

MIANXIANG DONGTAI SHUJU GAOXIAO SHUXING YUEJIAN SUANFA YANJIU

面向动态数据 高效属性约简算法研究

景运革○著



西南交通大学出版社

景运革 Ⅱ 著

面向动态数据高效属性 约简算法研究



西南交通大学出版社

• 成都 •

图书在版编目 (C I P) 数据

面向动态数据高效属性约简算法研究 / 景运革著.

—成都: 西南交通大学出版社, 2017.12

ISBN 978-7-5643-5968-3

I. ①面… II. ①景… III. ①数据采集—算法分析

IV. ①TP311.13

中国版本图书馆 CIP 数据核字 (2017) 第 317829 号

面向动态数据高效属性约简算法研究

景运革 著

责任编辑 张宝华

封面设计 何东琳设计工作室

印张 11 字数 164千

成品尺寸 170 mm × 230 mm

版次 2017年12月第1版

印次 2017年12月第1次

印刷 成都蓉军广告印务有限责任公司

书号 ISBN 978-7-5643-5968-3

出版发行 西南交通大学出版社

网址 <http://www.xnjdcbs.com>

地址 四川省成都市二环路北一段111号
西南交通大学创新大厦21楼

邮政编码 610031

发行部电话 028-87600564 028-87600533

定价 56.00元

图书如有印装质量问题 本社负责退换

版权所有 盗版必究 举报电话: 028-87600562

| 前 言 |

随着计算机网络技术、计算机存储技术和通信技术的迅猛发展,各行各业形成的大数据既为企业的发展提供了大好机遇,也给企业带来了严峻的挑战.如何能够及时有效地从大数据中发现有用的知识和信息已成为当前十分迫切的问题.另外,各种应用数据在现实中不断地动态变化,包括旧数据的删除和新数据的增加以及一些错误数据的修订等,如何设计基于大数据的实时属性约简高效算法是当前大数据处理研究领域人们普遍关注的一个核心问题.增量学习方法可以利用原有的计算结果对新进来的知识进行动态更新和修正,使更新后的知识更具有实时性,从而极大地降低大数据分析处理对时间和空间的需求.

本书围绕基于粗糙集动态属性约简算法这一目标进行研究,具体包括粗糙集理论的基本知识和属性约简的基本概念等内容.针对决策信息系统中对象动态变化时如何有效更新属性约简的问题,探讨了计算知识粒度和正域的增量更新机制,提出了对象变化后的动态属性约简算法;针对决策信息系统中属性随着时间动态增加时如何有效更新属性约简的问题,分析了计算知识粒度和正域的增量更新原理,设计了属性增加后的动态属性约简算法;针对决策信息系统中属性值动态变化时如何有效更新属性约简的问题,分析了对象的属性值发生变化后获取知识粒度和正域的增量更新机制,提出了对象值变化后动态属性约简算法;针对如何有效更新动态大数据属性约简的问题,利用粗糙集中多粒度的概念和“分而治之”方法分析大数据属性约简机制,并分析了决策信息系统知识粒度增量更新机制,设计了基于多粒度粗糙集模

型的属性约简算法和基于多粒度粗糙集模型的动态属性约简算法；针对决策信息系统中属性增加且属性值粗化时如何有效获取约简的问题，分析了属性增加和属性值粗化计算正域的增量更新机制，提出了属性增加且属性值粗化动态属性约简算法；针对决策信息系统中属性值粗化且对象增加时如何有效计算约简的问题，分析了属性值粗化且对象增加获取正域的增量更新原理，设计了属性值粗化且对象增加的动态属性约简算法。本书的出版对从事粗糙集理论和粒计算理论研究、知识发现和数据挖掘研究的科技工作者具有重要的参考价值。

本书在出版过程中得到了国家自然科学基金项目（项目编号：N0,61703363）和运城学院院级项目（项目名称：面向动态大数据高效属性约简算法研究）的资助，在此表示衷心的感谢。

由于作者学识水平有限和时间仓促，书中难免存在疏漏和不足之处，恳请各位读者批评指正。

景运革

2017年9月

| 目 录 |

第 1 章 绪 论	1
1.1 研究背景与研究意义	1
1.2 粒计算理论的研究现状	2
1.3 粗糙集理论的研究现状	4
1.4 基于粗糙集属性约简的研究现状	6
第 2 章 预备知识	13
2.1 经典粗糙集模型	13
2.2 多粒度粗糙集模型	15
2.3 粒度度量	16
2.4 经典粗糙集模型下基于知识粒度的属性约简定义和算法	20
2.5 经典粗糙集模型下基于正域的属性约简定义和算法	22
2.6 小结	23
第 3 章 对象变化时动态属性约简算法研究	24
3.1 对象增加时基于知识粒度的动态属性约简算法	24
3.2 对象增加时基于正域的动态属性约简算法	54
3.3 小结	60
第 4 章 属性增加时动态属性约简算法研究	61
4.1 属性增加时基于知识粒度的动态属性约简算法	61

4.2 属性增加时基于正域的动态属性约简算法	82
4.3 小结	88
第 5 章 属性值变化时动态属性约简算法研究	89
5.1 属性值变化时基于知识粒度的动态属性约简算法	89
5.2 属性值粗化时基于正域的动态属性约简算法	109
5.3 小结	116
第 6 章 多粒度粗糙集模型动态属性约简算法研究	117
6.1 多粒度粗糙集模型属性约简原理与算法	117
6.2 对象增加时多粒度粗糙集模型动态属性约简原理与算法	123
6.3 对象删除时多粒度粗糙集模型动态属性约简原理与算法	126
6.4 算法复杂度分析	128
6.5 实验方案与性能分析	130
6.6 小结	139
第 7 章 属性增加且属性值细化动态属性约简算法研究	140
7.1 属性增加且属性值细化动态属性约简原理与算法	140
7.2 实验方案与性能分析	143
7.3 小结	147
第 8 章 属性值细化且对象增加动态属性约简算法研究	148
8.1 属性值细化且对象增加动态属性约简原理与算法	148
8.2 实验方案与性能分析	151
8.3 小结	154
参考文献	156

第 1 章 绪 论

1.1 研究背景与研究意义

随着计算机网络技术、通信技术和计算机存储技术的迅速发展以及计算机运算能力的不断提高,金融投资、商业领域、产品制造、医疗卫生和休闲娱乐等各行各业产生和存储的数据量随着时间的变化在迅速增长,社会已经迈进大数据时代.对于现代信息社会中的众多行业而言,大数据就是它们共同面对的严峻挑战,同时也为它们的快速发展提供了难得的机遇^[1, 2].大数据分析和处理在社会管理、经济运行、图像处理、交通管理、股市预测等各个方面都得到了广泛的应用.研究者通过对大数据进行分析和挖掘,对各行各业做出正确的决策和创新,从而在未来的市场中占有主动权和获得对市场的支配权起到积极的作用^[3, 4, 155, 157, 161, 162, 163].通信、计算机存储和计算机网络等技术的快速发展使得大数据每天都在实时变化和增长,大数据的动态变化特点已经成为大数据挖掘和分析所面临的困难和挑战^[5, 6].如何对动态变化的大数据进行高效实时处理和设计动态高效属性约简算法已经成为当前人工智能研究领域中的热点和挑战性任务^[7, 8, 156].

粒计算是知识发现中非常有效的计算工具,主要用来解决那些具有复杂的、不完整的、不确定的、模糊的大数据知识发现的问题,它为信息处理提供了用于求解“粒”和“粒”间相关联问题的方案,是目前数据挖掘和人工智能领域中模拟人的思维模式及行为方式以及解决复杂问题的有效技术之一^[9, 10].它是通过把复杂问题分解成不同的“信息粒”的方法来实现复杂问题简单化的.利用粒计算理论中多层次、多视角和分而治之的简化方法可以对大数据进行高效分析与处理,从而提供从不同粒度层次上对

复杂问题进行分析和处理的高效方法,提高实时处理问题的计算效率.粗糙集是一种非常有效的数学计算工具,可以用来分析和处理具有不精确性和不确定性的复杂问题.它在处理分析复杂数据时不需要任何先验的知识,只需对数据集进行属性、属性值的约简,便可生成数据集的关联规则,进而引导企业制订正确的决策.粗糙集和粒计算理论与方法所具有的这些优势将为大数据环境下的属性约简算法及其优化策略提供坚实的理论依据和方法指导.

另外,由于科学技术和计算机网络的发展使大量的数据出现快速变化的趋势,每时每刻都有很多实时数据动态增加到数据集中或从数据集中删除,如何设计基于动态数据属性约简的高效算法是当前信息科学研究领域中研究者普遍关注的一个热点问题.增量学习方法能够模拟人的认知机理,可以利用原有的计算结果对新进来的知识进行动态更新和修正,使更新后的知识更具有实时性,从而极大地降低动态数据分析处理对时间和空间的需求.因此,利用增量学习方法研究基于粗糙集动态属性约简的高效算法,对于提高数据挖掘效率和实时处理能力具有重要的借鉴意义.本研究运用粗糙集理论中的属性约简算法及增量学习技术研究决策信息系统中属性集、对象集以及属性值随着时间变化动态更新情况下的数据挖掘方法,从理论和算法上探讨了在不同粒度层次之间增量更新机理和属性约简的高效算法.本研究成果将有助于拓展粗糙集属性约简的应用范围,对完善粗糙集理论与方法具有重要的意义.

1.2 粒计算理论的研究现状

粒计算是通过粒的分解和合成来处理复杂问题的计算方法的,它已经成为目前人工智能、数据挖掘等研究领域模仿人类思维和求解复杂问题的主要工具^[11, 12].通过寻求合适的求解粒度和问题描述,粒计算能够把复杂的问题分解成简单的模块,利用“分而治之”技术降低求解复杂问题的计算难度,使信息处理效率得到很大的提高.粒计算理论、方法和框架已

经广泛地应用于形式概念理论、粗糙集理论、商空间理论和三支决策理论等不同领域中。1979年, Zadeh 对模糊信息粒化问题进行了讨论和分析, 提出了模糊信息粒化中粒度、因果和组织三个基本概念^[13], 进而提出了人类认知领域中的主要概念如组织、粒度和因果, 之后又和 Lin 一起设计了粒计算方法^[14]。

近几年, 国内外许多学者在粒计算理论研究方面已经取得很多成果。Yao 阐述了粒计算理论并将粒计算和解释概念相结合, 把粒度集合之间的包含关系和 If-Then 关系联系到一起, 提出了求解一致分类问题的方法^[15, 16]。Lin 对粒计算中粗糙集和模糊集方法进行了分析和比较, 提出了基于邻域关系的粒计算方法^[17]。苗夺谦等依据粒计算的形式化描述问题, 在粒度空间中定义了论域、基和粒结构三层模型的概念, 并对基、覆盖和粒结构在多层次粒度下的相关性质进行了讨论和分析^[19]。梁吉业等给出了知识粒度度量方法的概念, 分析和比较了信息熵、粗糙熵、粒度度量和知识粒度之间的关系^[6]。Qian 等在模糊粒计算信息粒度理论的基础上, 讨论了二元粒结构的偏序关系、操作符、模糊信息系统的粒度和公理化方法^[24]。袁学海等把代数中的超群理论引入到粒计算理论中, 利用正规超群来刻画 Pawlak 的近似空间^[31]。刘清等介绍了基于 Rough 逻辑语义粒的定义, 探讨和分析了其相关性质, 设计了逻辑语义粒的归结和演绎推理方法^[32]。闫林等分析了粒和粒空间之间的关系, 提出了粒语义推理并进行了讨论^[33]。折延宏等重新给出了 Zoom-out 和 Zoom-in 两个不同算子的概念, 提出了一种基于覆盖的粒计算方法^[34]。邱桃荣等针对多值信息系统不确定性问题, 给出了属性值的近似表示方式, 设计了相容信息粒的计算方法, 构造了粒计算的多层次概念框架和获取方法^[36]。

在粒计算模型构建方面, Wang 等描述了不完备数据集中属性值的容差关系, 建立了容差关系粒度空间模型, 提出了不完备数据粒表示、粒分解及粒运算方法^[20]。Zheng 等根据人能够将知识泛化成不同大小粒度的能力, 探讨了粒之间关系对解决复杂问题的能力, 构建了相容粒度空间模型和算法^[21]。仇国芳等在两个完备格之间给出了外延内涵算子和内涵外延算子的概念, 构建了概念粒计算系统的模型, 并在该系统模型下给出迭代计算概念粒的算法^[25]。胡峰等针对不完备信息系统中的粒计算问题, 给出了

粒的表示形式、粒运算的规则及粒分解的运算方法,构造了属性必要性的判定模型,设计了不完备信息系统属性约简算法^[26]. Pedrycz 等构建了模糊神经网络粒计算模型,提出了模糊建模的多层次算法^[28]. 在信息系统发生变化时,张铃等将时间变量引入到商空间模型中,构建了动态商空间模型,并在拓扑结构保持不变而论域和属性发生变化以及拓扑结构发生变化而论域和属性保持不变两种情况下,讨论了信息系统动态商空间的属性和性质^[30].

在粒计算应用研究方面,苗夺谦等描述了知识粒度和属性重要度的概念,引入了系统协调度计算方法,并把这些概念和计算方法应用到属性约简算法和决策树构造等方面^[18]. Yager 利用粒计算理论及方法来分析社交网络问题,提出了智能社交网络分析方法^[22, 23]. Pedrycz 等利用粒计算理论和方法分析复杂系统建模问题,讨论了最优分配问题和信息粒的构建^[27]. Saberi 等将粒计算理论和方法应用到个人信用风险评估中,提出了评估在线用户信用状态的粒计算方法^[29]. Mauricio 等根据多元数据模糊信息粒化的特点,构建了一种基于粒计算的自适应模糊聚类方法^[35]. Skowron 等给出了不同粒度下近似空间的描述,并将其应用于机器学习和数据挖掘方法中^[37]. Dong 等用模糊聚类技术构造时间序列信息粒,并将粒计算理论用于解决时间序列分析问题^[38].

1.3 粗糙集理论的研究现状

由于粗糙集理论能够有效地解决那些具有不精确性和不确定性的复杂问题,它已经被广泛地应用于模式识别、知识发现等领域中.

近几十年,粗糙集在理论研究方面已经取得很大进展. 数学家 Pawlak 于 1982 年首次提出了粗糙集理论与方法^[39]. Yao 等利用信息粒定义目标概念的上、下近似集,并根据上、下近似集的概念把论域分成正域、负域和边界域互不相交的三部分^[40, 41]. 由于 Pawlak 提出的经典粗糙集模型在处

理复杂类型数据时存在一定的局限性,许多研究者提出了经典粗糙集扩展模型并将其应用于处理复杂类型数据的问题上^[42].针对决策信息系统中存在缺失数据问题,Kryszkiewicz等提出了基于容差关系的粗糙集模型^[43].尹旭日等提出了带约束相似关系的粗糙集模型^[44].针对相似关系和容差关系粗糙集模型在划分对象时存在的不合理问题,王国胤提出了基于限制容差关系的粗糙集模型^[45].Yao等把统计学中贝叶斯方法和粗糙集理论相结合,提出了一种决策粗糙集模型^[46],并进一步对经典粗糙集三支决策规则进行了分析和解释^[47].Ziarko在经典粗糙集中引入误分类参数,提出了变精度粗糙集模型^[48].Slowinski等给出了相容关系上、下近似集的定义,提出了相容关系粗糙集模型^[49].Greco等为了处理信息系统中具有优势关系的数据,提出了优势关系粗糙集模型^[50].Herbert等把数学中的博弈论运用到决策粗糙集模型中,提出了博弈论粗糙集模型^[51].Pawlak等把概率论方法与变精度粗糙集模型相结合,建立了概率粗糙集模型^[52].

另外,Inuiguchi等把概率粗糙集模型策略运用到优势粗糙集模型中,建立了变精度优势关系粗糙集模型^[53].Hu等把模糊集和优势集结合在一起,建立了模糊优势粗糙集模型^[54].针对信息系统具有多论域情况,Yan等建立了双论域粗糙集模型^[55].Sun等在模糊双论域相容关系粗糙集模型基础上,给出了两种扩展模糊粗糙集模型,并把它们运用到医疗诊断分析中^[56].胡军等针对覆盖粗糙模糊集模型中的不合理问题,提出了一种改进的覆盖粗糙模糊集模型^[57].Leung等针对区间信息系统中分类规则获取问题,建立了区间信息系统粗糙集模型^[68].在集值信息系统中,Chen等在变精度粗糙集模型中引入概率粗糙集模型,建立了函数集值粗糙集模型^[69].张文修等利用随机集描述和刻画粗糙集近似算子,建立了一种基于随机集的粗糙集模型^[70].证据理论在处理经验性和主观知识问题时具有很大优势,朱琼瑶等提出了证据理论粗糙集模型,并把它运用到水质分析预警中^[71].Yao等把经典粗糙集模型中对象等价类看成对象的邻域,建立了邻域算子粗糙集模型^[72].Yao等探讨了k-step邻域信息决策系统中近似算子操作及相关性质^[73].

另外,在多粒度粗糙集模型研究方面,Qian等在粗糙集模型中引入了

多粒度概念,建立了乐观多粒度及悲观多粒度两种粗糙集模型^[156],并进一步把决策粗糙集模型运用到多粒度粗糙集模型中,构建了多粒度决策粗糙集模型^[66]. Yang 等把优势关系模型引入到多粒度粗糙集模型中,建立了乐观多粒度及悲观多粒度两种优势关系粗糙集模型,并分析和讨论了它们之间的联系^[58]. Wang 等建立了不完备序决策信息系统多粒度粗糙集模型,从而扩展了多粒度粗糙集模型的应用范围^[59]. 张明等针对悲观多粒度粗糙集和乐观多粒度粗糙集在解决实际问题中存在一定的局限性,构建了可变多粒度粗糙集模型,并分析了其相关性质,设计了可变多粒度粗糙集模型属性约简算法^[60]. Lin 等构建了邻域多粒度粗糙集模型,设计了邻域多粒度粗糙集覆盖属性约简算法^[61],进一步定义了三种多粒度覆盖粗糙集的概念,分析和探讨了三种多粒度覆盖粗糙集的相关性质^[62]. 黄娟等在多粒度覆盖粗糙集中用最小描述生成两类上、下近似算子,设计了最小描述的属性约简算法^[63]. Xu 等把相容关系运用到粗糙模糊集中,结合多粒度概念,建立了相容关系的悲观多粒度及乐观多粒度两种粗糙模糊集模型,并给出了这两种多粒度粗糙模糊集的性质及关系^[64]. 李子勇把多粒度决策粗糙集理论引入到覆盖粗糙集理论中,设计了覆盖多粒度决策理论粗糙集方法^[65]. 杨习贝等定义了混合多粒度空间的概念,构建了一种混合多粒度粗糙集模型,并对其基本性质进行了深入探讨^[67].

上述粗糙集扩展模型可以有效地解决 Pawlak 经典粗糙集模型在处理复杂类型数据时存在的局限性问题,丰富了粗糙集理论,促进了它的蓬勃发展.

1.4 基于粗糙集属性约简的研究现状

1.4.1 粗糙集静态属性约简的研究现状

粗糙集理论中属性约简的作用就是删除决策信息系统中的冗余属性、降低数据维度. 针对特征维度高的数据,属性约简可以有效解决其计算的

准确性和复杂性问题,对于提高数据挖掘效率和减少数据计算时间与占用空间具有非常重要的作用,是信息预处理中的一个关键环节.国内外许多学者对它进行了深入研究和探讨且取得了一定成绩.

在差别矩阵属性约简算法研究方面, Skowron 等介绍了决策信息系统中差别函数和差别矩阵的概念,讨论了它们的相关性质,提出了属性核计算方法及属性约简算法^[74]. Wang 等给出了差别矩阵的概念,提出了一种完备属性约简算法,其属性约简的质量取决于预先定义的属性顺序^[75]. Zhang 等^[76]和 Wang 等^[77]分别对这种完备算法做了一些改进,使该属性约简算法同样适用于没有给出属性顺序的情况,提高了属性约简质量. Chen 等针对覆盖决策信息系统,给出了覆盖决策信息系统差别矩阵表示形式,提出了覆盖决策信息系统完备属性约简算法^[78]. 周献中等给出了不同目标下的广义属性约简算法和一般意义下的差别矩阵定义^[79]. 张颖淳等利用差别矩阵设计了启发式属性约简算法,并将其应用到数据挖掘中^[80]. Miao 等给出了一致和不一致决策信息系统差别矩阵的定义,并将其应用到基于粗糙集理论属性的约简算法中^[81].

在正区域属性约简算法研究方面,刘少辉等设计了一种能够快速计算决策信息系统的正区域算法^[82]. 徐章艳等提出了一种基于基数排序计算的决策信息系统正区域算法^[83]. 刘勇等分析了决策信息系统中的不一致度性质,提出了基于二次 Hash 表的正区域属性约简算法^[84]. 冯林等针对不一致决策信息系统中计算正区域扩展方法存在着缺陷问题,提出了一种改进的正区域属性约简算法^[85]. 我们也给出了决策信息系统中布尔矩阵的表示形式,并将其应用于基于正区域的属性约简算法中^[86].

在信息表示属性约简算法研究方面,苗夺谦等把信息论中的信息熵概念引入到粗糙集理论中,从互信息知识出发,在互信息量的基础上提出了一种新的属性约简算法^[87]. 王国胤等利用条件信息熵对粗糙集中的一些基本概念、性质及运算进行了探讨,并把条件信息熵运用到启发式属性约简算法中^[88]. 陈杰等利用扩展信息熵重新定义属性重要性的概念,并把它运用到扩展信息熵属性约简算法中^[89]. 杨明给出了决策信息系统条件信息熵

近似约简的概念,设计了相应的属性约简算法^[90].商琳等针对大部分属性约简算法只能处理离散值的属性问题,提出了一种可以处理连续值的属性约简算法^[91].黄兵等从信息熵的观点出发,在不完备信息系统下,给出了粗集粗糙性及粗集覆盖知识约简的定义,提出了新的度量方法^[92].徐久成等重新定义条件熵的概念,并将它们应用于基于信息熵的启发式属性约简中^[93].陈媛等把信息熵和粗糙集理论相结合,从信息熵角度讨论决策信息系统中的一些基本概念和性质,设计了基于信息熵的属性约简算法,该算法能够加快决策信息系统的运行速度^[95].

关于属性约简算法优化方面的研究,王国胤等将“分而治之”和并行策略运用到粗糙集属性约简算法中,可以有效解决大数据属性约简的问题^[97, 98].Chen 等首先给出了优势邻域粗糙集属性约简的性质,然后设计了一种基于优势邻域粗糙集的并行约简算法,为大数据属性约简问题提供了一种新的思路^[160].冯林等针对大部分属性约简算法只能处理离散值问题,设计了一种新的属性约简算法,它能够快速获得具有连续值的属性约简^[100],并把关系数据库中 SQL 语言操作引入到粗糙集理论中,设计了相应的属性约简算法^[101].Yang 等重新定义了覆盖广义粗糙集模型下近似空间和属性约简的概念,并把它应用到覆盖广义粗糙集属性约简算法中^[104].钱宇华等提出一种能够加速启发式属性约简算法的正区域加速器,可以有效解决传统属性约简算法在计算属性约简耗时过大问题^[105].黄兵等把模糊粗糙集引入到优势粗糙集中,构建了优势模糊粗糙集模型,并在该模型下设计了相应的属性约简算法,并将该算法应用到审计风险评估中^[110].Liang 等提出了一种更有效的属性约简算法,该算法将一个大的决策信息系统分成多个子决策信息系统,首先利用信息熵知识计算所有子决策信息系统属性约简,然后合并所有子决策信息系统属性约简,最后得到整个决策信息系统的近似属性约简^[113].桑妍丽等把分布约简概念引入到悲观多粒度粗糙集模型中,选择合适粒度,提出了悲观多粒度粗糙集分布约简算法^[114].

关于属性约简算法设计方面的研究,张文修等给出了不协调目标决策信息系统中最大分布约简的概念,提出了基于不协调目标决策信息系统的

属性约简算法^[96]. 胡清华等建立了邻域决策信息系统模型并讨论了其相关性质, 构造了数值型属性选择算法, 并将其用于知识约简中^[99]. 段洁等将粗糙集理论引入到多标记数据特征选择中, 在领域粗糙集模型下, 给出了依赖度和下近似集的计算方法, 提出了基于多标记分类的属性约简算法^[102]. Zhao 等提出了决策粗糙集模型下的一种属性约简算法^[103]. Kusunoki 等建立了变精度优势粗糙集模型并分析了其相关性质, 提出了一种适用于变精度优势粗糙集模型属性约简算法^[106]. Wu 等基于证据理论提出了一种适用于不完整信息系统或决策系统的属性约简算法^[107]. Yang 等设计了一种属性约简算法, 它可用来处理区间值决策信息系统属性约简问题^[108]. Hu 等构造了基于软模糊粗糙集模型下的依赖函数, 并将其用于特征选取与知识约简算法中^[109].

此外, 于洪等针对经典粗糙集中基于正区域的约简算法缺陷, 给出了概率粗糙集模型下属性重要型和风险最小化的概念, 设计了基于风险最小化的特征选择方法^[111]. 许韦等把相似关系、可变精度与多粒度粗糙集有机地结合在一起, 提出了一种基于相似关系的变精度多粒度粗糙集的近似分布约简方法^[112]. Qian 等介绍了多粒度粗糙集模型下近似属性约简的概念, 设计了多粒度粗糙集近似约简方法, 最后通过多粒度约简获取决策规则^[115]. 李顺勇等把粒度计算引入到粗糙集理论中, 建立了多粒度粗糙集属性约简模型, 并在该模型下设计了相应的属性约简算法^[116]. Liu 等提出了多粒度粗糙集模型下规则提取框架, 在该框架基础上可以得到“OR”的决策规则^[117]. Lin 等提出了邻域多粒度粗糙集特征选择方法^[118].

1.4.2 粗糙集动态属性约简的研究现状

许多决策信息系统在现实生活中都会随着时间的变化而变化, 用传统属性约简算法求解动态变化数据属性约简时, 需要耗费大量的计算时间并占用较多的内存空间, 导致运行速度较慢, 如何动态更新属性约简并提高数据挖掘效率成为学者们普遍关注的一个问题. 近年来, 许多研究者开始

研究动态决策信息系统属性约简高效算法,以提高计算速度和分类精度。由于决策信息系统中对象集、属性集和属性值在现实生活中都可能会发生变化,许多学者分别从这三大要素变化方面对粗糙集中动态属性约简算法进行了研究。

(1) 决策信息系统中对象集变化.

刘洋等针对决策信息系统对象集动态更新问题,分析了决策信息系统中对象增加时差别矩阵动态更新的机制,提出了一种对象集动态增加时增量特征选择方法^[119]。刘薇等分析了决策信息系统中对象增加时信息熵的增量更新机制和数学表示形式,提出了一种对象集动态增加时增量属性约简算法^[194]。Xu 等分析了一些对象增加到决策信息系统情况下属性约简的更新机理,在 0-1 运算基础上,提出了一种对象集动态增加时增量属性约简算法^[120]。Liang 等分析了一些对象增加到决策信息系统情况下三种信息熵的变化机制,提出了一种对象集动态增加时增量特征选择方法^[121]。杨明针对决策信息系统对象集动态更新问题,对差别矩阵进行改进,分析了决策信息系统中对象增加时改进差别矩阵动态更新的机制,设计了增量特征选择方法^[122]。罗来鹏根据决策信息系统对象动态增加情况,分析了二叉树增量更新机制,提出了基于对象动态增加时的增量属性约简算法^[123]。钱文彬等分析了决策信息系统对象增加时核属性的动态更新机制,并在信息熵的基础上,提出了基于对象动态变化下的核属性增量更新算法^[124]。Jiang 等根据决策树自适应算法,提出了增量属性约简算法,并把所提出的增量属性约简算法应用到网络入侵数据检测中^[125]。当决策信息系统增加一个对象时,Fan 等提出了决策规则增量更新策略^[126]。在覆盖粗糙集模型下,Lang 等针对对象集动态更新问题,提出了基于对象动态变化下的增量属性约简算法^[127]。谭旭在条件熵下定义了辨识矩阵的概念,分析了决策信息系统中对象增加时辨识矩阵动态更新机制,设计了增量特征选择方法^[128]。官礼和等根据决策信息系统中对象变化情况下的信息粒度增量更新原理,在属性序下辨识矩阵的基础上,设计了基于对象动态变化下的增量属性约简算法^[129]。