

/// 粒计算研究丛书 ///

# 动态知识发现与三支决策

## ——基于优势粗糙集视角

陈红梅 李少勇 罗 川 李天瑞 著



科学出版社

粒计算研究丛书

# 动态知识发现与三支决策 ——基于优势粗糙集视角

陈红梅 李少勇 罗川 李天瑞 著



科学出版社

北京

## 内 容 简 介

本书针对大数据的动态性,面向三支决策的知识维护,运用粒计算的思想,以经典优势粗糙集及其扩展模型为研究对象,借助增量学习方式和多核并行计算策略,构建大数据分析 with 挖掘的方法,力图展现优势粗糙集和三支决策视角下大数据分析处理与知识发现的最新进展。

本书可供计算机科学与技术、智能科学与技术、软件工程、自动化、控制科学与工程、管理科学与工程和应用数学等专业的教师、研究生、高年级本科生阅读,也可供科研和工程技术人员参考。

### 图书在版编目(CIP)数据

动态知识发现与三支决策:基于优势粗糙集视角/陈红梅等著.—北京:科学出版社,2017.6

(粒计算研究丛书)

ISBN 978-7-03-053058-5

I. ①动… II. ①陈… III. ①人工智能—计算方法 IV. ①TP18

中国版本图书馆 CIP 数据核字(2017) 第 125412 号

责任编辑:任 静/责任校对:郭瑞芝

责任印制:张 倩/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>

**三河市骏杰印刷有限公司** 印刷

科学出版社发行 各地新华书店经销

\*

2017 年 6 月第 一 版 开本:720×1000 1/16

2017 年 6 月第一次印刷 印张:9 3/4

字数:175 000

定价:60.00 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》。该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。该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”。该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。该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。该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。

该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，也就没有该丛书。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

## 序

2015年11月,在天津举办的2015国际粗糙集联合会议上,作为新当选的首批国际粗糙集学会会士,李天瑞教授做了一个精彩的特邀报告,题目为《PICKT: A Solution for Big Data Analysis》。该报告的主要内容是一个新颖的大数据处理 PICKT 方案。PICKT 一词来源于德语,对应英语的 peck,即啄食之意。李天瑞教授非常形象地将大数据中获得有价值的知识比喻为啄木鸟在森林里觅食。针对大数据的五个特征,采用相对应的策略: P (Parallel/Cloud Computing),运用并行计算或云计算应对数据大容量 (Volume); I (Incremental Learning),运用增量学习的方法解决数据高速率 (Velocity); C (Composite Rough Set Models),运用复合粗糙集处理数据多种类 (Variety); K (Knowledge Discovery),运用知识发现方法实现数据高价值 (Value); T (Three-way Decisions),运用三支决策处理数据的不确定性以获得最优决策。这个框架让我们对大数据分析和处理有了深度的认识。我个人非常喜欢 PICKT 方案,关于其中的 T,同李天瑞教授有过数次交流,得知他们准备出版一本专著,没想到这么快就面世了。

我有幸多次访问西南交通大学和四川省云计算与智能技术高校重点实验室,与本书的作者陈红梅教授、李天瑞教授、罗川博士和李少勇博士成为朋友。他们的科研态度、团队精神、学术成果及热情好客给我留下了深刻的印象。我很高兴在该书中看到他们在动态知识发现、大数据分析和三支决策等方面的进展和成果。

三支决策起源于经典粗糙集和决策粗糙集中正域、边界域和负域的语义解释和分析,运用贝叶斯理论的损失函数对于区间的划分进行有效的指导。最新研究显示,三支决策有更广泛的前景和意义,它给出了三元思维模式,并提供三分而治的复杂问题求解理论和方法。三支决策可以理解为基于三个粒的粒计算,反映粗糙集、区间集、阴影集和模糊集等不同领域的基本思想。在认识和解决很多实际问题时,我们常常采用三分而治的方式和方法。在新的认知时代,三支决策将会扮演一个重要角色。

该专著系统地探讨大数据处理、知识发现、粗糙集和动态三支决策的关系,给出实用的、有效的理论和方法。数据中各因素的时变性为决策有效性提出了挑战,对决策进行动态的维护以确保其有效势在必行。粗糙集中的上下近似集将论域空间划分为三个区域,从三个区域中分别诱导出三支决策中的正域、边界域和负域。近似集的动态维护本质上是对三支决策的动态维护。近似集的动态维护为三支决策理论注入了新的活力,成为其研究内容中一个分支。该书以优势关系粗糙集模型

为研究对象,以近似集动态维护和并行计算为目标,以增量学习为方法,精心选取优势关系粗糙集模型下近似集动态更新有特色的内容进行介绍,反映了优势关系粗糙集模型下近似集动态更新的最新研究成果。主要贡献包括三个方面:①全面刻画了近似集在论域、属性、属性值等不同因素变化下,动态性质的分析、数据的动态表示、算法设计、实验设计、性能的分析与测试;②选取了集值信息系统、不完备信息系统、多元关系融合的信息系统中扩展优势关系粗糙集模型,给出了对应的近似集动态更新方法的不同构建框架;③分析了优势关系粗糙集模型下,粒度、概念、近似集等的并行计算原理,设计了近似集更新的并行算法,并通过实验验证了其可行性。

该专著清晰地阐述了粗糙集和三支决策的重要概念、定理和实用算法,为大数据分析、处理与挖掘提供了理论支撑和技术支持。特别值得一提的是,作者用大量的图解和实例,突出基于粒计算和粗糙集理论应用于大数据分析的直观性,让读者能够充分理解书中的知识点。该书的出版充分展示了利用粒计算、粗糙集和三支决策理论解决大数据复杂问题的优势,同时对于推动大数据产业的快速发展具有重要的现实意义和应用价值。这是一本关于大数据、粗糙集和三支决策研究的必备参考书。我很高兴将这本优秀的著作推荐给每一位朋友和读者。

最后,我衷心地祝贺这本专著出版!

姚一豫

2017年5月

## 前 言

随着计算机、互联网等信息技术的迅猛发展，数据的产生和应用无处不在，渗透到社会生活的方方面面，如国家政策调控、医疗诊断、灾难预测与控制、教育娱乐资讯、商业运营与管理、人际交往环境与交流方式等。一方面，计算机提供了更好的数据分析和决策分析支持；另一方面，数据的产生呈指数级增长。计算能力的增强和数据的增长相辅相成。大数据成为又一鲜明的时代特征，它是极其宝贵的社会资源之一。大数据的复杂性、不确定性、随机性、动态性、体量巨大等特性对传统的数据存储、数据分析、统计学习和机器学习等都是一个严峻的挑战和机遇。精确和近似、全量（完备）和缺失、整体和部分等的衡量和取舍在大数据下需要重新考量。全面的数据分析需要新的智能信息处理技术的支撑。粗糙集理论是一个新兴的计算范式，它运用粒化的方法和近似的原理对不确定性概念进行定性近似描述，已经成功运用于数据挖掘、决策支持、图像处理等领域。以简驭繁的思想，层次化粒度空间的构造，近似区间的构建，对于大数据的处理和分析是大有裨益的。选择合适的粒度层次，摒弃微观层面的精度，提升宏观层面的洞察力，是粗糙集理论的思想精髓所在。近年来，粗糙集理论与方法得到了迅猛发展，特别是针对特定应用环境和复杂数据对象的特点，在模型构建、算法优化、方法融合等方面都得到了长足发展。

优势关系粗糙集是粗糙集理论的一个重要扩展，该方法着眼于数据值之间的偏序关系，由偏序关系诱导的粒对目标概念进行近似描述和推理。优势关系粗糙集方法能有效地应用于多属性多准则决策问题中。分析、权衡、审时度势、决策，是多属性多准则决策中的核心内容。优势关系粗糙集方法充分考虑了决策理论中属性值序关系的重要性，在粒度构建、特征选择、近似刻画、规则逻辑等各个方面将序的信息融入其中，是粗糙集理论中最为重要的扩展模型之一，并得到了很好的应用。条件属性值的优劣和粒度之间的优劣比较，以及最终决策方案的优劣比较是优势关系粗糙集方法对多属性多准则决策有效的支持。三支决策理论是不确定性数据处理和分析的重要理论范式，它将任意的论域空间根据实体评价函数划分为三个子空间，在不同子空间中分别做出接受、延迟和拒绝三种决策，给不确定性信息决策提供了有力而简洁的解决方案。在优势关系粗糙集中，通过优势粒（劣势粒）对决策的向上向下合集进行近似刻画，形成了下近似、边界域和上近似三个区间，从这三个区间中诱导出三种特性决策规则，是三支决策理论的一个具体体现和实施方案之一。



数据的动态性和体量巨大是大数据的重要特征。数据的产生、收集、更改、删除无时无刻不在悄无声息地进行着，与之密切相关的数据分析和处理亟需相应发展并完善以更好地支持决策演化。大数据的“1秒定律”，更是为数据的实时分析处理提出更高的要求。数据的体量巨大限制了批量处理的响应时间和成本，数据的动态性要求数据分析能够及时给予反馈。增量学习技术在大数据下尤为重要，成为实时处理的首选方案之一。本书从优势关系粗糙集模型中近似集的动态演化和实时维护入手，研究适应动态决策环境的数据表示、粒度变化、概念迁移、算法设计、实验分析、模型推广等内容，丰富和发展了粗糙集理论与方法，推动了粗糙集在决策领域的应用。书中选择了优势关系粗糙集方法中有特色的研究成果进行介绍，主要包括不同数据特征下不同粒度变化下基于经典优势关系粗糙集模型及其扩展模型（集值优势关系粗糙集模型、复合优势关系粗糙集模型、不完备优势关系粗糙集模型）的近似集增量更新方法研究，以及优势关系粗糙集下近似集的并行计算。这些成果不仅为动态环境下三支决策的动态维护提供了理论与实践支撑，而且为不确定性信息的分析和处理提供了新的视角和方法，为三支决策和粗糙集理论的发展拓宽了思路，同时为粗糙集理论在决策领域中应用给出了可行的方案。

全书共分七章，第1章综述了基于粗糙集的大数据分析、三支决策及优势关系粗糙集的研究现状；第2章给出了本书的预备知识，包括经典粗糙集理论，优势关系粗糙集及其扩展模型，以及粒度度量等；第3章介绍了经典优势关系粗糙集模型下对象集变化时近似集动态更新方法；第4章介绍了集值优势关系粗糙集模型下属性集变化时近似集动态更新方法；第5章介绍了不完备信息系统中属性值粗化细化时近似集动态更新方法；第6章介绍了基于容差和优势关系的复合粗糙集模型下的近似集动态维护方法；第7章介绍了基于优势关系粗糙集模型的近似集并行计算技术。

本书的工作得到了很多专家和同行的帮助，包括比利时国家核能研究中心阮达研究员，加拿大里贾那大学姚一豫教授和姚静涛教授，美国佐治亚州立大学潘毅教授，台湾科技大学洪西进教授，西南交通大学徐扬教授，秦克云教授和刘盾教授，重庆邮电大学王国胤教授，同济大学苗夺谦教授，山西大学梁吉业教授，李德玉教授和钱宇华教授，南京大学周献忠教授和商琳副教授，天津大学胡清华教授和代建华教授，浙江海洋学院吴伟志教授，河北师范大学米据生教授，河南师范大学徐久成教授，闽南师范大学李进金教授，电子科技大学祝峰教授等。

本书的出版受到国家自然科学基金项目（No. 61572406, 61573292, 61602327）的资助，在此表示衷心感谢。另外，由于水平有限，书中不足之处在所难免，敬请读者指正（联系方式：hmchen@swjtu.edu.cn）。

# 目 录

丛书序

序

前言

第 1 章 绪论	1
1.1 基于粗糙集的大数据分析	1
1.2 三支决策及其动态知识发现	2
1.3 优势关系粗糙集及其知识发现	4
1.4 本章小结	5
第 2 章 预备知识	6
2.1 经典粗糙集模型	6
2.1.1 信息系统	6
2.1.2 等价类	7
2.1.3 近似集	7
2.2 优势关系粗糙集模型	8
2.2.1 有序决策系统	8
2.2.2 近似集	9
2.3 集值优势关系粗糙集模型	11
2.3.1 集值有序决策系统	11
2.3.2 析取和合取优势关系	12
2.3.3 近似集	12
2.4 优势特性关系粗糙集模型	13
2.4.1 不完备有序决策系统	13
2.4.2 优势特性关系及关系矩阵	13
2.4.3 近似集	16
2.5 复合粗糙集模型	18
2.5.1 复合关系	18
2.5.2 近似集	19
2.6 不完备决策系统的知识粒度和信息熵	21
2.6.1 知识粒度及性质	21
2.6.2 粗糙熵及性质	22

2.7	本章小结	23
<b>第 3 章</b>	<b>优势关系粗糙集模型近似集动态更新方法</b>	<b>24</b>
3.1	对象集变化时近似集动态更新	25
3.1.1	添加对象时近似集动态更新	25
3.1.2	删除对象时近似集动态更新	29
3.2	算例	32
3.3	算法设计与分析	35
3.4	实验评估与性能分析	36
3.5	本章小结	40
<b>第 4 章</b>	<b>集值优势关系粗糙集模型近似集动态更新方法</b>	<b>41</b>
4.1	面向属性集更新的动态粗糙集模型	42
4.1.1	基于矩阵运算的近似集构造方法	42
4.1.2	添加属性时近似集动态更新	46
4.1.3	删除属性时近似集动态更新	47
4.2	算法设计与分析	48
4.3	算例	53
4.4	实验方案与性能分析	56
4.4.1	实验方案	56
4.4.2	性能分析	57
4.5	本章小结	60
<b>第 5 章</b>	<b>不完备信息系统中优势关系粗糙集模型近似集动态更新方法</b>	<b>61</b>
5.1	属性值粗化细化背景及定义	61
5.1.1	属性值粗化细化的背景	61
5.1.2	属性值粗化细化的定义	63
5.2	属性值细化时近似集的动态更新原理及算法	65
5.2.1	属性值细化时近似集的动态更新原理	65
5.2.2	属性值细化时近似集的动态更新算法	73
5.3	属性值粗化时近似集的动态更新原理及算法	81
5.3.1	属性值粗化时近似集的动态更新原理	81
5.3.2	属性值粗化时近似集的动态更新算法	86
5.4	算例	91
5.5	算法复杂度分析	92
5.6	实验方案及性能分析	92
5.6.1	实验方案	93
5.6.2	性能分析	94

---

5.7	本章小结	99
<b>第 6 章</b>	<b>复合关系粗糙集模型近似集动态维护方法</b>	<b>100</b>
6.1	复合关系粗糙集模型近似集动态更新	100
6.1.1	添加对象时近似集动态更新	100
6.1.2	删除对象时近似集动态更新	105
6.2	算例	108
6.3	算法设计与分析	111
6.4	实验方案与性能分析	113
6.5	本章小结	116
<b>第 7 章</b>	<b>优势关系粗糙集方法近似集并行计算算法研究</b>	<b>118</b>
7.1	属性集变化时近似集的并行计算原理	118
7.2	算例	120
7.3	算法设计与分析	125
7.4	实验方案与性能分析	128
7.5	本章小结	130
	<b>参考文献</b>	<b>131</b>

# 第1章 绪 论

## 1.1 基于粗糙集的大数据分析

近年来,随着信息技术的迅猛发展,计算机应用融入人们生活的方方面面,数据采集的可能性和驱动性不断攀升,大数据应运而生。大数据具有 Volume (海量)、Velocity (高速)、Variety (多样)、Value (低价值密度)、Veracity (真实性)等特点。一方面,信息不确定性的激增,使得智能信息分析处理的方法和原理的有效性变得尤为重要;另一方面,如何将智能信息处理的方法应用于大数据环境,是一个新的挑战。粗糙集理论是进行不确定信息分析的重要数学方法之一,它采用近似的原理,基于粒化的思想,对不确定信息进行分析,从而对决策分析与知识发现提供支持。粒化的思想、近似的原理在处理大数据中具有独特的优势,学者们对如何有效地将粗糙集理论应用于大数据的知识挖掘进行了有益的探索和研究,给出了可行的解决方案,为大数据的分析提供了行之有效的方案。针对大数据的动态性,构建增量学习模型,充分利用已有信息,是解决大数据动态特征的可行思路。针对大数据的体量大,采用分而治之、并行计算的方法,是现代计算分析的必由之路。针对大数据数据类型复杂、关系类型多等特点,研究复杂数据的分析处理模型迫在眉睫。大数据的出现,不仅是挑战,也是不确定信息分析理论发展的重要机遇,两者的发展相辅相成。在粗糙集理论中,针对大数据的动态性、多样性、体量大等特征,学者们进行了大量的研究,取得了丰硕的成果 [1]。

分布式计算、并行计算在一定程度上满足了人们对大规模数据处理的技术需求,是解决大数据挖掘问题的重要途径。如何对基于粗糙集理论的数据挖掘和知识发现算法进行并行优化引起了许多学者的研究兴趣。杨明等提出了一种基于水平分布的多决策表全局属性核求解方法,为分布式环境下的多决策表属性核求解提供了一条有效途径 [2]。Yang 等设计了一种基于 MapReduce 的属性约简并行算法 [3]。基于分辨矩阵, Liu 等设计了一种并行化决策规则挖掘的增量算法 [4]。Zhang 等提出了基于粗糙集理论的并行知识发现方法 [5],结合 Hadoop 云计算平台中 MapReduce 并行计算框架,分析了信息系统中条件划分和决策分类的并行计算方法,提出了一种基于 MapReduce 的粗糙近似集并行求解算法 [6],并在 Twister、Phoenix、Hadoop 等不同 MapReduce 平台下比较了它的性能 [7]。考虑到大数据的动态性特征, Zhang 等进一步将增量学习方法和并行计算策略相结合,提出了一种云计算环境中粗糙近似集的并行增量求解算法 [8] 和不完备决策信息系统中并行增量计算近似集的

算法<sup>[9]</sup>。钱进等通过引入信息系统中可辨识和不可辨识对象对的概念,设计了一种基于 MapReduce 计算框架的等价类并行求解算法,并提出了基于数据并行的知识约简算法<sup>[10]</sup>。通过分析基于正域、差别矩阵和信息熵的启发式知识约简中的可并行化操作,钱进等进一步利用 MapReduce 并行计算框架提出了一种新的知识约简算法<sup>[11]</sup>。Qian 等讨论了不同粒度层次下分层决策表之间的关系,并分别提出了 MapReduce 并行计算环境下基于属性重要度和不可辨识矩阵的属性约简并行求解算法<sup>[12, 13]</sup>。徐菲菲等针对大数据环境中数据的分布式存储架构,将近似约简概念引入到基于属性依赖度和基于互信息的区间值启发式约简方法中,并提出了基于多决策表的区间值全局近似约简方法,极大地降低了大数据分析的难度<sup>[14]</sup>。Li 等进一步在多核集群计算环境中提出了优势粗糙集模型中近似集的并行计算方法<sup>[15]</sup>。Zhang 等提出了复合关系粗糙集中基于 GPU 的近似集并行计算方法<sup>[16]</sup>。

在粗糙集中构建增量学习的方法以有效地解决动态数据的维护问题得到了学者们的关注<sup>[17]</sup>。李天瑞等指出动态数据主要有三种更新方式,分别为:①数据对象的动态插入和删除;②属性特征的动态增加和移除;③数据取值的动态更新和修改,并给出了动态维护知识的系列方法<sup>[1]</sup>。Liang 等研究了当对象集加入论域中信息熵的动态更新原理,进而给出了动态更新约简的算法<sup>[18]</sup>。Wang 等分别研究了对象集增加和属性集增加时信息熵增量更新原理,进而给出了约简动态维护算法<sup>[19, 20]</sup>。Shu 等研究了不完备决策信息系统中当增加一个对象时约简的更新方法<sup>[21, 22]</sup>。Yang 等研究了当对象增加时模糊粗糙集、变精度粗糙集模型中约简的动态维护<sup>[23, 24]</sup>。Sang 等提出了基于局部粗糙集和动态粒度的原理动态维护决策粗糙集模型下属性约简的方法<sup>[25-27]</sup>。Raza 等提出了在粗糙集中运用增量依赖类来更新属性约简<sup>[28]</sup>。Fan 等提出了增加对象时规则的动态诱导算法<sup>[29]</sup>。

从上述分析可以看出,寻找有效的途径来改善基于粒计算与粗糙集理论的知识获取模型及算法,使之可以更好地解决大数据环境中数据的体量大、动态性问题,已成为当前一项紧迫而又重要的课题。

## 1.2 三支决策及其动态知识发现

在大数据中,数据的时变性变大,信息的不确定性激增,有意义的精确描述与决策难度增加,非此即彼的二支决策对于不确定信息不能给出有效的决策。在管理科学、决策理论以及日常生活中,人们常常用到三支决策的方法而不是非此即彼的二支决策,如医疗诊断、论文及项目评审、风险评估、垃圾网页判断、分类聚类等领域。基于决策粗糙集模型, Yao 提出了三支决策理论,并将三支决策推广为一般化理论<sup>[30, 31]</sup>。在不同应用中,通过不同的准则对对象进行评价,确定其是否属于

接受、拒绝或延迟区域,从而给出不同的决策<sup>[32, 33]</sup>。

三支决策的理论一经提出就得到了众多学者的广泛关注,在三支决策应用和三支决策模型推广方面都取得了很多成果。Li 等在决策粗糙集模型下将三支决策应用于风险决策中<sup>[33]</sup>。Savchenko 等将序列三支决策理论和粒计算应用于物体识别<sup>[34]</sup>。Zhang 等将三支决策方法应用于推荐算法<sup>[35]</sup>。Qi 等研究了经典三支决策和概念格之间的联系<sup>[36]</sup>。Peters 等提出了近似三支决策并应用于社会网络研究中<sup>[37]</sup>。Li 等将三支决策应用于软件缺陷的检测<sup>[38]</sup>。Liang 等在决策粗糙集中运用语言评价建立三支决策理论并应用于群决策<sup>[39]</sup>。Dutta 等将三支决策应用于闭合供应链中的供应及反馈<sup>[40]</sup>。Chen 等在邻域粗糙集的约简中考虑三支决策<sup>[41]</sup>。Yu 等将三支决策理论应用于增量聚类<sup>[42]</sup>。Li 等在不完备的形式背景的约简和近似概念构造中运用三支决策理论<sup>[43]</sup>。Hu 等研究了三支决策中的三个基本要素:决策度量、决策条件和评价函数,并建立了多粒度下的三支决策空间和三支决策<sup>[44]</sup>,进一步研究了三支决策空间的集成<sup>[45]</sup>。Deng 等提出了模糊集下的决策三支近似集<sup>[46]</sup>。Zhao 等提出了四种基于不同参数概率的模糊粗糙集模型和概率区间模糊粗糙集模型,并基于贝叶斯理论将三支决策理论应用于不同概率模糊粗糙集的参数确定<sup>[47]</sup>。Liang 等探讨了如何从直觉模糊粗糙集中诱导三支决策的方法<sup>[48,49]</sup>。Li 等提出通过不同粒度构建三支认知概念格的方法<sup>[50]</sup>。

粗糙集理论是不确定信息分析的一种有效理论,它运用近似的原理,用确定的信息去描述不确定的概念。上、下近似算子是粗糙集理论的主要算子。基于上、下近似算子将论域划分为三个区域:正域、负域和边界域。姚一豫教授提出的三支决策理论,对粗糙集中的正域、负域和边界域的决策语义给出了合理解释:即从正域中诱导出接受规则,从负域中诱导出拒绝规则,而从边界域中诱导出延迟规则。这些规则集能有效地指导决策,即三支决策。随着数据的动态演化,构成信息系统的各个因素亦随之变化,即属性集变化、对象集变化和属性值变化,因而信息系统的粒度结构将发生变化,进而导致近似集发生变化。近似集的变化和信息系统中三个区域的变化,以及三支决策的变化是对应的。近似集的动态维护实际上是对三支决策的动态维护。近年来,近似集的动态维护得到了更多人们的重视,涌现出许多成果。

Chen 等分析了对象增加和属性值变化时粒度的变化对近似集的影响,并给出了高效更新近似集的方法<sup>[51]</sup>。Zhang 等提出了邻域粗糙集模型中对象增加时近似集动态更新的方法<sup>[52]</sup>,并利用云计算中的 MapReduce 技术,提出了对象集变化时动态更新经典粗糙集模型中近似集的并行算法<sup>[6]</sup>。Luo 等讨论了在集值有序信息系统中近似集动态更新的增量方法<sup>[53]</sup>。Liu 等在变精度粗糙集模型中定义了覆盖矩阵和精度矩阵,提出了随着对象迁入移出时获取感兴趣知识的增量算法<sup>[54]</sup>。Li 等分析了属性变化时上边界和下边界的变化,提出了增加多个属性时近似集动态更

新的有效方法 [55]。Cheng 提出了在粗糙模糊集中当多个属性增加和删除时增量更新近似集的方法, 有效地利用边界域增量更新近似集和利用截集近似集和粗糙模糊集近似集的关系来增量更新近似集 [56]。在集值信息系统中, Zhang 等提出了基于矩阵方法的近似集计算, 并通过矩阵的增量更新完成近似集的动态维护 [57]。Li 等研究了优势关系粗糙集模型中泛化决策在属性动态变化时的更新原理, 从而提出了基于优势关系矩阵的更新近似集方法 [58]。Cui 针对不完备信息系统中的非对称相似关系粗糙集, 提出了基于上下边界的近似集更新原理和方法 [59]。在集值序信息系统中, Luo 等研究了属性集泛化时, 基于矩阵的近似集快速更新算法 [60]。Chen 等研究了当属性值粗化细化时, 粒度的动态变化, 提出了经典粗糙集模型、不完备信息系统和优势关系粗糙集模型中动态更新近似集的方法 [61-63]。王磊研究了信息系统中等价类和近似集的布尔矩阵表示和计算方法, 提出了对象集、属性集和属性值分别动态变化时变精度粗糙集模型中上、下近似集的增量更新矩阵方法 [64]。

以上研究成果表明, 针对数据动态变化情形, 考虑不同粗糙集模型的特点, 借助多粒度多层次学习的框架, 构建不同的数据表示方法和增量学习策略, 人们在不同粗糙集模型下近似集增量更新算法的研究工作中已取得了众多成果, 但还有待于在实际应用中进一步验证和推广。同时, 可以看出近似集动态维护为三支决策理论及应用发展提供了有力的支持。

### 1.3 优势关系粗糙集及其知识发现

在多属性多准则决策中属性值具有偏序关系, 如信用评价、风险估计和可行性学习等决策问题 [65-69]。经典粗糙集模型中对象之间的等价关系没有考虑属性值之间的偏序关系, 因此 Greco 等提出了基于优势关系的粗糙集模型 (Dominance-based Rough Set Approach, DRSA) 以适应多属性多准则决策中的应用 [69], 并将其推广到模糊粗糙集中, 提出了基于优势关系的模糊粗糙集 [70]。近年来, 优势关系粗糙集模型的扩展、应用和动态知识发现等方面的研究都得到了广泛的关注。

Chakhar 等将优势关系粗糙集模型应用于群决策中 [65]。Inuiguchi 等提出了变精度优势关系粗糙集模型 [71]。Qian 等研究了区间值信息系统和集值信息系统中的偏序关系, 建立了这两种信息系统中基于偏序关系的粗糙集模型 [72, 73]。针对优势关系粗糙集模型中数据的缺失, Yang 等定义了相似优势关系, 随之给出了保持两种不同定义的相似优势关系的约简方法 [74]。Yang 等将偏序关系推广到区间值信息系统中, 讨论了区间值系统中的三种数据丢失情况和数据补齐的方法, 在此基础上进行规则的抽取 [75]。Hu 等提出了模糊优势关系粗糙集模型 [76]。Kotlowski 等研究了随机优势关系粗糙集模型 [77]。Huang 等在区间模糊信息系统中定义了梯度优势区间值关系, 并研究了该模型下的两种约简方法 [78]。骆公志和杨晓江定义



了不完备信息系统中的限制优势关系<sup>[79]</sup>。Zhang 等通过直觉模糊加法算子定义了广义优势关系和广义优势粗糙集模型<sup>[27]</sup>，并研究了该广义优势粗糙集模型下的启发式约简方法。Du 等研究了条件准则具有偏序关系而决策类不仅是模糊而且具有偏序的有序模糊决策系统中优势关系粗糙集模型<sup>[26]</sup>。

综上所述，人们在优势粗糙集模型及其知识发现方面的研究工作中已取得了一定的成果<sup>[80-82]</sup>。但在实际大数据环境中如何有效地应用优势关系粗糙集模型进行科学的三支决策是值得未来继续探讨的重要课题。

## 1.4 本章小结

大数据体量大、动态性强、不确定性和多源异构等特点使得传统数据挖掘技术从大数据中获取有价值知识面临着一系列的巨大挑战。粒计算是一种基于信息化的大规模复杂问题求解范式，粗糙集是不确定信息近似处理的一种重要粒计算模型，而三支决策是基于符合人类认知的一种决策模式。本章综述了当前数据挖掘的研究现状，回顾了粒计算、粗糙集和三支决策理论发展动态，并系统分析、归纳了粒计算和粗糙集以及三支决策应用于海量、动态、复杂数据挖掘与分析的相关研究工作。