数理逻辑基础”、“可计算性与计算复杂性”、“形式语言与自动机理论”等一系列课程。5年来,高级研讨班受到全国广泛关注、响应和支持,先后吸引了几十所大学三百多人次的师生参加听讲和学习,目前已经发展为由教育部批准资助、15所大学联合主办的高级研讨班。实践证明,高级研讨班为中国高等学校计算机科学教学改革和教育质量的提高正在发挥其独特的作用,并得到国内外一大批学者的充分肯定和好评。高级研讨班正在成为按照一级学科办学和教学改革的要求,对计算机科学系教师进行高起点、高标准、正规化研究生学位课程和本科重点课程培训的一个师资培训模式,有可能对未来计算机科学教育产生深远的影响。试想,如果高等学校的教师和培养的研究生普遍具有高级研讨班所开设的3门~4门课程的基础,不仅科学办学面临的主要困难迎刃而解,各大学科研学术队伍的素质也将得到显著提高。

一些学者对高起点研究生学位课程的必要性提出疑问:是否这些课程都要学习?我们认为,应该看到在高等教育界从来就存在着两种不同的教育观,一种是专才教育观,一种是通才教育观。持这两种教育观的人尽管都主张基础知识的重要性,但在对学以致用原则的理解和解释方面存在差异。一般地说,专才教育观主张在一定的基础上,通过深入钻研某一方向的学问,逐步扩展和加深自己的知识,缺什么基础补什么知识,学以致用,逐步成长为一个学科的专家。通才教育观则不同,它不主张在具备一定的基础后,就匆忙沿着某一方向钻研学问,单线独进,而是主张在一级学科的范围,通过尽可能系统地掌握从事本学科各个重要方向的研究所需要的基础知识,能够站在学科的各个至高点上,沿着学科的一个方向,以单线独进、多线并进或整体推进的观点,逐步扩展和加深一级学科的知识,融会贯通,学以致用,逐步成长为一个学科的专家。两种教育观都有其代表人物。迄今为止,高等教育中研究生教育主要以培养专才为主,专才教育观是主流。但是,两种不同的教育观各有其特点。一般地说,当一个学科的发展处于早期时,专才教育比较容易跟上学科的发展步伐,比较容易出成果,也比较容易迅速地达到较深的学术层面。而当一个学科的发展比较成熟,发展速度比较平稳时,通才教育的优势就比较明显。因为,通才教育培养的人才可以在一级学科的范围比较容易地向任何一个方向转向。特别是在胜任高难度重大创新人才的培养方面,在出综合性的大成果方面,在创立一套科学理论和开辟一个研究方向方面,通才教育的多种优势往往是专才教育所不具备的。当然,两种教育观谁优谁劣迄今并无定论,根据两种教育观的特点和现实情况,选择哪一种教育观实际仅反映了师生的一种选择策略。不过,实践告诉我们,尽管通才教育观的操作实现比较困难,但作为师资补充的来源,通才教育培养的人才更容易适应大学教学与科研

的双重要求,理应更多地受到研究大学的青睐。在科学技术日益深化、高度分化又高度综合的今天,放眼未来,在高精尖学科中,通才教育观无疑有着更为宽广而美好的发展前景。

伴随着学科教学改革理论研究与实践探索的推进,社会热切地期待着一套与教学改革方案相配套的高质量系列教材问世。总结过去教材建设成功和失败的经验和教训,使我们清楚地认识到:教材建设必须建立在科学研究基础之上,按照科学的运作程序,动员在第一线从事科学研究、功底深厚、学有所长的优秀教师参与到教材的创作中来,才有可能推出高质量并符合学科发展要求的系列教材。我们的主张是:“让大学中的科学家来创作教材。”

冬去春来,年复一年。当我们终于从跟踪、学习、盲从西方大学教程的发展模式中走出时,感受到了一种从未有过的释然与激动,一种走自己的发展道路,独立自主的自豪与喜悦。这条道路虽然艰难,但前景光明。连续5年在贵州大学举办的全国计算机科学与技术高级研讨班的成功实践更进一步地坚定了对内涵发展模式与科学办学之路的认识与追求。

今天,经过编审委员会、作者和出版社的共同努力,《现代计算机科学与技术教材系列》终于开始陆续出版发行。在新年即将到来的时候,我们怀着喜悦的心情,向祖国和人民,向计算机科学界献上一份完全由华夏学者按照他们对高等教育和计算机科学的理解与认识,倾力创作的新年贺礼,一份建立在科学研究基础之上的教育与教学改革最新成果。

新的世纪已经到来,愿《现代计算机科学与技术教材系列》的出版能够为新一代的莘莘学子攀登现代科学技术的高峰成就未来。

《现代计算机科学与技术教材系列》编审委员会
2006年10月

前 言

当前,众多企业正面临越来越多的管理数据,如何从海量的数据中挖掘、提炼出企业的有效决策支持信息,已经成为各企业发展及培训、招揽人才的主攻方向,而高校人才培养的计划中也高度注意了这一点。本书针对这一需求,以数据挖掘主要方法为主线组织教科书内容,结合国际跨行业数据挖掘标准流程(GRISP-DM),以企业管理面临的现实问题作为应用案例,由浅入深介绍数据挖掘方法在这些问题的数据理解、数据处理、数据分析、数学建模、模型结果评估等过程,引入应用广泛的数据挖掘Clementine软件辅助问题或案例的解决,使读者不仅可以集中地学习数据挖掘的主要方法理论,而且可以学习在实际问题中如何应用数据挖掘方法的数学建模和了解解决问题的过程,更可以学习如何应用现有软件辅助解决问题。本书把理论、案例、建模、和软件辅助结合一体统一叙述,既考虑高校学生的学习需要,分本科生与研究生学习层次,也考虑企业管理者的应用与实践需要。书中融进作者近年来在多个项目中应用数据挖掘方法解决实际问题的教学与科研成果,突出数据挖掘方法应用与数学建模过程和软件辅助的有机结合。

本书的初稿是教案,曾在应用数学专业(应用软件方向)、信息管理与信息系统专业、信息与计算科学专业的《管理与决策支持系统》的本科生教学中使用,深受学生欢迎。从人才需求反馈信息了解到,企业非常重视能熟练掌握数据挖掘方法及其软件应用的人员。因为市场调查与预测、营销(含供应链管理和客户管理)、生产、采购、库存以及财务、质量等管理部门都有这一强烈的应用需求,传统的已知数据模型关系进行管理的计划与控制已不能完全满足企业的战略需求,需要从大量的业务数据中通过专门的算法进行挖掘、提炼,使潜在的数据关系模型明确化,为企业的决策支持使用。因此,掌握数据挖掘理论、方法与技术的学生社会就业中常常得到企业的青睐。

本书初稿也曾作为中国移动通信广东分公司的管理层培训材料。对于数据挖掘方法及软件在客户分类管理、客户流失分析、营销策略关联分析、神经网络辅助市场分析与预测、营销布点优化、促销方案制定等实际问题,引起了移动通信管理层的极大兴趣,使他们从传统的统计分析提高到可以通过大量销售数据、客户数据发现有效决策支持信息的水平。

本书的特点是配置适合高年级本科生的基本数据挖掘理论和来自实际管理的问题,注重论述数据挖掘方法在解决实际问题中的数学建模过程,注重应用数据挖掘软件 Clementine 辅助建模过程的实现,注重模型结果分析与决策信息的提取,在一定程度上满足数据挖掘理论与技术的教学、实践、应用、提高的需要。

本书以数据挖掘标准流程的主要方法和参照“商业理解、数据理解、数据处理、模型建立、结果评估、应用部署”的六个步骤,并按“问题理解、数据处理、模型建立、模型检验”一环扣一环地论述数据挖掘和数学建模的关键过程。全书共分为 8 章,第 1 章主要论述当前信息化发展趋势及决策支持面临的问题,论述数据挖掘与数学建模的关系,概述基于数据挖掘的数学建模主要步骤;第 2 章主要论述当前应用较多的统计分析方法,包括回归分析、逻辑回归分析、主成分分析与因子分析等;第 3 章论述数据处理应用较广的聚类分析方法;第 4 章论述神经网络及模糊神经网络辅助数学模型建立及其应用;第 5 章论述决策树、模糊决策树及其在分类问题及知识规则提取中的应用;第 6 章论述关联分析、时序关联、关联分类、模糊关联分析及其在不确定性问题中的应用;第 7 章论述遗传算法及其在解决优化问题中的应用;第 8 章论述贝叶斯网络与模糊贝叶斯网络的建立及其在不确定性知识推理与诊断中的应用。每章除了在知识点论述配有相应例题外,还配有来自实际问题的综合案例分析,包括数学建模过程与数据挖掘软件 Clementine 辅助实现过程。

华南理工大学研究生李晶、黄志炜、焦厚昌、谢静、付辉、陈潇潇、唐仲华、古慧敏、郑佳碧、陈波、陈望宇、刘晋中、晏文娟等为本书的资料查阅、内容整理与案例数据处理做了大量工作,在此表示感谢。

由于水平有限,书中难免有疏漏和错误,恳请读者指正。

编著者
2009 年 10 月
于华南理工大学

目 录

第 1 章 数据挖掘与数学建模关系概述	1
1.1 当前信息化发展的趋势与面对问题	1
1.2 数据挖掘发展及其应用	2
1.3 基于数据挖掘的数学建模	3
1.4 数据挖掘软件 Clementine 的基本操作概述	14
1.4.1 Clementine 数据挖掘的基本思想	14
1.4.2 Clementine 的基本操作方法	15
参考文献	22
第 2 章 统计分析	23
2.1 问题概述	23
2.2 回归分析及其应用	23
2.2.1 回归分析概述	23
2.2.2 一元线性回归及其模型建立	24
2.2.3 多元线性回归及其建模过程	27
2.2.4 Clementine 辅助多元回归分析	30
2.3 二项逻辑回归	36
2.3.1 二项逻辑回归概述	36
2.3.2 二项逻辑回归模型	36
2.3.3 二项逻辑回归方程中回归系数的含义	37
2.3.4 二项逻辑回归方程的检验	38
2.3.5 Clementine 辅助 Logistic 回归模型	40
2.4 主成分分析	47
2.4.1 主成分分析概述	47
2.4.2 主成分分析的数学模型	47
2.4.3 主成分计算步骤及应用	49
2.5 因子分析	52

2.5.1	因子分析概述	52
2.5.2	因子分析的数学模型	53
2.5.3	因子载荷阵的估计方法	54
2.5.4	因子旋转	55
2.5.5	因子得分	56
2.5.6	因子分析与主成分分析的联系与区别	60
2.6	管理胜任力的案例分析和数学建模	61
2.6.1	问题提出	61
2.6.2	数据处理	61
2.6.3	模型建立	62
2.6.4	模型检验	68
2.6.5	模型应用	69
	参考文献	69
第3章	聚类分析	70
3.1	问题概述	70
3.2	聚类分析概述	70
3.3	基于距离的聚类相似度	73
3.4	系统聚类法	74
3.5	C - 均值(C - Means)聚类算法	80
3.6	Clementine 辅助 K - Means 聚类	80
3.7	模糊聚类	86
3.7.1	模糊 C - Means(FCM)算法	86
3.7.2	WFCM 算法	88
3.8	聚类有效性	88
3.8.1	基于可能性分布的聚类有效性函数	89
3.8.2	基于模糊相关度的聚类有效性函数	91
3.9	医疗建设评价的案例分析与数学建模	93
	参考文献	98
第4章	神经网络及其应用	100
4.1	问题概述	100
4.2	神经网络概述	100
4.3	神经网络的基本模型	101

4.3.1	神经网络的理论依据	101
4.3.2	神经网络的组成	102
4.3.3	感知机模型	104
4.4	误差逆传播神经网络模型	107
4.4.1	BP 神经网络的基本原理	107
4.4.2	BP 神经网络的学习算法	110
4.4.3	Clementine 辅助 XOR 问题的模型建立	111
4.5	RBF 神经网络	113
4.5.1	RBF 神经网络的基本原理	114
4.5.2	RBF 神经网络的学习算法	115
4.5.3	新型轮胎性能的 RBF 神经网络预测	116
4.6	自组织特征映射(SOM)神经网络	118
4.6.1	SOM 模型的基本原理	118
4.6.2	SOM 模型的学习算法	119
4.6.3	Clementine 辅助 Kohonen 模型建立	121
4.7	神经网络的案例分析与数学建模	123
4.7.1	城市医疗能力评价的 BP 神经网络	123
4.7.2	地下燃气管网安全风险的综合评价	130
4.8	模糊神经网络	137
4.8.1	模糊基本概念	137
4.8.2	模糊神经网络概述	139
4.8.3	模糊神经网络学习算法	140
4.8.4	食品安全管理主任评价模糊神经网络模型	141
	参考文献	149
第 5 章	决策树及其应用	150
5.1	问题概述	150
5.2	决策树概述	150
5.2.1	决策树基本算法	151
5.2.2	CLS 算法	151
5.2.3	信息熵	152
5.3	ID3 算法	153
5.3.1	基本思想	153
5.3.2	ID3 算法应用实例	154

5.4	C4.5 算法	157
5.4.1	基本思想	157
5.4.2	基于信息增益率建模的决策树	158
5.5	CART 算法	159
5.5.1	基本思想	159
5.5.2	基于 CART 算法建模的决策树	160
5.6	Clementine 辅助决策树建立	162
5.7	决策树的评价标准	170
5.8	决策树的剪枝	171
5.9	决策树的优化	174
5.10	燃气管网安全预测案例分析与数学建模	174
5.11	模糊决策树	181
5.11.1	模糊决策树基本原理	181
5.11.2	模糊决策树 ID3 算法	182
5.11.3	绩效评估的模糊决策树建立	183
	参考文献	186

第 6 章	关联分析	188
6.1	问题概述	188
6.2	关联分析概述	188
6.3	Apriori 关联规则算法	190
6.4	Clementine 辅助 Apriori 算法实现关联分析	194
6.5	基于 Apriori 算法的改进算法	195
6.6	基于分类搜索的关联规则算法	197
6.7	基于频繁树的关联规则算法	200
6.8	关联规则的推广	203
6.8.1	层次关联规则算法	203
6.8.2	三种层次关联规则的算法	204
6.9	时序关联规则算法	205
6.10	Clememtin 辅助时序关联分析	210
6.11	多值属性关联规则算法	213
6.12	增量关联规则算法	216
6.12.1	方法概述	216
6.12.2	算法描述	216

6.13	基于关联规则的分类算法	217
6.13.1	关联规则分类算法概述	217
6.13.2	关联分类	218
6.14	关联分类算法的改进	221
6.15	模糊关联分类算法与应用	224
6.15.1	属性的模糊划分	224
6.15.2	基于类别的模糊关联分类算法	224
6.16	关联规则案例分析与数学建模	226
6.16.1	基于移动通信客户群的关联分类	226
6.16.2	移动通信客户成长模糊关联分析	230
6.16.3	地下燃气管网影响因素与安全级别的关联分析	232
6.16.4	电子商务发展影响因素的时序关联分析	235
	参考文献	239
第7章	遗传算法	241
7.1	问题概述	241
7.2	遗传算法概述	241
7.2.1	基本概念	241
7.2.2	标准遗传算法	242
7.2.3	选择算子	243
7.2.4	交叉算子	245
7.2.5	变异算子	246
7.2.6	用标准遗传算法求解函数最大值	247
7.2.7	遗传算法优化神经网络参数	249
7.3	模式定理	251
7.4	改进的遗传算法	254
7.4.1	分层遗传算法	255
7.4.2	CHC 算法	256
7.4.3	messyGA 算法	256
7.4.4	自适应遗传算法	258
7.4.5	基于小生境技术的遗传算法	259
7.4.6	混合遗传算法	261
7.5	遗传算法案例分析与数学建模	264
7.5.1	地下燃气管网安全风险的优化控制	264

7.5.2	网络广告收益的优化控制	271
7.5.3	政府资源计划与控制体系的优化设计	275
7.5.4	ERP 系统物料需求计划的遗传优化	279
7.5.5	基于遗传算法的决策树优化	285
参考文献		289
第 8 章 贝叶斯网络与知识推理		292
8.1	问题概述	292
8.2	贝叶斯网络原理概述	292
8.2.1	贝叶斯网络定义	292
8.2.2	贝叶斯网络的知识推理模式	293
8.2.3	贝叶斯网络建立的主要步骤	293
8.2.4	贝叶斯网络的结构学习	294
8.2.5	贝叶斯网络的参数学习	295
8.3	主要贝叶斯网络模型	297
8.3.1	朴素贝叶斯网络	297
8.3.2	TAN 贝叶斯网络	301
8.3.3	无约束贝叶斯网络	305
8.4	管理胜任力的贝叶斯网络知识推理模型建立	307
8.5	Clementine 辅助贝叶斯网络的建立	315
8.6	地下燃气管网安全风险的贝叶斯网络知识推理与诊断	323
8.7	模糊贝叶斯网络	336
8.7.1	模糊概率	336
8.7.2	模糊全概率公式和贝叶斯公式	337
8.7.3	模糊贝叶斯网络模型建立	337
8.8	电子商务发展水平的模糊贝叶斯网络知识推理模型	341
8.9	数据挖掘软件 WEKA 辅助无约束贝叶斯网络的建立	345
8.9.1	WEKA 简介	345
8.9.2	应用 WEKA 建立贝叶斯网络模型	346
参考文献		352