灰色粗糙集模型及其应用

吴顺祥 /著

科学出版社
www.sciencep.com

灰色粗糙集模型及其应用

吴顺祥　著

科学出版社

北京

内 容 简 介

本书介绍了粗糙集与灰色系统的理论、方法与应用，并针对粗糙集理论与灰色系统理论的数据融合理论与技术进行了研究，较系统地介绍了基于区间灰色集的粗糙集的各种模型、方法及应用，这是对传统不完备信息系统的有效拓展，为粗糙集理论与灰色系统理论的研究提供了一个全新的视角。全书内容分为十章，包括粗糙集理论的基本概念与基本理论，灰色系统理论的基本概念与基本理论，区间灰集的表征及其运算法则，灰色粗糙集模型及其性质，灰色信息系统的粗糙集拓展模型，基于(α,β)-灰相似关系的粗糙集模型，基于构造性方法的灰色粗糙集模型，基于灰色信息系统的优势关系及其属性约简方法，一种基于连续属性值的灰色决策表的属性约简方法，以及一种基于灰色区间的 BP 神经网络算法等。

本书可作为高等院校信息科学、应用数学及管理科学等相关专业高年级本科生及研究生教材，也可作为相关专业教师、科技工作者、工程技术人员和企业管理人员的参考书。

图书在版编目(CIP)数据

灰色粗糙集模型及其应用/吴顺祥著. —北京：科学出版社，2009
ISBN 978-7-03-023988-4

Ⅰ. 灰… Ⅱ. 吴… Ⅲ. 粗糙集-应用-灰色模型 Ⅳ. N94

中国版本图书馆 CIP 数据核字(2009)第 017128 号

责任编辑：姚庆爽 / 责任校对：陈丽珠
责任印制：赵 博 / 封面设计：王 浩

科 学 出 版 社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

铭浩彩色印装有限公司印刷

科学出版社发行 各地新华书店经销

*

2009 年 2 月第 一 版 开本：B5(720×1000)
2009 年 2 月第一次印刷 印张：10 1/2
印数：1—2 500 字数：208 000

定价：35.00 元

(如有印装质量问题，我社负责调换〈路通〉)

前　言

粗糙集理论是20世纪80年代初由Pawlak提出的一种处理不确定性问题的一种有效的软计算方法，其最大特点是在处理不确定性问题时不需要提供问题所需处理的数据集合之外的任何先验信息，适合于发现数据中潜在的、隐含的知识，为信息系统的约简、决策表规则的提取提供了一整套方法，随着其理论的拓展和不断完善，在各方面的应用已取得了很大进展。

灰色系统理论同样是在20世纪80年代初由邓聚龙教授提出的一种解决不确定问题一种有效的数学工具，由于其应用的广泛性和有效性而备受关注，其基础理论研究和应用方面都取得了很大的进展。灰色系统理论主要针对信息不完备、不确定，可用数据少的不确定性问题，研究部分信息已知，部分信息未知或非确知的"小样本、贫信息"不确定性系统，主要通过对部分已知信息的生成、开发，提取有价值的信息，实现对系统运行行为、演化规律的正确描述和有效监控。灰色系统理论已诸多领域成功地解决了大量实际问题。

粗糙集理论和灰色系统理论都是处理不完全、不精确及不确定性信息的有效工具，通过对这两种理论进行有益的融合来研究处理不确定性问题的更有效和更一般化的方法，无疑是一项有重要现实意义的工作。

本书深入分析了粗糙集和灰色系统理论的发展和研究现状，详细介绍了粗糙集和灰色系统理论的基本知识，在此基础上对灰色系统理论和粗糙集理论相结合进行信息处理的方法进行了研究。提出了区间灰色集的基本概念，构建了一种新的区间灰色集的表征与运算法则体系，提出了基于区间灰色集的粗糙集的各种模型，在此基础上，研究并提出了基于区间灰色集的灰色容差关系、非对称的灰色相似关系和灰色拟序关系及其属性约简方法，这是对传统不完备信息系统的有效拓展；本书针对灰色信息系统和灰色决策表，提出了基于(α,β)-灰相似关系的粗糙集模型，利用认知程度模式来描述正域、边界域、负域、相对约简等概念，对灰色决策表中的不一致性问题从认知程度模式的角度提出了一种解决方法，其结果表明该模型能增加对不确定性信息的描述精度，增强灰色信息覆盖的能力，能更加有效、准确地发现和提取灰色决策表中所蕴涵的规则或规律；本书最后采用构造性方法来研究灰色粗糙集，定义了四对灰色粗糙集近似算子，对其相关基础理论进行研究。在应用方面提出了灰色区间关联聚类与粗糙近似组合决策方法，有效地解决了单一使用灰色关联聚类分析方法或粗糙集方法不能精确地刻画不确定性决策问题的本质，决策结果不够准确或不能更好地贴近实际的问题，具有更好的普适性和

有效性。本书讨论了灰色信息系统中的基于优势关系的属性约简方法和辨识矩阵的构造方法及基于优势度排序的方法。提出一种基于连续属性值的离散化处理方法和灰色决策表的属性约简方法。本书结合 FBBP 算法和灰色区间，提出了一种缩小输入灰度的 GBP 算法，该算法将传统的权值调整方法与输入矢量调整的方法结合起来，并对输入矢量进行灰度调整，在保持原有分类的不变的情况下，提高了训练速度和泛化能力，实现了更加有效的学习。

本书的研究工作是对粗糙集和灰色系统理论的有效融合及应用这个崭新的研究领域的有益探索，为深入研究粗糙集和灰色系统理论提供了更广阔的思路。

本书的出版得到了国家“十一五”科技支撑计划项目(2007BAK34B04)和厦门大学 985 二期信息创新平台项目的资助，在此表示衷心的感谢！

由于作者知识水平的局限性，书中不妥之处再所难免，恳请各位专家和读者批评指正。

作　者

2008 年 8 月

此为试读，需要完整PDF请访问：www.ertongbook.com

本书符号表

符号	
U	有限非空论域
R	二元关系
l	划分
2^U	集合 U 的幂集
ϕ	空集
$\|X\|$	集合 X 的基数
U/R	U 关于等价关系 R 的划分
Ω	发生背景
$\mathrm{ind}(B)$	信息系统中由集合 B 所确定的不可区分关系
$[x]_B$	对象 x 所在的 $\mathrm{ind}(B)$ 等价类
$\bigvee_i L_i$	析取范式
$\underline{R}(X)$	集合 X 的 R 下近似集
$\overline{R}(X)$	集合 X 的 R 上近似集
$\mathrm{bn}_R(X)$	集合 X 的 R 边界域
$\mathrm{neg}_R(X)$	集合 X 的 R 负域
$\mathrm{pos}_R(X)$	集合 X 的 R 正域
$\alpha_R(X)$	集合 X 的近似精度
$\rho_R(X)$	集合 X 的 R 粗糙度
$\mathrm{core}(P)$	集合 P 的核
$\mathrm{red}(P)$	集合 P 的所有约简
$\mathrm{pos}_P(Q)$	集合 Q 的 P 正域
$\mathrm{core}_Q(P)$	集合 P 的 Q 核
$\sigma_{(C,D)}(C')$	子集 $C'\subseteq C$ 关于 D 的重要性
$\kappa_{(C)}(D)$	知识 D 依赖于知识 C 的程度
$C_D(i,j)$	区分矩阵
Δ	区分函数
$\otimes$	灰数
$\widetilde{\otimes}$	灰数$\otimes$的白化数

$\otimes\Big\vert_{P}^{Q}$	以集合 P 为下界，集合 Q 为上界的区间灰集
$\mu(\otimes)$	灰数⊗的测度
$g^{\circ}(\otimes)$	灰数⊗的灰度
$\upsilon(\otimes)$	灰数⊗的不确定度
f	信息函数
$\otimes(a)$	表示以 a 为白化值的灰数
$\gamma(x_i(k),x_j(k))$	两个序列 X_i 与 X_j 之间的灰关联系数
$\underline{R}(\otimes)$	灰集⊗的下近似
$\overline{R}(\otimes)$	灰集⊗的上近似
$\underset{\sim}{R}(\otimes)$	灰色下近似
$\tilde{R}(\otimes)$	灰色上近似
$R^{\sim}(\otimes)$	上近似粗糙灰集
$R_{\sim}(\otimes)$	下近似粗糙灰集
$\dot{R}(\otimes)$	关系 R 下由可定义集导出的灰色粗糙集
$\alpha(\otimes,R)$	近似空间(U,R)的灰色精度
$\rho(\otimes,R)$	近似空间(U,R)的灰色粗糙度
$\upsilon(\otimes,R)$	近似空间(U,R)的不确定度
T_B	集合 B 上的灰色相容关系
S_B	集合 B 上的非对称灰色相似关系
P_B	集合 B 上的灰色拟序关系
$\underset{\sim}{\approx}$	灰色粗下相等
$\overset{\sim}{\approx}$	灰色粗上相等
$\approx$	灰色粗相等
$\underset{\sim}{\subset}$	灰色粗下包含
$\overset{\sim}{\subset}$	灰色粗上包含
$\subset$	灰色粗包含
$\underset{\sim}{\approx}'$	粗糙灰下相等
$\overset{\sim}{\approx}'$	粗糙灰上相等
$\approx'$	粗糙灰相等
$\underset{\sim}{\subset}'$	粗糙灰下包含
$\overset{\sim}{\subset}'$	粗糙灰上包含
$\subset'$	粗糙灰包含
$\Leftrightarrow$	等价；当且仅当
$\Rightarrow$	蕴涵

目　录

此为试读,需要完整PDF请访问: www.ertongbook.com

第一章　粗糙集理论的基本概念与基本理论

1.1　粗糙集理论的研究现状

20 世纪 70 年代初，波兰学者 Pawlak 和波兰科学院、华沙大学的逻辑学家们组成了研究小组，开始了对信息系统逻辑特性的长期基础性研究。针对从实验中得到的以数据形式表述的不精确、不确定、不完整的信息和知识进行了分类分析，这一研究成为粗糙集(rough sets，简称“粗集”)理论产生的基础。1982 年，Pawlak 针对 Frege 的边界线区域思想提出了该处理含糊和不确定性问题的新型数学工具——粗糙集，其最大特点是无须提供问题所需处理数据以外的任何先验信息，即“让数据自己说话”　经典论文 *Rough Set* 的发表宣告了粗糙集理论的诞生。直至 20 世纪 80 年代，许多专家学者对粗糙集理论进行了深入的研究，不过主要集中在对粗糙集理论的数学性质及逻辑系统的分析研究。1991 年，Pawlak 的专著 *Rough Sets* 的出版是粗糙集理论研究的一个里程碑。1992 年 Slowinski 主编的关于粗糙集应用及其与相关方法比较研究的论文集的出版，对这一时期的工作成果做了很好的总结，也促进了粗糙集理论在应用领域的推广。1995 年，ACM Communication 将其列为新浮现的计算机科学的研究课题。

1992 年在波兰 Kiekrz 召开了第 1 届国际粗糙集讨论会，这次会议着重讨论了集合近似定义的基本思想及其应用，其中粗糙集环境下机器学习的基础研究是这次会议的四个专题之一。从此每年召开一次以粗糙集理论为主题的国际研讨会。1993 在加拿大 Banff 召开了第 2 届国际粗糙集与知识发现(RSD'93)研讨会，这次会议极大地推动了国际上对粗糙集理论与应用的研究，其主题是粗糙集、模糊集与知识发现。特别值得一提的是在 1995 年召开的第 4 届模糊理论与技术国际研讨会(Fuzzy Theory & Technology'95)上，针对粗糙集与模糊集合的基本观点与相互关系展开了激烈的讨论，较大地促进了粗糙集的研究。1996 年底在日本东京召开了第 5 届国际粗糙集研讨会，这是第一次在亚洲地区召开的范围广泛的粗糙集研讨会。1998 年 6 月在波兰华沙召开了第 1 届粗糙集和计算的当前趋势学术会议。1999 年 11 月在日本召开了第 7 届粗糙集、模糊集、数据挖掘和粒度-软计算的国际学术研讨会(RSFDGrC'99)，阐述了当前粗糙集、模糊集的研究现状和发展趋势。2000 年 10 月在加拿大召开了第 2 届粗糙集和计算科学的当前趋势学术会议。2001 年在日本召开了粗糙集理论和粒度计算国际研讨会。2002 年在美国召

开了第 3 届粗糙集和计算科学的当前趋势学术会议。RSFDGrC2003 和 CRSSC2003 国际学术会议是在中国重庆召开的。2005 年在加拿大 Regina 大学召开了第 10 届粗糙集、模糊集和粒计算国际学术研讨会(RSFDGrC'2005)。2006 年在日本召开第五届粗糙集和计算科学的当前趋势国际学术研讨会(RSCTC'2006)。目前,在关于人工智能、模糊理论、信息管理与知识发现等国际学术会议上经常可以看到许多涉及粗糙集的论文。

粗糙集理论的研究在中国也得到了积极推广。2001 年 5 月在重庆召开了第一届中国 Rough 集与软计算学术研讨会,邀请了创始人 Pawlak 教授做大会报告。2002 年 10 月在苏州召开了第二届中国 Rough 集与软计算学术研讨会。2003 年 10 月在重庆召开了第三届中国 Rough 集与软计算学术研讨会,并同时举办第九届粗糙集、模糊集、数据挖掘和粒度-软计算的国际会议。2003 年成立了中国人工智能学会粗糙集与软计算专业委员会。2004 年 10 月在舟山召开了第四届中国 Rough 集与软计算学术研讨会。2005 年 8 月在鞍山召开了第五届中国 Rough 集与软计算学术研讨会。2006 年 7 月在中国重庆召开了首届粗糙集和知识发现国际学术研讨会(RSKT'2006),并于 2006 年 10 月在浙江金华召开了第六届中国 Rough 集与软计算学术研讨会。第七届中国 Rough 集与软计算学术会议(CRSSC2007)于 2007 年 8 月在山西太原召开。我国近年来在此领域的研究发展速度非常快,Rough 集的研究队伍也更加壮大,每年的 CRSSC 系列研讨会在规模和质量上均呈良好的增长趋势,研究成果在深度和广度上有了更大的发展。其理论模型得到不断完善和发展,并渗透到很多学科,成为研究数据挖掘、知识约简和粒计算的理论基础。Rough 集理论自身也已成为完整、独立的科学领域。此外,Rough 集理论与其他一些软计算理论,诸如 Fuzzy 集、粒计算、神经网络、遗传算法等均已经成为当前国内外计算机及相关专业的研究热点。

目前,国内外对粗糙集理论的研究主要集中在以下方面。

1. 粗糙集的数学性质

对粗糙集数学性质的研究,主要包括研究集合和分类近似的性质、决策表性质、代数结构、粗糙集代数、粗糙集逻辑、粗糙集拓扑结构及其收敛性问题,它们是粗糙集理论形成和发展的基础。研究粗糙集数学性质的方法主要有构造性方法和公理化方法。构造性方法是从一个二元关系出发定义一对上、下近似算子,通过不同的二元关系定义不同的近似算子,从而构造不同类型的粗糙集代数,如序列粗糙集代数、反射粗糙集代数、对称粗糙集代数、传递粗糙集代数、Pawlak 粗糙集代数等。Yao 系统地研究了上述各种特殊类型的粗糙集代数及其相应近似算子所具备的特性,基于普通的邻域关系,引入了邻域系统的概念,并系统地研究了邻域系统与粗糙近似之间的关系,证明了在邻域系统表示的近似算子意义下 Pawlak 近似

算子和模态逻辑的可能性和必然性算子是一致的，这些工作为建立近似模型提供了强有力的工具；而公理化方法是通过定义满足不同的特定的公理系统的一对对偶近似算子来刻画不同类型的粗糙集代数的，这种方法又称粗糙集代数方法；Liu等给出了一个粗糙集公理组，并证明了公理组的可靠性；祝峰等简化了该公理组，也证明了简化公理组的可靠性；在此基础上，孙辉等进一步研究了粗糙集公理组的极小化问题，得到了两个简化的粗糙集公理组，并讨论了它们的可靠性和极小性；Pawlak定义了粗糙逻辑和决策逻辑；Skowron研究了粗糙概念逻辑和近似逻辑，强调了这种逻辑的完备性；Charabory提出了带粗糙量词的粗糙逻辑，并建立了一套近似推理的逻辑工具；史开泉等提出了奇异粗糙集（简称S-粗集）的概念、数学结构及单向S-粗集和双向S-粗集的性质；吴志伟等提出了粗糙模糊集的构造与公理化方法；Nakamura定义了一种粗糙层次模态性，把粗糙逻辑、模糊逻辑和模态逻辑融为一体。

2. 数据的预处理

粗糙集理论只能处理离散数据，这一局限大大限制了粗糙集理论的应用范围，因此连续属性的离散化成为粗糙集理论的主要问题之一，也是研究粗糙集理论实用性的瓶颈之一。因此，实际应用中，必须先对连续属性值进行离散化处理。粗糙集理论中的离散方法基于两类。一类基本上很少或不考虑粗糙集理论的特殊性，只是把机器学习等其他学科中的离散化问题借用到粗糙集理论上来，离散化效果并不突出；另一类注意到了粗糙集理论对决策表的特殊要求，采取结合方法来解决离散化问题。目前在粗糙集理论中，连续属性的离散化方法很多，如非监督离散方法中的等宽度离散化、等频率离散化，监督离散方法中的单规则离散器、统计检验方法、信息熵方法、自适应量化法及预测值最大算法及非线性组合构造超曲面的方法等。连续属性的离散化使得粗糙集理论对离散和连续的属性都能处理，扩大了粗糙集理论的应用范围。

数据预处理的另一个重要内容是对不完备信息表的完备化。在很多情况下，得到的待处理的信息表并不是一个完备的信息表，表中的某些属性值是被遗漏的，且无从知道其原始值，这也是信息系统不确定性的一种主要原因。对于这种情况，目前主要通过以下途径来对信息表中的遗漏数据进行补齐。一种途径是简单地将存在空缺（遗漏）属性值的实例记录删除，从而得到一个完备的信息表。虽然这种方法不是严格意义上的数据补齐，然而在信息表数据巨大的并且有遗漏属性值的实例记录的数量远远小于信息表所包含的记录数时，这种方法在删除不完整记录之后并不太影响信息表中信息的完整性，是一种可取的处理方法。但是，当信息表中的信息较少、存在遗漏信息的实例相对较多时，这种方法就会严重影响信息表中的信息量，这时就不能采用这种方法了。第二种途径是将空缺属性值作为一种特

殊的属性值来处理，它不同于其他任何属性值，这样就能实现不完备信息表的完备化。第三种途径是采用统计学原理，根据信息表中其余实例在该属性上的取值的分布情况来对一个空缺属性值进行估计补充，这样不会影响信息表中包含的信息量。第四种途径是根据粗糙集理论中数据不可分辨关系来对不完备的数据进行补齐处理。传统的粗糙集理论和方法已经成功地用于处理不精确、不一致、不确定的数据或知识，但它存在一个假定的前提，即所有可以获得的个体对象由这个属性集合给出完全的描述。换句话说，用 $U=\{x_1,x_2,\cdots,x_n\}$ 表示个体对象的有限集合，$A=\{a_1,a_2,\cdots,a_m\}$ 表示属性集合，则对于任意 $a\in A,u\in U$，属性值 $a(x)$ 总是存在的，即 $a(x)\neq\phi$。这个假设虽然是合理的，但与很多现实情况有差异。在这些情况下，由于不可能得到一部分属性值（例如，集合 U 是关于病人的集合，属性是一些临床检验，则并非所有的检验结果在给定时间内都可以得到），或者由于存储介质的故障、传输媒体的故障、一些人为因素等等，导致关于对象集合 U 的描述是不完全的。这样，就导致了不完全信息系统的出现。

3. 核与约简的求取

核和约简是粗糙集理论中的两个核心概念，对数据的约简起着重要作用。约简往往不止一个，求出所有的属性约简是 NP(non-deterministic polynomial)难题，常采用启发方法找出一个最优或次优约简，其中基于“属性重要性”思想的启发式算法得到了广泛的研究。核是最重要的属性集合，寻找核的意义在于：从核开始求取属性约简，会大大减少求属性约简的计算量。最初提出该算法的是 Hu，使用核作为计算的初始约简，引入“属性的重要性”这样一个度量作为启发信息，按照属性的重要程度的大小逐个将属性加入约简集，直到该集合是一个约简为止；此外，属性约简也成为数据挖掘的一项重要工作，属性约简的方法和技术也不断成熟，取得了许多可喜的研究成果。

在粗糙集理论的各种应用中，属性约简算法具有重要意义，是知识发现的重要课题，因而对属性约简算法的研究一直是粗糙集理论研究中的核心作用问题之一。根据粗糙集中的定义寻找属性的最小约简，会导致组合爆炸问题，已被证明是一个 NP-hard 问题，因此需要研究更为有效的约简算法，而运用启发信息来简化计算是最直接的思想。目前最常采用的是基于启发方法找出一个最优或次优约简，其中基于“属性重要性”思想的启发式算法得到了广泛的研究。最初提出该算法的是 Hu，使用核作为计算的初始约简，引入“属性的重要性”这样一个度量作为启发信息，按照属性的重要程度的大小逐个将属性加入约简集，直到该集合是一个约简为止。该算法可以很简单直观地计算一个最好的或用户指定的最小约简。Jakub 提出了基于遗传算法去寻找系统的最小约简。Kryszkiewiez 和 Rybinski 研究了在复合信息系统中寻求约简的问题，通过寻求子系统的约简最终求出复合系统的约

简。其主要思想是将布尔函数的化简问题转化成集合空间中的边界搜索问题，从而在已知子系统的约简的情况下，简化复合系统的搜索空间。Starzyk 等提出强等价的概念，进而发展为扩展法则，用于快速简化区分函数。Bazan 等提出动态约简方法，该方法能够有效的提高约简的抗噪声能力。此外，属性约简也成为数据挖掘的一项重要工作，属性约简的方法和技术也不断成熟，取得了许多可喜的研究成果。

4. 粗糙集模型的拓广

由于经典的粗糙集理论是基于完备信息系统的假设，即每个对象的所有属性值都是确定的。而在现实生活中，由于数据测量的误差、数据理解或获取的限制等原因，会遇到噪声、数据缺失、大数据量、连续属性离散化等具体问题，要保证对象属性的完整性是非常困难的，有时也是不必要的，因此基于经典的粗糙集理论的等价关系将不再成立。为了使粗糙集理论能够适应于不完备信息系统的处理，目前其处理方法主要有两种：第一种是间接处理方法，它的特点是通过一定的方法（一般是基于概率统计）把不完备信息系统转化为完备信息系统，即数据补齐；第二种是直接处理方法，它的特点是对经典粗糙集理论中的相关概念在不完备信息系统下进行适当的扩充，由于粗糙集所基于的论域上的关系必须是经典关系，其等价关系的要求过于严格，从而限制了经典粗糙集的应用，于是学者们对等价关系进行推广，将等价关系放宽为相容关系、相似关系，甚至于一般意义下的二元关系，提出了一些泛化的粗糙集模型，从而得到各种不同的扩展的粗糙集模型。可变精度粗糙集模型通过引入两个集合的相等误分类度，把集合的普通包含关系放宽为多数包含关系。在允许的分类误差下对概念的上近似和下近似重新定义；相容关系模型针对经典粗糙集中等价关系条件太强的缺陷，提出了用相容关系代替传统的等价关系的相容关系模型；概率粗糙集模型通过条件概率来定义概念的上、下近似，弥补了早期的经典粗糙集模型没有考虑与不确定性分类问题有关的概率分布信息的缺陷。此外还有基于覆盖的粗糙集模型、粗糙模糊集和模糊粗糙集模型等，这些模型的提出丰富了粗糙集理论，为粗糙集理论的应用研究拓宽了道路。

在对于信息缺省的不完备性研究方面，Kryszkiewicz 提出了基于容差关系的粗糙集扩展模型；Stefanouski 和 Tsoukeas 在容差关系的基础之上提出了非对称相似关系和基于量化容差关系的粗糙集扩展模型；王国胤提出了限制性的容差关系模型；黄兵和周献中提出了基于联系度的粗糙集模型。而针对属性取值不确定性问题方面，罗党利用灰色聚类决策的机制构建决策表，然后利用粗糙集理论从决策表中挖掘出极小化决策算法；张文修等利用可辨识矩阵给出了一种基于集值信息系统的知识约简方法和将不协调集值决策信息系统转化为广义协调近似空间的方法；杨小平提出了基于属性集值的不完备信息系统的三种基本关系及其属性约

简方法；吴陈等根据几种不同的上、下近似集定义，提出了一种基于集值信息系统的决策方法。这些思想与方法都拓展了传统完备/不完备信息系统的思想，在处理不完备信息系统方面显示出更好的适用性和应用前景。

5. 粗糙集理论的应用研究

近几年来，粗糙集理论在机器学习、知识发现、决策支持与分析、专家系统、智能控制、模式识别等学科领域获得了成功应用并得到了交叉发展。粗糙集理论在许多应用领域也得到了成功的应用，如医学、图像处理、环境改善与保护、市政工程、电力系统、地质分析等等。目前国际上已经开发出了一些基于粗糙集理论的KDD系统。

实践证明粗糙集理论是处理模糊和不确定性知识的很好的工具。粗糙集理论的方法在社会生活的许多领域都有重要作用。粗糙集理论也被证明是完备而有效的。在今后的研究中，粗糙集理论的研究主要将集中在与其他前沿学科的交叉结合和应用方面，如粗糙集与数据挖掘、人工智能等前沿学科领域的交叉结合。随着Internet的迅速扩展，Web页面的增加，利用粗糙集进行Web知识发现将是今后重要的研究课题之一。

1.2 集合论的基本知识

1.2.1 集合论概述

自从19世纪康托(Cantor)创立了集合论以来，集合论就已成为现代数学的基础。集合的表示通常有三种。其一是列举法，也就是集合的元素全部枚举出来，这只有元素数目少的情况下用这种方法。例如，中国的直辖市＝{北京，天津，上海，重庆}。其二是性质法，就是用集合中的元素具有某种共性来描述集合。例如，具有 $x<10$ 的特性的奇数的集合，$A=\{x \mid x<10 \wedge x$ 是奇数$\}$。其三是特征函数法，即集合中的元素与特征值0和1对应起来表示。例如，一个学习小组有6个人，分别用 x_1, x_2, x_3, x_4, x_5 和 x_6 表示，如果 x_i 是男，则记成 $1/x_i$，若 x_i 是女，则记成 $0/x_i$，于是可将这个集合记成 $A=\{0/x_1, 1/x_2, 0/x_3, 1/x_4, 1/x_5, 1/x_6\}$。

一个集合通常用大写字母 A 表示，而集合的元素用小写字母 a 表示。某个元素 a 属于集合 A，或者不属于 A，分别记作 $a \in A$ 或 $a \notin A$。

一个集合的全部元素数目称做该集合的基数，记成 $|A|$ 或 $K(A)$，$\mathrm{card}(A)$。其基数可以是无限的。例如集合{北京，天津，上海，重庆}和$\{x \mid x$ 是奇数$\}$分别为有限集和无限集。对于无限集，通常不能用数字来写出它的基数。

如果集合 B 中的元素全部都能在集合 A 中找到，则集合 B 被称为集合 A 的

此为试读，需要完整PDF请访问：www.ertongbook.com