

四川省2013年重点图书出版规划项目

FUGAI
CUCAOJI
DE JISHU YU FANGFA

覆盖粗糙集 的技术与方法

汤建国 余堃 祝峰 著



电子科技大学出版社

四川省2013年重点图书出版规划项目

FUGAICUCAOJI
DE JISHU YU FANGFA

覆盖粗糙集 的技术与方法

汤建国 余堃 祝峰 著



电子科技大学出版社

图书在版编目(CIP)数据

覆盖粗糙集的技术与方法 / 汤建国, 余堃, 祝峰著.

—成都: 电子科技大学出版社, 2013. 1

ISBN 978-7-5647-1465-9

I. ①覆… II. ①汤… ②余… ③祝… III. ①数据模型—建立模型 IV. ①TP311.13

中国版本图书馆 CIP 数据核字 (2013) 第 012115 号

内容提要

覆盖粗糙集是一种处理不确定性问题的重要工具, 本书通过对其存在的几个关键问题进行深入分析和研究, 建立了三种覆盖粗糙集的扩展模型, 提出了一种覆盖细化的思想, 并将拟阵论引入粗糙集的研究中, 从多个角度构建了粗糙集的拟阵结构, 设计了基于拟阵的知识约简算法。本书观点新颖, 研究内容为该领域的前沿问题, 其成果不仅丰富了覆盖粗糙集的理论研究, 而且也促进了它在实际问题中的应用, 具有重要的理论和现实意义。

覆盖粗糙集的技术与方法

汤建国 余堃 祝峰 著

出版: 电子科技大学出版社(成都市一环路东一段 159 号电子信息产业大厦 邮编: 610051)

策划编辑: 张 琴 吴艳玲

责任编辑: 周 岚

封面设计: 墨创文化

主 页: www.uestcp.com.cn

电子邮箱: uestcp@uestcp.com.cn

发 行: 新华书店经销

印 刷: 成都蜀通印务有限责任公司

成品尺寸: 185 mm×260 mm 印张 8.5 字数 192 千字

版 次: 2013 年 1 月第一版

印 次: 2013 年 1 月第一次印刷

书 号: ISBN 978-7-5647-1465-9

定 价: 28.00 元

■ 版权所有 侵权必究 ■

◆ 本社发行部电话: 028-83202463; 本社邮购电话: 028-83201495。

◆ 本书如有缺页、破损、装订错误, 请寄回印刷厂调换。

前言

智能信息处理技术已成为人工智能、机器学习以及数据挖掘等众多领域中的热点研究内容。粗糙集凭借其在处理不确定性问题中的优良表现,被公认为是一种重要的智能信息处理技术,吸引了国内外众多学者的研究兴趣。随着现实世界中大量应用数据在结构和形式上日益复杂化和多样化,经典粗糙集已不再适应很多实际问题的处理需求。因此,许多学者从不同的角度对粗糙集进行了扩展性的研究。覆盖粗糙集便是粗糙集的一种重要扩展,它将经典粗糙集中的划分拓广成更为一般的覆盖,不仅丰富了粗糙集理论的研究内容,而且增强了粗糙集处理复杂数据的能力,拓宽了粗糙集的应用领域。

覆盖粗糙集在最近几年取得了较大发展,但在挖掘覆盖的深层特性、构建覆盖粗糙集的扩展模型,以及覆盖粗糙集代数结构和公理化等方面的研究还有待进一步地探索和改进。基于这些问题,本书从关键技术和方法入手进行深入的分析和研究,建立了三种覆盖粗糙集的扩展模型,提出了一种覆盖细化的思想,并将拟阵论引入粗糙集的研究中,从多个角度构建了粗糙集的拟阵结构,设计了基于拟阵的知识约简算法。本书主要的创新性研究成果体现在以下几个方面:

1. 提出了一种覆盖细化的方法,提高了覆盖近似描述精度。根据元素是否隶属于多个覆盖块,将元素分为不确定元素和确定元素。一个覆盖块中的确定元素具有该覆盖块的独有特性,利用该特性将覆盖块中的确定元素与不确定元素进行重新组合,从而实现对覆盖块的细化。通过对多种覆盖粗糙集模型在覆盖细化前后的比较发现,覆盖的细化有效地提高了各模型对目标集合的近似描述精度。

2. 建立了一类新的覆盖粗糙模糊集模型。通过充分考虑元素与其最小描述之间的关系,以及其在给定模糊集中的隶属度,提出了模糊覆盖粗糙隶属度,并建立了一类新的覆盖粗糙模糊集模型。该模型在对给定模糊集进行描述时,不仅考虑了一个元素与其他元素之间的关系,也体现了该元素在给定模糊集中的隶属度的重要作用。因此,相比其他三类模型,新模型对给定模糊集的描述更为全面。实验结果也进一步验证了这一结论。

3. 建立了覆盖 Vague 集模型。Vague 集在描述对象的模糊性时比模糊集提供的信息更为丰富。从目标集合的覆盖上、下近似与论域中各元素之间存在的不确定关系出发,构建了目标集合的覆盖 Vague 集,从一种新的角度展现了论域中各元素与目标集合之间的从属关系,对存在于覆盖粗糙集中的一些不确定现象有了更清晰的认识,即:一个集合的覆盖下近似中的元素并非完全从属于该集合,而它的覆盖上近似之外的元素也并非完全不从属于该集合。

4. 建立了基于覆盖的软粗糙集模型。通过剖析软集的特征及其与覆盖之间的内在联系后,提出了补参的概念,利用它建立了软覆盖近似空间,从而进一步构造了基于覆盖的软粗糙集模型。给出了不同软覆盖近似空间具有相同近似运算的充要条件,针对软覆盖近似空间中存在冗余的参数化子集提出了一种参数约简方法,并证明对于论域上的任

意集合，一个软覆盖近似空间与其约简产生的关于该集合的软上近似、软下近似相等。

5. 将拟阵论引入粗糙集的研究中，构建了粗糙集的三种拟阵结构，并提出了基于拟阵的知识约简方法。拟阵论是一个具有完善的公理化体系的数学工具，将其引入粗糙集的研究中，以期对粗糙集以及覆盖粗糙集的公理化研究提供新的方法。本书从拟阵中的均匀拟阵以及图论中的完全图和圈三个不同角度出发，分别构建了粗糙集的三种拟阵结构。用拟阵的方式等价刻画了粗糙集中的上、下近似等重要概念，发掘出粗糙集一些新的性质。基于完全图和圈的粗糙集拟阵结构之间的关系，证明了这两种不同的拟阵恰好为一组对偶拟阵。同时，本书提出了基于拟阵的知识约简方法，简捷有效地实现了知识表达系统的属性约简以及属性值的约简，并设计了相关的知识约简算法。

综上所述，本书从覆盖粗糙集存在的关键问题出发，分别建立了基于模糊集、Vague集和软集的三类覆盖粗糙集扩展模型，提出了一种覆盖细化的方法，并将拟阵论引入粗糙集理论的研究中，建立了三类粗糙集拟阵结构，设计了基于拟阵的知识约简算法。这些创新成果进一步丰富了覆盖粗糙集的理论研究，同时也促进了它在实际问题中的应用，并为其后续相关研究奠定了可靠的基础。

目 录

第 1 章 绪论.....	1
1.1 研究背景及意义.....	1
1.1.1 研究背景.....	1
1.1.2 研究意义.....	2
1.2 粗糙集与覆盖粗糙集的研究现状.....	3
1.2.1 粗糙集的研究现状.....	4
1.2.2 覆盖粗糙集的研究现状.....	6
1.3 本书的主要内容和创新之处.....	7
第 2 章 背景知识回顾.....	9
2.1 粗糙集理论.....	9
2.1.1 基本概念.....	9
2.1.2 粗糙集近似算子的性质.....	11
2.2 覆盖粗糙集理论.....	12
2.2.1 基本概念.....	12
2.2.2 覆盖近似算子的性质.....	13
2.2.3 几类主要的覆盖粗糙集模型.....	14
2.2.4 覆盖的约简.....	15
2.3 本章小结.....	16
第 3 章 覆盖粗糙集的扩展.....	17
3.1 引言.....	17
3.2 覆盖粗糙集与模糊集.....	18
3.2.1 模糊集基础知识.....	18
3.2.2 三类覆盖粗糙模糊集模型.....	19
3.2.3 三类覆盖粗糙模糊集存在的不足.....	20
3.3 一类新的覆盖粗糙模糊集模型.....	21
3.3.1 第四类覆盖粗糙模糊集.....	21
3.3.2 第四类覆盖粗糙模糊集的性质.....	24
3.3.3 第四类覆盖粗糙模糊集的退化.....	25
3.3.4 四类覆盖粗糙模糊集的比较分析.....	26
3.4 覆盖粗糙集与 Vague 集.....	32

3.4.1	Vague 集基础知识.....	32
3.4.2	覆盖 Vague 集.....	33
3.4.3	覆盖 Vague 集的性质.....	35
3.4.4	覆盖 Vague 集与覆盖粗糙集之间的关系.....	36
3.4.5	覆盖 Vague 集的退化.....	37
3.5	覆盖粗糙集与软集.....	38
3.5.1	软集基础知识.....	38
3.5.2	基于覆盖的软粗糙集.....	39
3.5.3	软覆盖近似空间的运算及其性质.....	42
3.5.4	软覆盖近似空间的约简.....	43
3.6	本章小结.....	44
第 4 章	覆盖的细化.....	46
4.1	引言.....	46
4.2	覆盖的细化.....	47
4.2.1	覆盖细化的概念.....	47
4.2.2	主要结论和性质.....	48
4.3	覆盖等价类.....	51
4.4	覆盖细化中的约简问题.....	53
4.5	覆盖的细化算法.....	54
4.6	覆盖细化前后各类覆盖粗糙集模型比较.....	56
4.7	本章小结.....	58
第 5 章	粗糙集的拟阵结构.....	60
5.1	引言.....	60
5.2	拟阵和图.....	61
5.2.1	拟阵.....	61
5.2.2	图.....	63
5.3	基于均匀拟阵的粗糙集拟阵结构.....	64
5.3.1	划分导出的均匀拟阵集合.....	64
5.3.2	P -UMS 的结合.....	66
5.3.3	粗糙集的拟阵结构.....	68
5.4	基于图的粗糙集拟阵结构.....	72
5.4.1	基于完全图的粗糙集拟阵结构.....	72
5.4.2	基于圈的粗糙集拟阵结构.....	78
5.4.3	完全图拟阵结构与圈拟阵结构的关系.....	85
5.5	本章小结.....	87

第 6 章 拟阵近似空间下的粗糙集知识约简	88
6.1 引言	88
6.2 粗糙集知识约简	89
6.2.1 知识表达系统	89
6.2.2 属性约简	91
6.2.3 属性值约简	93
6.3 拟阵近似空间	96
6.3.1 拟阵近似空间	96
6.3.2 拟阵近似空间下的粗糙近似算子	97
6.4 拟阵近似空间下的属性约简	99
6.4.1 基于拟阵的信息系统属性约简	99
6.4.2 基于拟阵的决策系统属性约简	100
6.4.3 基于拟阵的属性约简算法	101
6.5 拟阵近似空间下的属性值约简	104
6.5.1 基于拟阵的信息系统属性值约简	104
6.5.2 基于拟阵的决策系统属性值约简	105
6.5.3 基于拟阵的属性值约简算法	106
6.6 本章小结	109
第 7 章 结论与展望	110
7.1 结论	110
7.2 后继工作展望	111
参考文献	112

第 1 章 绪 论

1.1 研究背景及意义

1.1.1 研究背景

随着人类步入信息化社会,数字信息的规模也在急剧扩大,各种海量数据库不断涌现,无论是在数量上,还是在数据的维度和复杂度等方面都已远远超越过去。面对如此浩瀚的数据,传统的数学理论与方法已很难有效应对,而依靠人力去处理这些海量数据更是成为一种不可能完成的任务。特别是在一些高端科学的研究中,如生物技术、太空探测以及基因分析等,产生的数据量都非常巨大,数据的结构也异常复杂。这些高端科技的研究不仅会对人类社会的发展有重要的推动作用,同时也蕴藏着巨大的商机。因此,世界各国的科学家们都开始积极寻求和开发新的智能信息处理技术,利用计算机模拟人类的智能来处理各种复杂而庞大的数据,以提高对大规模数据的处理速度和利用效率。

然而,在智能信息处理的研究中存在着一个非常大的难题,那就是如何有效地应对普遍存在的信息不确定性。李德毅院士认为,不确定性的涵义很广泛,主要包括随机性、模糊性、不完全性、不稳定性和不一致性五个方面,其中随机性和模糊性是不确定性的最基本内涵。^[1]概率论是研究随机性的重要手段之一,它通过概率的方式对随机性进行量化分析,并利用分布函数来研究随机现象的统计特征。Dempster 和 Shafer 提出了证据理论^[2, 3],通过引入信任函数和似然函数来描述知识的不确定性。Shannon 将物理学中熵的概念引入信息论中,把概率和信息冗余联系起来,提出了信息熵的概念,并用它来刻画和描述离散型随机变量的随机不确定性程度。^[4]信息熵被提出后,许多学者通过对其进行扩展来解决更多具体的问题。王国胤等^[5]以条件信息熵为启发知识来设计决策表的启发式约简算法。杨明^[6]将条件信息熵引入决策表的约简问题中,提出了基于信息熵的近似约简概念,并设计了相关的约简算法。Liang 等^[7]提出互补熵的概念,将其用于度量信息系统的随机不确定性和模糊不确定性。Hu 等^[8]将信息熵扩展成为模糊信息熵,用来度量模糊信息系统中的信息量。

模糊性是由概念本身的不明晰造成的,即无法确切地给出概念的明确含义和界限,它不是由人的主观认识造成的,而是事物的一种客观属性。在现实世界中,模糊的概念大量存在,而有效处理模糊问题的方法还非常有限。针对这一现象,美国数学家 Zadeh 在 1965 年提出了模糊集理论^[9],用隶属度的方式来描述概念的这种模型性,将经典集合论中元素与集合之间明确的从属关系扩展为非明确的从属关系,给出了一套处理模糊信息的方法体系。不过,模糊集理论虽然能够较好地反映一个概念所具有的模糊性,但对

于概念的模糊性边界却很难进行刻画。1982年,波兰学者 Pawlak 提出了粗糙集理论^[10],通过一对精确的近似集合,即上近似和下近似,来对目标集合进行逼近,从而得到一个近似描述。粗糙集理论利用边界域的思想,很好地解决了概念模糊边界的描述问题。1990年, Dubois 和 Prade^[11]将模糊集和粗糙集相结合,创造性地提出了模糊粗糙集和粗糙模糊集理论。1993年, Gau 等^[12]提出了 Vague 集理论,它将模糊集中的隶属度扩展为真隶属度和假隶属度,从支持程度和反对程度两个方面去体现一个对象从属于某个集合的程度,其对概念的模糊边界是通过真、假隶属度之和与 1 之间的差值来反映。与模糊集相比, Vague 集对事物的描述更为丰富和全面。

目前,国内外虽然已经出现了许多用于处理不确定性问题的理论和方法,但在很多实际问题的解决中仍然难以达到让人满意的效果。造成这种结果的原因主要有两个方面:一方面是由于原始数据本身的不完整和不完备放大了问题的不确定性程度,增加了对其进行处理的难度;另一方面则是由于各种处理不确定性问题的理论或方法虽然都具有各自独特的性能,但同时也存在各自的不足,当所面对的数据较为复杂时,单一的理论或方法很难给予全面有效的处理。于是,人们逐渐意识到,任何单一的理论和方法都很难独自解决所有的问题,甚至很难全面地反映和处理某一个问题。因此,将这些理论和方法结合起来,互取所长,互补所短,已经成为各国学者在这个领域的一种研究共识。

由此可见,在各种数字信息急剧增加的今天,人们迫切需要一些更有效的方法来处理这些数据,而数据的不确定性对这些方法又提出了更高的要求。现有的一些方法虽然在理论和应用方面都取得了长足发展,但仍然很难单独地去处理现实中的许多问题,尤其是在复杂问题的处理上就更显得捉襟见肘,所以还需要对它们做进一步的完善和扩展。鉴于单一的方法在处理不确定性问题上的能力有限,研究多种方法之间的互补性,实现它们之间的有机结合,相互取长补短,不仅具有重要的理论意义,也有很重要的现实意义。

1.1.2 研究意义

粗糙集理论为解决数据中的不确定性问题提供了一种有效方法,它通过一对精确的集合,即下近似和上近似,来对不确定的目标集合进行确定的近似描述。在粗糙集中,认为目标集合的不确定性主要是由其上、下近似之间的差集造成的,该差集越大,则不确定性也越大;反之,不确定性则越小。与其他一些处理不确定性问题的理论相比,粗糙集最大的特点在于它无需提供问题所需处理的数据集合之外的任何先验信息,所以对问题不确定性的描述或处理可以说是比较客观的。^[13, 14]正因如此,粗糙集被广泛应用于机器学习、人工智能以及数据挖掘等众多领域,成为一种重要的智能信息处理技术。

粗糙集中的近似运算是建立在数据划分基础上的,而现实中的很多数据却都是以覆盖的形式存在。比如,一个男人在家庭中所处的角色,他可能只代表父亲或者儿子,但也有可能两者都是。再比如一堆积木,如果按照颜色分类,那么有些具有多种颜色的积木就会同时属于多个类,等等。在这些常见的生活实例中,对象的分类就会形成论域上的一个覆盖,即不同类之间可能存在非空交集。此外,在粗糙集的很多扩展研究中,如基于邻域的粗糙集^[15~19]、基于相似关系的粗糙集^[20]、基于相容关系的粗糙集^[21, 22]以及基

于优势关系的粗糙集^[23, 24]等, 它们研究的问题虽然在形式上是以关系等外在形式来表示, 但实质上都是论域上的覆盖数据, 只是产生覆盖的方式有所不同。由此可见, 不论是在现实生活中, 还是在科学理论研究中, 都广泛存在着覆盖类型的数据。因此, 研究如何对覆盖数据进行有效处理将是一件非常重要且很有必要的工作。

覆盖广义粗糙集是由 Zakowski^[25]最先提出的, 它将粗糙集中由等价关系确定的划分推广为覆盖, 是对粗糙集理论的一种重要扩展, 已成为一种处理覆盖数据的有效方法。之后, 又有许多学者陆续对覆盖粗糙集进行深入研究, 取得了很多有意义的研究成果^[26-34]。尽管对覆盖粗糙集的研究已经有了较大的发展, 但还存在许多问题有待进一步解决和完善。

首先, 已有的多类覆盖粗糙集模型对于目标集合的近似描述过于粗糙且稳定性不强。其原因是由于这些模型在建立时, 对于元素与覆盖块之间的关系没有予以充分重视和深入分析, 没能对覆盖所蕴涵的知识做进一步的挖掘。因此, 如何从给定的覆盖中尽可能挖掘出更多的知识, 是覆盖粗糙集研究中一个值得重视和需要探究的关键问题。

其次, 将覆盖粗糙集与其他方法进行结合研究的工作还不够丰富, 已有的一些相关研究也还存在不足之处需要改进。随着实际应用中数据的结构与类型愈加复杂化, 用单一的方法已很难对这些数据进行有效处理, 所以必须综合运用多种方法来共同解决问题。在覆盖粗糙集的扩展研究中, 模糊集虽然已被许多学者引入覆盖粗糙集的研究中^[32, 34-42], 但其他理论与方法还很少被用于此类探索。因此, 在覆盖粗糙集与其他理论结合研究方面, 还需要做更多深入和探索性的工作来加以丰富和完善。

最后, 覆盖粗糙集的公理化研究还显得不足。众所周知, 公理化对于一个理论体系的逻辑推理有着重要指导作用。Zhu 等最早对覆盖粗糙集的公理化展开了研究, 并做了大量的工作。^[28, 43-46]随后, Zhang 等^[47]和 Liu 等^[48]在 Zhu 的基础上又做了一些补充和完善。但由于覆盖的定义过于宽泛, 而对其进行公理化的方法又过于单一, 以及公理化研究本身就是一个循序渐进的过程, 使得这方面的研究并不完整且很难继续往下进行, 还有待进一步地完善和推进。因此, 探究一些新的方法或手段来研究覆盖粗糙集的公理化, 已成为覆盖粗糙集研究中一个亟待解决的关键问题。

随着现实应用中海量数据库在数量和规模上的不断增加和扩大, 覆盖结构的数据也越来越多, 从而对基于覆盖的数据处理技术产生更多的需求, 提出了更大的挑战。研究覆盖粗糙集的关键技术将有助于突破覆盖粗糙集中存在的技术瓶颈, 极大地促进和推广它在现实问题中的应用, 带动其理论和应用更好地向前发展。

1.2 粗糙集与覆盖粗糙集的研究现状

粗糙集从 1982 年提出到现在已跨过了 30 年的时间, 在这期间, 粗糙集的研究地域从刚开始的只局限于东欧部分国家和地区, 到现在已几乎传播到世界各国, 经历了一个快速壮大的发展阶段。当然, 除了研究地域的迅速扩大之外, 粗糙集自身在理论与应用两方面的研究也都得到了极大的发展和丰富, 出现了大量有意义的研究成果, 被广泛地应用于人工智能、机器学习以及数据挖掘等众多领域, 成为一种重要的处理不确定性问题的理

论工具, 并被列为世界上三个最重要的粒计算模型之一^[49, 50]。

1.2.1 粗糙集的研究现状

1991 年, 关于粗糙集理论的第一本专著《Rough Sets: Theoretical Aspects of Reasoning about Data》正式出版, 揭开了粗糙集理论发展的新篇章, 使粗糙集理论从此迅速走向世界, 被世界各国学者所熟知。尤其是中国学者正是从这个时期开始认识粗糙集, 并展开对粗糙集理论的广泛研究, 逐渐形成了一支人数庞大且实力雄厚的粗糙集研究队伍, 在国际粗糙集研究领域产生了重要的影响力。我国许多粗糙集研究方面的知名学者, 如刘清、张文修、梁吉业、苗夺谦、祝峰、吴伟志以及王国胤等, 基本上都是从这个时期开始进行粗糙集的研究, 他们不仅做出了很多出色的工作, 而且极大地推动了粗糙集理论在我国的传播和发展。此外, 美国圣荷西州立大学的 T.Y. Lin 教授、加拿大里贾纳大学的 Y.Y. Yao 和姚静涛两位教授以及日本前桥工科大学钟宁教授等许多国外知名学者, 也对我 国粗糙集理论的研究和发展提供了积极的支持和帮助。

近几年来, 粗糙集在国际学术界的影响力日益增强, 有关粗糙集的国内外会议也越来越多。自从 1992 年在波兰召开“第一届国际粗糙集研讨会”以来, 已陆续在世界各地定期召开了以粗糙集为主题的多个国际会议, 如 RSCTC、RSFDGrC 和 RSKT, 为粗糙集理论的广泛传播和交流提供了良好途径和氛围。在我国, 从 2001 年开始每年都举办“中国粗糙集与软计算学术会议 (CRSSC)”。2007 年, 又召开了“中国 Web 智能学术研讨会 (CWI)”和“中国粒计算学术研讨会 (CGrC)”两个会议, 成为国内三个有关粗糙集理论研究最重要的学术会议, 对我国粗糙集理论的研究和发展起到了积极的推动作用。此外, 一些国际重要学术期刊, 如 Information Sciences、Transactions on Rough Set、International Journal of Approximation Reasoning、Knowledge-based System 等, 也都大量刊登与粗糙集相关的论文, 成为了解和传播粗糙集最新研究动态和成果的重要平台。

目前, 国内外越来越多的学者已投入到粗糙集理论的研究中, 做出了大量有价值的科研成果, 使得有关粗糙集的各类文献数量大幅增加。通过对国际权威论文数据库 ScienceDirect 进行统计, 从 2008 年到现在的近五年时间里, 以“粗糙集”为主题的各类文献数目为 700 余条, 而所有关于粗糙集的文献的总数为 1600 余条。而 IEEE 网络电子期刊数据库的统计结果表明, 近五年刊登的有关粗糙集的文献数目为 2700 余条, 而所有关于粗糙集的文献总数为 4800 余条。另外, 在中国知网 CNKI 数据库中, 这一统计结果分别为 5600 余条和 10 300 余条。由此可见, 不论是在国外还是在国内, 最近几年研究粗糙集的规模都在急剧扩大, 尤其是在我国, 这一趋势更加明显。在这些文献中, 有许多是讨论粗糙集在现实问题中的应用方式, 如利用粗糙集来对知识表达系统进行属性约简和特征提取等。其余的大部分文献则是侧重于粗糙集理论方面的研究, 其内容主要包括以下四个方面:

(1) 对粗糙集模型的直接推广。Pawlak 粗糙集是建立在论域划分基础上的, 而划分又与论域上的等价关系一一对应, 所以从二元关系的角度, 也可以说粗糙集是建立在等级关系上的。由于划分和等价关系的约束过于苛刻, 很难满足大量现实问题的实际需要, 有许多学者都通过弱化划分或等价关系所要求的条件来扩展粗糙集理论。于是, 就出现

了将论域的划分扩展为覆盖,建立了覆盖粗糙集模型。^[25]通过将等价关系中的条件弱化,建立了基于相容关系^[21, 22]和基于相似关系^[20]的粗糙集模型,以及更为一般的二元关系下的粗糙集模型^[18, 51, 52]。还有学者根据属性值的偏好关系,提出了基于优势关系的粗糙集模型^[23, 24]。

(2) 将其他理论与方法引入粗糙集的研究中。大量实践证明,任何单一的理论和方法都很难独自解决所有的问题,甚至很难全面地反映和处理某一个问题。自从 Dubois 和 Prade^[11]提出了粗糙模糊集和模糊粗糙集之后,便启发了许多学者进行这方面的研究。一方面,他们进一步地研究了粗糙集与模糊集之间更深层次的关系,将之应用到实际问题中并取得了很好的效果^[36, 53~64]。另一方面,许多其他理论与方法被引入粗糙集的研究中以增强粗糙集的应用性,如概率论、证据理论、神经网络和遗传算法等^[8, 65~76];还有一些理论,如拓扑、布尔代数、概念格和拟阵论等,则被用于研究粗糙集的数学结构和性质,以期从中挖掘出粗糙集一些新的特性。^[77~86]

(3) 决策粗糙集的研究。决策粗糙集模型是由 Yao 在基于 Bayes 决策论的基础上最先提出的^[65, 87~90],与 Ziarko^[91]基于包含度提出的变精度粗糙集不同的是,决策粗糙集模型是从概率的角度对粗糙集中的一些核心概念,如属性约简、正域、负域和边界域等,给出了更具语义的一般性解释,所以从某种意义上说,变精度粗糙集是决策粗糙集的一个特例^[92]。决策粗糙集模型中的三支决策思想^[93, 94]精确地反映了粗糙集的近似原理,对现实中的决策问题能给予更为合理的解释,从而被广泛应用于一些实际问题的解决中,并取得了很好的效果^[95~99]。

(4) 基于粗糙集的粒度计算模型。粒计算的思想是由 Zadeh 最先提出的^[100],他起初将其称为“Granular Mathematics”,后经 T.Y. Lin 的建议,Zadeh 将这种思想改称为粒计算(Granular Computing,简称 GrC)。虽然学术界对于粒计算还未能给出一个正式的、精确的、能够广泛接受的定义,但这并未影响粒计算在最近几年的快速发展。粗糙集作为三个最重要的粒计算模型之一,受到了许多学者的关注。T.Y. Lin^[101~103]和 Y.Y. Yao^[104~106]对基于粗糙集的粒度计算模型做了大量的前期探索工作,并对粒计算做了更清晰的阐释,推动了粒计算的研究与发展;Qian 等^[107, 108]基于粗糙集提出了乐观多粒度粗糙集和悲观多粒度粗糙集;Qian 等^[109~111]通过提出正向近似的概念,研究了动态粒度下的概念表示和决策表示,设计出了通用粗糙特征选择加速器,提高了特征选择的效率;Yang 在 Qian 研究工作的基础上,进一步提出了更具弹性的基于优势关系的乐观多粒度粗糙集模型和悲观多粒度粗糙集模型^[112]。

粗糙集的快速的发展,不仅培养了一大批优秀的科研人才,也极大地丰富了粗糙集理论的研究,拓广了粗糙集的应用领域。不过,粗糙集的研究虽然取得了不少辉煌成就,但还存在一些棘手的问题和面临许多严峻的挑战。特别是在粗糙集的应用方面,尽管粗糙集被广泛应用于机器学习和数据挖掘等众多领域,但与处理同类型问题的方法相比,其优势在很多情况下并不明显。要想改变这一状况,就需要从更深层次去找出问题存在的根本原因,并能寻求到一种巧妙的方法予以有效解决。因此,粗糙集的未来研究任重而道远,仍需要通过各国学者的不懈努力和群体的智慧来推动它进一步地向前发展。

1.2.2 覆盖粗糙集的研究现状

覆盖广义粗糙集是由 Zakowski^[25]最先提出的,它将 Pawlak 粗糙集中论域上的划分推广为论域上的覆盖,并以此建立了一对粗糙近似运算。作为对粗糙集的一种直接拓广,覆盖粗糙集可适用于更多实际问题的数据处理工作。由于覆盖的定义较为宽泛,很难从中挖掘出更多的规律,所以有许多学者从不同的角度建立了多种基于覆盖的粗糙集模型。Pomykala 在 Zakowski 研究工作的基础上又提出了两对具有对偶性的近似算子^[26]。Tsang 等^[30]利用最小描述提出了一类新的覆盖粗糙集模型,Zhu 等^[113]对这类模型的性质进行深入研究。Zhu 等^[31]通过强化上、下近似之间的依赖性,提出了一类新的覆盖粗糙集模型。Zhu 等^[45]和 Liu 等^[48]分别利用邻域又提出了两类覆盖粗糙集模型。Qin 等^[114]利用邻域建立了 5 对具有对偶性的覆盖粗糙上、下近似算子。此外,还有其他一些学者也都从不同角度定义了一些新的覆盖粗糙集模型^[115~118]。

覆盖块的冗余问题是覆盖粗糙集研究中的一个重要内容。Zhu 等^[28]首先对这个问题进行了阐述和研究,他们发现在一些覆盖粗糙集模型中,对于论域上的任意子集,基于不同的覆盖对其进行描述,却可以得到相同的上、下近似。在对这个现象进行深入分析后,他们提出了可约元的概念,并给出了覆盖约简的一套方法。该方法不仅完美地回答了这个问题,而且还意外地解决了覆盖粗糙集研究中的一些其他关键问题^[28, 31, 43, 44, 113, 119~121]。此外,Zhu 等还对覆盖粗糙集的公理化做了大量工作^[28, 43~46],Zhang 等^[47]和 Liu 等^[48]在 Zhu 研究工作的基础上又分别做了一些补充和完善,解决了 Zhu 等提出的一些开放性问题。

Yao^[18, 122]根据串行二元关系提出了后继邻域和前趋邻域,并在分别由这两种邻域构成的覆盖上研究了 Pomykala^[26]提出的对偶近似运算。Huang 等^[123]在覆盖约简的基础上,利用信息熵来研究知识和粗糙度的测度。Li^[29]从拓扑的角度研究了覆盖粗糙集。Xu 等^[117]研究了覆盖粗糙集的一些性质。Feng 等^[37]研究了基于覆盖的粗糙模糊集。Xu 等^[32]研究了覆盖粗糙集的模糊性,并提出一类覆盖粗糙模糊集。余美真等^[124]提出了覆盖的相对约简概念,并通过实例说明相对约简与 William 所提出的覆盖约简之间的区别。孙士保等^[125, 126]研究了变精度的覆盖粗糙集。巩增泰等^[127]建立了基于覆盖关系的概率粗糙集模型,提出了该模型下的 Bayes 决策方法和应用实例。张倩倩等^[128]建立了基于覆盖的粗糙 Vague 集模型。胡军等^[42]通过对已有两种覆盖粗糙模糊集模型进行改进,提出了一种新的覆盖粗糙模糊集模型。汤建国等^[129]利用集值映射研究双论域下的覆盖粗糙集。周圣意等^[130]利用覆盖粗糙隶属函数建立了基于邻域隶属度的覆盖粗糙集模型。Zhu 等^[119~121, 131]分别研究了四类覆盖粗糙集的性质。Tang 等^[132, 133]通过深入分析元素和覆盖块之间的内在关系,提出了覆盖的细化,并在覆盖细化的基础上对多种覆盖粗糙集模型进行了比较分析。Tang 等^[132]将软集和覆盖粗糙集相结合,提出了基于覆盖的软粗糙集。Wang 和 Zhu 等^[83, 84, 134]利用拟阵研究了覆盖粗糙集的结构。Yao 等^[34]通过总结已有的各种覆盖粗糙集模型,构建了关于一个覆盖的四种邻域、六个新覆盖和两个子系统,从而提出了一个基于覆盖的粗糙近似统一框架。Wang 等^[135]研究了六类覆盖粗糙集模型之间的关系。Ge 等^[136]对论域上的覆盖从拓扑的角度进行刻画,将四类覆盖粗糙集模型的上近似算子转化为拓扑中

的闭包算子。

现实世界中覆盖数据的急剧增加,扩大了现实应用对处理此类型数据的技术需求,从而推动了覆盖粗糙集理论的快速发展。最近几年,国内外对于覆盖粗糙集的研究也逐渐趋热,出现了大量有关这方面的研究成果,为其日后的发展奠定了扎实基础。在覆盖粗糙集未来的研究中,对于覆盖数据的本质探索、覆盖粗糙集与其他理论的有机结合以及基于覆盖粗糙集的高效数据挖掘算法等,将是覆盖粗糙集研究中亟待突破的几个重要方面,这需要国内外学者对这些方面投入更多的关注和研究。

1.3 本书的主要内容和创新之处

本书针对覆盖粗糙集研究中存在的一些关键问题,展开了三个方面的研究工作。首先,针对已有几类覆盖粗糙模糊集模型中的不足,设计了一种新的覆盖粗糙模糊集模型。同时,将 Vague 集和软集分别与覆盖粗糙集相结合,提出了覆盖 Vague 集和基于覆盖的软粗糙集。其次,对一个覆盖所蕴涵的知识进行深入挖掘,提出了一种覆盖细化的思想。最后,将拟阵论引入粗糙集的研究中,建立了三种粗糙集的拟阵结构,为后续构建覆盖粗糙集的拟阵结构以及覆盖粗糙集公理化的研究工作打下了坚实基础。后续章节的内容大致如下:

第 2 章回顾了粗糙集和覆盖粗糙集中的一些基本知识,介绍了几类主要的覆盖粗糙集模型以及覆盖粗糙集中的约简理论,归纳整理了相关的一些覆盖粗糙集中的重要研究成果。

第 3 章分别利用模糊集、Vague 集和软集,构建了三个覆盖粗糙集的扩展模型。首先,通过深入分析已有三类覆盖粗糙模糊集模型存在的问题,提出了模糊覆盖粗糙隶属度,并在此基础上建立了一类新的覆盖粗糙模糊集模型。理论和统计实验结果均表明,这类新的覆盖粗糙模糊集与给定模糊集的贴近度均高于其他三类已有模型。其次,从覆盖粗糙集中上、下近似与论域中各元素之间关系存在的不确定性出发,构建了任意覆盖上给定集合的覆盖 Vague 集,从一种新的角度展现了论域中各元素与给定集合之间的从属关系。最后,通过深入分析软集的特征及其与覆盖之间的内在联系,提出了补参的概念,建立了软覆盖近似空间,并构造了基于覆盖的软粗糙集模型。

第 4 章从分析元素与覆盖中覆盖块之间的关系入手,探索从一个覆盖中挖掘出更多信息的方法。通过提出确定元素和不确定元素的概念,给出了覆盖细化的定义和实现覆盖细化的算法,探讨了覆盖细化与覆盖粗糙集中一些已有的重要概念之间的关系,研究了覆盖细化的一些基本性质,并讨论了覆盖细化前后,六种覆盖粗糙集模型的上、下近似的变化情况。

第 5 章将粗糙集与拟阵论相结合,从三个不同的角度出发,建立了三种粗糙集的拟阵结构。首先,将划分中的每个等价类分别转换为一个秩为 1 的均匀拟阵,再利用拟阵中的直和运算将它们结合成一个新的拟阵。在此基础上,用拟阵的方式等价刻画了粗糙集中的上、下近似等重要概念,建立了粗糙集的拟阵结构,发掘了粗糙集中一些新的性质。其次,通过剖析完全图与划分之间的关系,并根据图与拟阵之间的密切关系,建立

了基于完全图的粗糙集拟阵结构。类似地,通过发掘圈与划分之间的内在联系,建立了基于圈的拟阵结构,从另外一种图的角度阐释了粗糙集与拟阵之间的密切联系。最后,基于完全图和圈的粗糙集拟阵结构之间的关系,证明了这两种不同的拟阵恰好为一组对偶拟阵。

第 6 章在拟阵近似空间下研究了粗糙集中的知识约简问题。通过分析知识表达系统中由属性值确定的对象之间的关系,定义了一种特殊的拟阵极小圈——二元圈。利用二元圈进一步建立了拟阵近似空间,并将知识表达系统直接转换为一族拟阵近似空间,从而在一个完全的拟阵环境中来刻画粗糙集知识约简理论。此外,利用拟阵中极小圈的一些特性,针对信息系统和决策系统分别设计了基于拟阵的属性约简算法和属性值的约简算法。

第2章 背景知识回顾

粗糙集理论为解决数据中的不确定性问题提供了一种有效方法,它通过一对精确的集合,即下近似和上近似,来对不确定的目标集合进行确定的近似描述。在经典粗糙集理论中,知识是建立在等价关系基础上论域的划分,以此建立的计算模型不仅易于理解和计算,形式化也非常简洁。但在许多现实问题中,划分无法准确地表达出实际的知识,大量的知识是以覆盖的形式存在。于是,许多学者开始将经典粗糙集中的划分推广为更为一般的覆盖,建立了覆盖粗糙集理论,在理论和应用两个层面上都对粗糙集进行了丰富和促进。本章将对粗糙集理论和覆盖粗糙集理论的基本知识进行回顾,并对目前几种主要的覆盖粗糙集模型进行介绍和评述。

2.1 粗糙集理论

本节将介绍粗糙集理论中的一些基本概念,如近似空间、不可分辨关系、上近似和下近似等,并阐述这些概念之间的关系以及它们具备的一些性质。相关内容可参考文献[10]、[137]、[138]。

2.1.1 基本概念

定义 2.1 (近似空间) 设 U 是一个非空有限集合,称为论域, R 是 U 上一簇等价关系。称序对 $S = (U, R)$ 为一个近似空间或知识库。

对于一个近似空间 (U, R) 来说, R 中的任意一个等价关系 R 都可以将论域 U 分成若干个互不相交的非空子集,并且这些子集的并等于 U ,我们将由 R 得到的这些子集形成的集合称为论域 U 上的一个划分,记为 U/R 。此时,称 U/R 中的任意集合为论域 U 上一个关于 R 的等价类。

在粗糙集中, U 的任意一个子集被称为是一个概念或者范畴,若干个概念的集合称为一个知识。因此,在近似空间 (U, R) 中, R 中的任意一个等价关系 R 对论域 U 的划分都可以看成是 U 上的一个知识,划分中的每个等价类则可以看做这个知识的一个概念或者范畴。

定义 2.2 (不可分辨关系) 设 $S = (U, R)$ 是一个近似空间, $P \subseteq R$ 是一个非空子集。则称 $IND(P)$ 是 P 上的不可分辨关系,如果

$$IND(P) = \{(x, y) \in U \times U : \forall R \in P, (x, y) \in R\} \quad (2-1)$$

不可分辨关系 $IND(P)$ 可以理解为 P 中所有等价关系的交集。因此, $IND(P)$ 仍然是一个等价关系,其对论域的划分可以看成是 P 中所有等价关系对论域的划分的交集,即