

/// 粒计算研究丛书 ///

覆盖粒计算模型与方法

——基于粗糙集的视角

胡 军 王丽娟 刘财辉 王国胤 著



科学出版社

粒计算研究丛书

覆盖粒计算模型与方法 ——基于粗糙集的视角

胡 军 王丽娟 刘财辉 王国胤 著

科学出版社

北 京

内 容 简 介

粒计算是一种模拟人类解决复杂问题的理论方法,是人工智能研究领域的一个重要分支。本书从覆盖的角度基于粗糙集理论对粒计算理论方法进行系统的总结和归纳,具体内容包括:研究覆盖近似空间中概念近似的各种方法,并给出这些近似方法的主要特点;研究基于覆盖的知识表示的知识粒度层次关系,从定性比较和量化度量两个方面进行分析;研究多粒度覆盖近似空间中概念的描述方法,并给出不同方法所构成的格结构;基于覆盖粒计算理论研究知识获取的方法。

本书可供计算机、自动化等相关专业的研究人员、教师、研究生、高年级本科生和工程技术人员参考。

图书在版编目(CIP)数据

覆盖粒计算模型与方法:基于粗糙集的视角/胡军等著. —北京:科学出版社, 2018.7

(粒计算研究丛书)

ISBN 978-7-03-058024-5

I. ①覆… II. ①胡… III. ①人工智能-计算方法 IV. ①TP18

中国版本图书馆 CIP 数据核字(2018)第 131820 号

责任编辑:任 静 / 责任校对:郭瑞芝

责任印制:师艳茹 / 封面设计:华路天然

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>

河北鹏润印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2018 年 7 月第 一 版 开本: 720×1 000 1/16

2018 年 7 月第一次印刷 印张: 11 1/2

字数: 214 000

定价: 68.00 元

(如有印装质量问题, 我社负责调换)

《粒计算研究丛书》编委会

名誉主编：李德毅 张 钹

主 编：苗夺谦 王国胤 姚一豫

副 主 编：梁吉业 吴伟志 张燕平

委 员：(按拼音排序)

陈德刚 代建华 高 阳 胡清华

胡学钢 黄 兵 李德玉 李凡长

李进金 李天瑞 刘贵龙 刘 清

米据生 史开泉 史忠植 王飞跃

王 珏 王熙照 徐久成 杨 明

姚静涛 叶东毅 于 剑 张 铃

张文修 周献忠 祝 峰

秘 书：王睿智 张清华

丛 书 序

粒计算是一个新兴的、多学科交叉的研究领域。它既融入了经典的智慧，也包括了信息时代的创新。通过十多年的研究，粒计算逐渐形成了自己的哲学、理论、方法和工具，并产生了粒思维、粒逻辑、粒推理、粒分析、粒处理、粒问题求解等诸多研究课题。值得骄傲的是，中国科学工作者为粒计算研究发挥了奠基性的作用，并引导了粒计算研究的发展趋势。

在过去几年里，科学出版社出版了一系列具有广泛影响的粒计算著作，包括《粒计算：过去、现在与展望》《商空间与粒计算——结构化问题求解理论与方法》《不确定性与粒计算》等。为了更系统、全面地介绍粒计算的最新研究成果，推动粒计算研究的发展，科学出版社推出了《粒计算研究丛书》。丛书的基本编辑方式为：以粒计算为中心，每年选择该领域的一个突出热点为主题，邀请国内外粒计算和该主题方面的知名专家、学者就此主题撰文，来介绍近期相关研究成果及对未来的展望。此外，其他相关研究者对该主题撰写的稿件，经丛书编委会评审通过后，也可以列入该系列丛书。丛书与每年的粒计算研讨会建立长期合作关系，丛书的作者将捐献稿费购书，赠给研讨会的参会者。

中国有句老话，“星星之火，可以燎原”，还有句谚语，“众人拾柴火焰高”。《粒计算研究丛书》就是基于这样的理念和信念出版发行的。粒计算还处于婴儿时期，是星星之火，在我们每个人的爱心呵护下，一定能够燃烧成燎原大火。粒计算的成长，要靠大家不断地提供营养，靠大家的集体智慧，靠每一个人的独特贡献。这套丛书为大家提供了一个平台，让我们可以相互探讨和交流，共同创新和建树，推广粒计算的研究与发展。本丛书受益于粒计算研究每一位同仁的热心参与，也必将服务于从事粒计算研究的每一位科学工作者、老师和同学。

《粒计算研究丛书》的出版得到了众多学者的支持和鼓励，同时也得到了科学出版社的大力帮助。没有这些支持，也就没有本丛书。我们衷心地感谢所有给予我们支持和帮助的朋友们！

《粒计算研究丛书》编委会

2015年7月

前 言

粒计算是专门研究基于粒结构的思维方式、问题求解方法、信息处理模式的理论、方法和技术的学科,是当前智能信息处理领域中一种新的计算范式。从人工智能角度来看,粒计算是模拟人类思考和解决大规模复杂问题的自然模式,从实际问题的需要出发,用可行的满意近似解替代精确解,达到对问题简化、提高问题求解效率等目的。从数据分析与处理层面看,粒计算通过将复杂数据进行信息粒化,用信息粒代替原始数据作为计算的基本单元,可大大提高计算效率。

自粒计算提出以来,人们分别从不同的角度对粒计算理论进行了研究,如粗糙集、模糊集、商空间、云模型等,并形成了多部本领域重要的著作,包括王国胤等撰写的《云模型与粒计算》、张燕平等撰写的《粒计算、商空间及三支决策的回顾与发展》、徐久成等撰写的《粒计算及其不确定信息度量的理论与方法》、徐伟华等撰写的《基于包含度的粒计算方法与应用》、李天瑞等撰写的《大数据挖掘的原理与方法——基于粒计算与粗糙集的视角》、刘盾等撰写的《三支决策与粒计算》等。本书主要集合了作者近年来的研究成果,从覆盖的角度对粒计算理论模型和方法进行了归纳和整理,本书的出版将进一步丰富粒计算的理论和研究方法研究。

全书由胡军、王丽娟、刘财辉统稿,王国胤审稿。在编写的过程中,本书得到了研究生刘赛男、邵瑞、黄思好、潘皓安、张淳茜、马康等的参与。全书的组织结构如下:第1章绪论,由胡军、刘赛男编写;第2章覆盖粗糙集,由王丽娟编写;第3章覆盖近似空间的约简,由胡军、黄思好编写;第4章覆盖粗糙模糊集,由胡军、邵瑞、黄思好编写;第5章覆盖决策粗糙集,由胡军、潘皓安编写;第6章多粒度覆盖粗糙集和第7章多粒度覆盖粗糙模糊集,由刘财辉编写;第8章覆盖粒计算模型与知识获取,由王丽娟编写。

在本书的编写和出版过程中,得到了国家重点研发计划课题(No. 2017YFB0802303)、国家自然科学基金(61472056、61502211、61471182、61663002、61533020、61379114)、江西省自然科学基金(No. 20171BAB202034)等项目,以及重庆邮电大学出版基金的资助,在此一并致谢。

由于作者水平有限,加之时间仓促,书中难免存在疏漏之处,恳请读者批评指正。

目 录

丛书序

前言

第 1 章 绪论	1
1.1 引言	1
1.2 粒计算研究概述	1
1.3 粒计算的主要理论模型	3
1.3.1 词计算理论	3
1.3.2 商空间理论	6
1.3.3 粗糙集理论	8
1.3.4 云模型	10
1.4 基于覆盖的粒计算模型研究	11
1.5 本书的组织结构	12
参考文献	13
第 2 章 覆盖粗糙集	21
2.1 覆盖粒计算模型的研究概述	21
2.2 覆盖粗糙集模型的基本概念	23
2.3 主要的覆盖粗糙集模型	23
2.4 六种覆盖粗糙集模型间的关系	25
2.5 本章小结	32
参考文献	32
第 3 章 覆盖近似空间的约简	36
3.1 引言	36
3.2 覆盖近似空间的知识约简模型	37
3.2.1 覆盖近似空间的绝对约简	37
3.2.2 覆盖近似空间的相对约简	39
3.2.3 覆盖近似空间的知识约简	40
3.2.4 实例分析	43
3.3 覆盖近似空间的扩展及其约简	44
3.3.1 覆盖近似空间的扩展	44

3.3.2	基于扩展的覆盖近似空间的约简	47
3.3.3	实例分析	50
3.4	本章小结	51
	参考文献	51
第4章	覆盖粗糙模糊集	52
4.1	引言	52
4.2	覆盖粗糙模糊集模型	53
4.2.1	三种覆盖粗糙模糊集模型	53
4.2.2	三种覆盖粗糙模糊集的关系	56
4.2.3	覆盖粗糙模糊集模型在模糊决策中的应用	57
4.3	不同知识粒度下的覆盖粗糙模糊集	60
4.3.1	两个覆盖生成相同覆盖粗糙模糊集的充要条件	60
4.3.2	覆盖粗糙模糊集的不确定性度量	62
4.3.3	不完备信息系统的模糊决策	67
4.4	本章小结	70
	参考文献	70
第5章	覆盖决策粗糙集	73
5.1	引言	73
5.2	决策粗糙集基础	74
5.3	覆盖决策粗糙集模型	76
5.4	覆盖决策粗糙集的约简	78
5.4.1	覆盖的约简	78
5.4.2	覆盖概率决策信息系统的约简	79
5.5	实例分析	80
5.6	本章小结	81
	参考文献	82
第6章	多粒度覆盖粗糙集	83
6.1	引言	83
6.2	基本概念及性质	84
6.2.1	Pawlak 粗糙集模型的基本性质	84
6.2.2	覆盖粗糙集模型	85
6.2.3	多粒度粗糙集模型	85
6.3	多粒度覆盖粗糙集模型	86

6.3.1 四种多粒度覆盖粗糙集模型的定义	86
6.3.2 四种多粒度覆盖粗糙集模型的性质	89
6.3.3 在覆盖粒空间及其约简上生成的上、下近似的关系	96
6.4 四种多粒度覆盖粗糙集模型的关系	99
6.4.1 四种多粒度覆盖粗糙集的上、下近似的关系	99
6.4.2 四种多粒度覆盖粗糙集模型的格关系	101
6.5 多粒度覆盖粗糙集在银行信用卡审批过程中的应用	104
6.6 本章小结	105
参考文献	106
第 7 章 多粒度覆盖粗糙模糊集	108
7.1 引言	108
7.2 基本概念及性质	109
7.2.1 I 型覆盖粗糙模糊集模型	109
7.2.2 II 型覆盖粗糙模糊集模型	109
7.2.3 III 型覆盖粗糙模糊集模型	110
7.3 多粒度覆盖粗糙模糊集模型及其性质	110
7.3.1 三种多粒度覆盖粗糙模糊集的定义	110
7.3.2 三种多粒度覆盖粗糙模糊集的性质	113
7.3.3 在覆盖粒空间及其约简上生成的上、下近似的关系	122
7.4 三种多粒度覆盖粗糙模糊集模型的关系	124
7.5 本章小结	126
参考文献	127
第 8 章 覆盖粒计算模型与知识获取	129
8.1 覆盖粗糙集模型中的知识发现方法	129
8.2 多粒度覆盖粗糙集模型中的知识发现方法	132
8.3 多覆盖多粒度模型中的知识发现方法	144
8.3.1 多覆盖多粒度模型	145
8.3.2 多覆盖多粒度模型的实例分析	149
8.3.3 多覆盖多粒度模型中的近似约简	157
8.4 扩展的邻域系统粒计算模型中的知识发现方法	162
8.4.1 邻域系统的约简	164
8.4.2 扩展的邻域系统粒计算模型中的近似约简	166
8.5 本章小结	168
参考文献	168

第1章 绪 论

1.1 引 言

粒的思想广泛地存在于现实问题中,如自动机与系统论中的“分解与划分”、最优控制中的“不确定性”、区间分析中的“区间数运算”,以及D-S证据理论中的“证据”都与信息粒密切相关。人们很早就对信息粒的相关问题进行了深入的研究。1979年,Zadeh提出信息粒的概念,并将信息粒描述为一个模糊集合^[1]。1985年,Hobbs讨论了粒的分解和合并,以及如何得到不同大小的粒,并提出了产生不同大小粒的模型^[2]。Yager教授在1998年指出:发展信息粒的操作方法是当前粒计算研究的一个重要任务^[3]。

粒计算(granular computing)一词最早出现在1997年Zadeh发表的一篇题为“Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic”的文章中,并被定义为粒数学的子集和词计算的超集^[4]。Zadeh指出人的认知基础包括三个基本概念:粒化、组织及因果关系。其中,粒化是将整体分解为部分;组织是将部分合并为整体;而因果关系是指原因与结果间的联系。这些思想为以后的粒计算研究奠定了基础。

1.2 粒计算研究概述

现在,粒计算已经成为一个受到学术界广泛研究的领域,IEEE计算智能学会于2004年成立了粒计算特别小组(Task Force on GrC),并从2005年开始召开国际粒计算学术年会(IEEE International Conference on Granular Computing),我国也从2007年开始召开国内粒计算学术年会,还有RSFDGrC、RSCTC、RSKT、JRS等国际国内会议都设置了专门的粒计算专题。到目前为止,研究者已经对粒计算的理论及其应用作了大量的有意义的探索,主要工作如下。

Lin在1988年提出邻域系统并研究了邻域系统与关系数据库之间的关系^[5]。此后,Lin发表了一系列关于粒计算与邻域系统的文章,主要研究了二元关系(邻域系统、粗糙集和信任函数)下的粒计算模型,论述基于邻域系统的粒计算在粒结

构、粒表示和粒应用等方面的问题,讨论了粒计算中的模糊集和粗糙集方法,并将粒计算方法引入数据挖掘和知识发现^[6,7]。

在 Lin 的研究基础上, Yao 结合邻域系统对粒计算进行了研究^[8-10],发表了一系列研究成果^[11-14],并将它应用于数据挖掘等领域,建立了概念之间的 IF-THEN 规则与粒集合之间的包含关系,提出利用由所有划分构成的格求解一致分类问题,为数据挖掘提供了新的方法和视角。结合粗糙集理论, Yao 还探讨了粒计算方法在机器学习、数据分析、数据挖掘、规则提取、智能数据处理和粒逻辑等方面的应用,并给出了粒计算的三个层面^[15]:

- (1) 从哲学角度看,粒计算是一种结构化的思想方法;
- (2) 从应用角度看,粒计算是一个通用的结构化问题求解方法;
- (3) 从计算角度看,粒计算是一个信息处理的典型方法。

针对复杂问题求解,张铃和张钹依据人们在解决问题时能从不同的粒度世界分析和观察同一问题,并且可以容易地从一个粒度世界跳转到另一个粒度世界,建立了一种复杂问题求解的商结构形式化体系,称为商空间理论^[16,17]。近年来,在粒计算研究的热潮中,商空间理论被人们广泛认识和推广。2003 年,张铃和张钹将 Zadeh 的模糊概念与商空间理论结合,提出模糊商空间理论,为粒计算提供了新的数学模型和工具,并成功应用于数据挖掘等领域^[18-22]。

基于粗糙集理论,王国胤等提出了基于容差关系的粒计算模型,利用属性值上的容差关系给出了不完备信息系统的粒表示、粒运算规则和粒分解算法,同时结合粗糙集中的属性约简问题,提出了不完备信息系统在粒表示下属性必要性的判定条件,并探讨了粒计算在规则提取方面的应用^[23-26]。张清华、王国胤等基于商空间理论,研究了分层递阶粒计算模型,并在分层递阶商空间的不确定性度量、基于分层递阶结构的知识获取等方面取得了研究成果^[27-29]。

为了探讨粗糙集理论在各种环境下的应用, Skowron 等以包含度概念来研究粒近似空间上的 Rough 下近似和 Rough 上近似^[30-32]。刘清等在粗糙逻辑的基础上,提出了粒逻辑的概念(G-逻辑),构造了粒逻辑的近似推理系统,并将其应用于医疗诊断^[33-35]。苗夺谦等对知识的粒计算进行探讨,引入属性的重要度,并在求最小属性约简方面得到应用^[36]。王飞跃对词计算和语言动力学进行了探讨,以词计算为基础,对问题进行动态描述、分析、综合,提出了设计、控制和评估的语言动力学系统^[37]。依据人类根据具体的任务特性把相关数据和知识泛化或者特化成不同程度、不同大小的粒的能力,以及进一步根据这些粒和粒之间的关系进行问题求解的能力,郑征等提出了相容粒度空间模型,并在图像纹理识别和数据挖掘中取得了成功^[38-41]。卜东波等从信息粒度的角度剖析聚类 and 分类技术,试图使用信息粒度原理的框架来统一聚类和分类,指出从信息粒度的观点来看,聚类是在

一个统一的粒度下进行计算,而分类却是在不同的粒度下进行计算,并根据粒度原理设计了一种新的分类算法,在大规模中文文本分类的应用实践中表明这种分类算法有较强的泛化能力^[42]。Zhang 等对粒神经网络进行了探讨,并在高效知识发现中得到很好的应用^[43,44]。李道国等研究了基于粒向量空间的人工神经网络模型,在一定程度上提高了人工神经网络的时效性、知识表达的可理解性^[45]。杜伟林等根据概念格与粒度划分在概念聚类的过程中都是基于不同层次的概念结构来进行分类表示,而且粒度划分本身构成一个格结构的特点,研究了概念格与粒度划分格在概念描述与概念层次转换之间的联系,通过对概念的分层递阶来进行概念的泛化与例化,使概念在递阶方面忽略不必要的冗余信息^[46]。Yager 探讨了基于粒计算的学习方法和应用,并指出发展信息粒的操作方法是当前粒计算研究的一个重要任务^[47]。Bargiela 和 Pedrycz 也从各个侧面对粒计算的根源和实质进行了详细的探讨与总结^[48]。

总而言之,粒计算的研究范围非常广泛,所有与粒度相关的理论、方法、技术和工具都可归为粒计算的研究范畴^[49],其主要目标是发展基于认知信息的自动推理能力,是认知计算理论的基础,通过对细节的抑制而简化系统,并将系统中最重要的关系完整保留^[50]。

1.3 粒计算的主要理论模型

粒计算作为一种理论思想,其研究必须借助具体的研究理论方法,同时为这些研究理论方法提供指导。从现有的粒计算研究工作来看,当前粒计算理论和应用研究主要依托四种理论,即词计算理论、商空间理论、粗糙集理论和云模型。

1.3.1 词计算理论

高标准的精确表达,普遍存在于数学、化学、工程学和另外一些“硬”科学之中,而不精确表达却普遍存在于社会、心理、政治、历史、哲学、语言、人类学、文学、文艺及相关的领域中^[51]。针对复杂且非明晰定义的现象,无法用精确的数学方法来描述,如高、胖、秃头、美丽等,但可以用一些程度词语,如很不可能、十分不可能、极不可能等,来对某些模糊概念进行修饰。尽管普通的精确方法(如数学)在某些科学领域应用相当广泛,也一直尝试着应用到人文学科中,但人们在长期的实践中已经清楚地认识到精确的方法应用到人文学科有很大的局限性。传统方法过于精确的特点,导致它们在某些人文系统中的应用出现异常和失败。面对巨大而又复杂的人文学科系统,区别于传统方法的新方法——模糊计算方法被 Zadeh 提出。在人类的认识中,粒的模糊性直接源于无区别、相似性、

接近性以及功能性等这些概念的模糊性。此外，它也是人类的心智及感观有限的处理细节和存储信息的能力所必需的。从这一点来看，模糊信息粒化可以看作压缩缺损数据的一种方式。人类具有在不精确性、部分知识、部分确定以及部分真实的环境下作出合理决策这一不同寻常的能力，而模糊信息粒化正是这种能力的基础。在模糊逻辑中，模糊信息粒化是语言变量、模糊“IF-THEN”规则以及模糊图的基础。事实上，模糊逻辑能有效、成功地处理实际问题，在很大程度上得益于模糊信息粒化方法的使用。

词计算(computing with words)是用词语代替数进行计算及推理的方法^[52]。如何利用语言进行推理判断，这就要进行词计算。信息粒化为词计算提供了前提条件，词计算在信息粒度、语言变量和约束概念上产生了自己的理论与方法，意在解决模糊集合论的数值化隶属度函数表示法的局限性、表达的概念缺乏前后联系、逻辑表达和算子实现的复杂性问题，使它们能够更符合人类的思维特点。词计算有狭义和广义两个方面的概念。狭义的词计算理论是指利用通常意义下的数学概念和运算(如加、减、乘、除等)构造的带有语义的模糊数值型的词计算的理论体系；广义的词计算理论统指用词进行推理、用词构建原型系统和用词编程，前者是后者的基础。模糊逻辑在词计算中起核心作用，它可以近似地被认为与词计算相同。之所以要进行词计算，主要是因为两点：一是当可得到的信息不够精确，而使用传统方法得到的数值误差较大时必须进行词计算；二是当实际问题不需要精确解，允许利用不精确性、不确定性及部分真值的方法使问题简化，便于处理，从而获得鲁棒性，降低求解费用且能较好地与现实一致时，最好采用词计算。在词计算中存在两个核心问题：模糊约束的表现问题和模糊约束的繁殖问题。它们是模糊信息粒化的基本准则。

粒化是粒计算的基本问题之一，信息粒化(information granulation)是粒化的一种形式。在众多的信息粒化中，非模糊粒化的方法很多，如将问题求解空间形成划分空间，每个粒子都是精确的。但这种粒化方法不能解决很多现实问题，如将人的头部粒化为脸、鼻子、额头、耳朵、头盖、脖子等粒子，这些粒子之间没有明确的分界线，它们都是模糊的粒子。模糊信息粒化是传统信息粒化的一种推广。模糊信息粒化理论^[53](theory of fuzzy information granulation, TFIG)建立在模糊逻辑和信息粒化方法基础之上，是从人类利用模糊信息粒化方式中获得的启发，其方法的实质是数学。模糊信息粒化可以看作适用于任何概念、方法和理论的一种广义化方式，它所涉及的广义化方式如下。

(1) 模糊化(F-广义化)：这种广义化方式是用模糊集取代普通集，是普通粒化方法的推广。

(2) 粒化(G-广义化)：这种广义化方式实质是将集合分割成粒。

(3) 随机化(R-广义化): 这种方式是以随机变量去取代变量。

(4) 通常化(U-广义化): 这种方式用“通常(X 是 A)”取代“ X 是 A ”的命题。

模糊化与粒化的结合具有特殊重要性, 称为模糊粒化(F-G 广义化)。作为一种推广方式, 它可适用于任何概念、方法和理论。如果将模糊粒化应用于变量、函数及其关系等概念, 即可得到模糊逻辑中的语言变量、模糊规则集以及模糊图等基本概念。正是因为有这些概念才产生了广义的模糊逻辑。模糊粒化理论的出发点是广义约束的问题, 主要的约束类型有等式约束(equality constraint)、可能约束(possibilistic constraint)、真实约束(veristic constraint)、概率约束(probabilistic constraint)、概率值约束(probability value constraint)、随机集约束(random set constraint)、模糊图约束(fuzzy graph constraint)等。由于自然语言表示的命题包含的信息过于复杂, 如果只用简单的、非模糊的约束方式来表达命题, 无法完整地表达其所有语义, 所以需要这些更为详细的广义约束。广义约束为概念的模糊粒的分类提供了基础。在模糊信息粒化理论中, 粒的类型由定义它的约束类型来确定, 主要有可能粒、真实粒、概率粒和广义粒。

模糊信息粒化方法, 特别是它的表现形式, 如语言变量、模糊“IF-THEN”规则以及模糊图在模糊逻辑的应用方面都起着主导作用。Zadeh 指出, 除模糊逻辑外, 没有一种方法能提供概念框架及相关技术, 它能在模糊信息粒化中起主导作用。继 Zadeh 之后, 许多学者开始了有关词计算的研究工作, Wang 编写了 *Computing with words* 一书^[54]。李征等^[55,56]通过研究模糊控制器的结构, 认为模糊控制实际上是应用了信息粒化和词计算技术, 但却只是应用了该技术的初级形式, 而基于信息粒化和词计算(IGCW)的模糊控制系统, 将具有更强的信息处理和推理判断能力, 是对人类智能更高程度的模拟。他们指出, 基于信息粒化和词计算的模糊控制系统是通过信息粒化和重组、多层次的思维决策, 动态地改变下层控制器的参数和推理方法或控制规则, 因而使控制器具有变结构和多模态的特性。信息太多会延误推理计算的时间, 给系统带来不必要的处理任务; 而信息太少, 则会降低推理结果的完善性。因此, 提出了合理重新组织信息的研究课题。随着近年来智能信息处理的不断深入与普及, 特别是处理复杂系统分析与评估时的迫切需要, 人们越来越发现排除自然语言的代价太大了。首先, 从应用角度来看, 人类已习惯于用自然语言描述和分析事物, 特别是涉及社会、政治、经济和管理中的复杂过程。人类可以方便地利用以自然语言表示的前提进行推理和计算, 并得到用自然语言表达的结果。其次, 从理论角度来看, 不利用自然语言, 现有的理论很难甚至不能够处理感性信息, 而只能处理测度信息。感性信息或知识通常只能用自然语言来描述, 由于人类分辨细节和存储信息的认知能力的内在限制, 感性信息在本质上是不精确的^[37,57-59]。王飞跃利用自然语言知识和信息, 建立了

以词计算为基础的语言动力学系统(linguistic dynamic systems, LDS)。王飞跃通过融合几个不同领域的概念和方法,提出基于词计算的语言动力学系统的计算理论框架^[37]。根据这个计算理论框架,利用常规或传统数值动力学系统中已有的成熟概念和方法,对语言动力学系统进行动力学分析、设计、控制和性能评估。这些研究的目的是建立连接人类的语言知识表示与计算机的数字知识表示的桥梁,成为下一代智能化人机交互的理论基础之一。

1.3.2 商空间理论

张铃和张钹在研究问题求解时,提出了商空间理论^[16],他们指出“人类智能的公认特点,就是人们能从极不相同的粒度上观察和分析同一问题。人们不仅能在不同粒度的世界上进行问题求解,而且能够很快地从一个粒度世界跳到另一个粒度的世界,往返自如,毫无困难。这种处理不同世界的能力,正是人类问题求解的强有力的表现”。如果能够将人类的这种能力形式化,并使计算机也具备类似的能力,对于开发机器智能来讲,意义十分重大。商空间粒计算理论的主要内容包括复杂问题的商空间描述、分层递阶结构、商空间的分解与合成、商空间的粒计算、粒度空间关系的推理以及问题的启发式搜索等。商空间理论建立了一种商结构的形式化体系,给出一套解决信息融合、启发式搜索、路径规划和推理等领域问题的理论和算法,并已有一些相关研究和应用。

商空间理论模型可用一个三元组来表示,即 (X, F, T) 。其中, X 是论域, F 是属性集, T 是 X 上的拓扑结构。当取粗粒度时,即给定一个等价关系 R (或说一个划分),得到一个对应于 R 的商集(记为 $[X]$),它对应于三元组 $([X], [F], [T])$,称为对应于 R 的商空间。商空间理论就是研究各商空间之间的关系、各商空间的合成与分解和在商空间中的推理。在这个模型下,可建立对应的推理模型,并满足两个重要的性质:“保假原理”和“保真原理”。所谓“保假原理”是指若一个命题在粗粒度空间中是假的,则该命题在比它细的商空间中也一定为假;所谓“保真原理”是指若一个命题在两个较粗粒度的商空间中是真的,则在(一定条件下)其合成的商空间中对应的问题也是真的。这两个原理在商空间模型的推理中起到了很重要的作用。设在两个较粗空间 X_1, X_2 上进行求解,得出对应的问题有解,利用“保真原理”可得,在其合成的空间 X_3 上问题也有解。设 X_1, X_2 的规模分别为 s_1, s_2 ,因为一般情况下, X_3 的规模最大可达到 $s_1 s_2$ 。于是将原来要求解规模为 $s_1 s_2$ 空间中的问题,化成求解规模分别为 s_1, s_2 的两个空间中的问题。即将复杂性从“相乘”降为“相加”。张铃又将统计学上的一些方法移植到商空间粒度分析上来,得到了“弱保假原理”,即若在某商空间上问题无解,则在 X 上问题无解的概率大于 $1-a$ 。并指出,若在 X 上有解的可信度等于 d ,则在 $[X]$ 上对应

的问题有解的可信度大于或等于 d 。为了将精确粒度下的商空间的理论和方法推广到模糊粒度计算中,张铃和张钺又将模糊集合论引入商空间,证明了利用模糊等价关系可以将原来的商空间理论推广成模糊商空间理论,并指出下面的叙述是等价的^[22]:

- (1) 在 X 上给定一个模糊等价关系;
- (2) X 的商空间 $[X]$ 上给定一个归一化的等腰距离 d ;
- (3) 给定 X 的一个分层递阶结构;
- (4) 给定一个 X 的模糊知识基。

他们还指出,所有模糊等价关系构成一个完备半序格。这些结论为粒计算提供了有力的数学模型和工具。模糊商空间理论能够更好地反映人类处理不确定问题的若干特点,即信息的确定与不确定、概念的清晰与模糊都是相对的,都与问题的粒度粗细有关。因此,构造合理的分层递阶的粒结构,可以高效地求解问题 and 处理信息。他们提出扩展模糊商空间理论的途径,即可从三个方向推广商空间理论成为模糊商空间理论:

- (1) 研究的论域 X 是模糊空间;
- (2) 研究的结构 T 是模糊拓扑;
- (3) 研究的等价关系是模糊等价关系。

并得出结论:任何一个模糊的概念必存在一个相应的粒度空间,在其上该概念是清晰的;任何一个清晰的概念必存在一个相应的粒度空间,在其上该概念是模糊的。这深刻地揭示了模糊和清晰的辩证关系。

近几年来,基于商空间的粒度计算模型的应用也得到推广。徐峰等采用距离度量空间的手段研究了商空间的模糊粒度聚类,结合信息融合技术用不同粒度合成聚类结果,认为聚类可以以非均匀粒度来描述样本集^[60,61]。张旻等将商空间粒度理论应用于对数据仓库中的数据进行粒度分析,取得了很好的结果^[62]。张持健等利用商空间理论,不仅解决了模糊控制规则指数爆炸问题,而且通过控制粒度的不断变化,模糊控制系统可以同时兼顾控制的精度和速度指标,为高精度模糊控制器获得理想的稳态和瞬态性能指标提供了很好的理论方法^[63]。刘岩等针对航空相机快速返回定位问题,应用商空间理论提出了一种快速无超调定位模糊控制算法^[64]。毛军军等利用商空间理论,对生物信息学中序列比较的若干问题加以改进,从粒度角度说明生物序列比较的本质是分层递阶结构,即商空间的有序链^[65]。刘仁金等从商空间粒度理论角度分析图像分割,提出图像分割的商空间粒度原理和基于粒度合成原理的复杂纹理图像的分割算法^[20]。张燕平等^[21]在研究粒度世界的描写、划分、粒度确定以及不同粒度世界之间的关系的的基础上,给出了粒度世界的描述实例——互联网中的路由算法和称球游戏算法,有效地验证了商空间方

法处理实际问题的高效性。Zhang 等将商空间理论应用到多值信号处理中, 得到一种区别于小波分析的多值信号处理模型^[19]。

这些利用商空间理论来解决实际问题的例子说明, 当人们在面对实际复杂的、难以准确求解的问题, 或者求精确解的代价很大, 以及实际不需要精确解的问题时, 通常不是采用系统的、数学的、精确的方法去追求问题的精确解或最优解, 而是通过粒化的思想, 将实际问题的解空间转化为商空间, 再在商空间上继续求解问题, 最终利用商空间理论的“保真原理”和“保假原理”, 得到符合实际问题的较优解。人类就是采用这种自顶向下, 形成一个分层递阶的解空间结构, 使得解空间的复杂度由相乘变为相加, 避免了计算复杂度高的困难, 使得看似难以求解的问题迎刃而解。但是, 商空间理论同样缺乏实现粒度与粒度之间、粒度与粒度世界之间、粒度世界与粒度世界之间转换的高效方法。

1.3.3 粗糙集理论

Pawlak 在 20 世纪 80 年代提出了粗糙集理论, 他假设人的智能(知识)就是一种分类的能力^[66]。给定一个论域上的等价关系等价于给定了一个论域上的划分, 划分中的分块由在给定等价关系下不可分辨的元素组成, 它们构成论域上的一个知识。Pawlak 称之为在论域上给定了一个知识基, 然后讨论一个一般的概念(论域中的一个子集)如何用知识基中的知识来表示。对那些无法用知识基中的元素的并表示的概念, 借用拓扑学中的内核和闭包的概念, 引入了一对近似算子, 称为下近似和上近似算子。这一对算子实际界定了概念在给定知识基中确定和不确定的部分, 构成对概念的一个近似描述, Pawlak 将其称为粗糙集。

粗糙集理论的研究, 经过近三十年的时间, 无论是在理论模型的建立还是应用系统的研制开发上, 都已经取得了大量的成果, 形成了一套较为完善的粗糙集理论体系^[67,68]。目前, 粗糙集理论已成为处理不确定、不精确和不完备问题的重要数学理论工具, 并在机器学习、知识获取、决策分析、数据挖掘、专家系统、决策支持、归纳推理、矛盾归结、模式识别、模糊控制和医疗诊断等应用领域取得了大量的成果, 也已成为粒计算研究的主要工具之一。

Pawlak 在不分明关系和粗糙隶属函数的基础上, 基于同一等价类中的元素具有相同的隶属函数的思想, 探讨了知识粒的结构和粒度问题, 对利用不精确概念进行推理也作了讨论, 指出近似和独立是同一问题的两个方面, 将模糊集理论和粗糙集理论有机地结合在一起形成了新的处理不确定问题的方法^[69]。Polkowski 和 Skowron 使用 Rough Mereology 方法和神经网络技术, 基于知识粒化思想, 提出了一个粗糙神经计算(RNC)模型, 将粗糙集的知识基(划分块)和神经网络相结合, 形成一种高效的神经计算方法^[70]。Skowron 进一步完善了基于粗糙集的神