

粒计算研究综述

王国胤^{1,2}, 张清华^{1,2}, 胡 军^{1,3}

(1. 重庆邮电大学 计算机科学与技术研究所, 重庆 400065; 2. 西南交通大学 信息科学与技术学院, 四川 成都 610031;
3. 西安电子科技大学 电子工程学院, 陕西 西安 710071)

摘 要:粒计算(granular computing)是当前计算智能研究领域模拟人类思维 and 解决复杂问题的新方法. 它覆盖了所有有关粒度的理论、方法和技术, 是复杂问题求解、海量数据挖掘、模糊信息处理的有效工具. 首先回顾了粒计算研究和发展状况, 介绍了粒计算的基本组成和问题, 综述了粒计算的基本模型和方法, 并讨论了它们之间的相互关系, 最后探讨了构建统一的粒计算模型、复杂问题空间的粒化、粒层之间的转换、高效的粒计算方法、新的粒计算模型、动态粒计算模型、自主粒计算模型、粒计算方法的模糊化以及粒计算模型的应用和推广等几个方面的关键问题.

关键词:粒计算; 数据挖掘; 智能信息处理; 粗糙集; 模糊集; 商空间

中图分类号: TP18 **文献标识码:** A **文章编号:** 1673-4785 (2007) 06-0008-19

An overview of granular computing

WANG Guo-yin^{1,2}, ZHANG Qing-hua^{1,2}, HU Jun^{1,3}

(1. Institute of Computer Science & Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; 2. School of Information Science & Technology, Southwest Jiaotong University, Chengdu 610031, China; 3. School of Electronic Engineering, Xidian University, Xi'an 710071, China)

Abstract: In the field of computational intelligence, granular computing (GrC) is a new way to simulate human thinking to help solve complicated problems. GrC involves all the theories, methodologies and techniques of granularity, providing a powerful tool for the solution of complex problems, massive data mining, and fuzzy information processing. In this paper, first the current situation and the development prospects of GrC are introduced, then the fundamental and existing problems related to GrC are presented and its basic models and methods summarized. Finally, some future research topics about GrC are presented, such as, uniform granular computing model, granulation of complex problem space, transformation between granule spaces, efficient granular computing algorithm, novel granular computing model, dynamic granular computing model, data-driven granular computing model, fuzzy granular computing method, and the applications of granular computing models, etc.

Key words: granular computing; data mining; intelligent information processing; rough sets; fuzzy sets; quotient space

自 Zadeh 1979 年发表论文“Fuzzy sets and information granularity”以来^[1], 研究人员对信息粒度化的思想产生了浓厚的兴趣. Zadeh 认为很多领域都存在信息粒的概念, 只是在不同领域中的表现形式不同. 自动机与系统论中的“分解与划分”、最优

控制中的“不确定性”、区间分析里的“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. Hobbs 在 1985 年直接用“粒度(granularity)”作为论文题目发表^[2], 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. 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^[3]. 1996 年, 他在 UC-Berkeley 大学访问时, 向 Zadeh 提出作“granu-

收稿日期: 2007-04-02.

基金项目: 国家自然科学基金资助项目(60573068); 重庆市教委科学技术研究资助项目(KJ060517).

lar computing”的研究,Zadeh称之为“granular mathematics”,Lin改称为“granular computing”,并缩写成 GrC.他发表了一系列关于粒计算与邻域系统的论文^[4-10],主要是研究二元关系(邻域系统、Rough 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机器发现.依据人们在解决问题时能从不同的粒度世界去分析和观察同一问题,并且很容易地从一个粒度世界转到另一个粒度世界,张钹和张铃在1990年针对复杂问题求解,建立了一种复杂问题求解的商结构形式化体系,给出了一套解决信息融合、启发式搜索、路径规划和推理等问题的理论和算法^[11-12].1997年,Zadeh进一步指出^[13],世上有3个基本概念构成人类认知的基础:粒化、组织及因果关系.其中,粒化是整体分解为部分,组织是部分结合为整体,而因果关系则涉及原因与结果间的联系.物体的粒化产生一系列的粒子,每个粒子即为一簇点(物体),这些点难以区别,或相似、或接近、或以某种功能结合在一起.一般来说,粒化在本质上是分层次的,时间可粒化为年、月、日、小时、分、秒就是大家熟悉的例子.

在Lin的研究基础上,Yao结合邻域系统对粒计算进行了详细的研究^[14-16],发表了一系列研究成果^[17-22],并将它应用于知识挖掘等领域,建立了概念之间的if-then规则与粒度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.Yao给出了粒计算的3种观点^[22]:

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.

随着粒计算研究的发展,近年来国内外又有很多学者加入到了粒计算研究的领域.为了探讨粗糙集理论在各种环境下的应用,Skowron^[23-27]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.刘清^[28-30]在Rough逻辑的基础上,提出了粒-逻辑的概念(G逻辑),构造了这种逻辑的近似推理系统,并应用于医疗诊断.近几年来,在掀起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^[31-35].2002年苗夺谦等人^[36]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.王飞跃等人^[37]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、分析和综合,提出了设计、控制和评估的语言动力学系统.王国胤等人^[38-44]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,对粒计算方法在规则提取方面进行了探索.郑征等人^[45-47]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,他们认为,人类具有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.卜东波等人^[48]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,大规模中文文本分类的应用实践表明,这种分类算法有较强的泛化能力.Zhang等人^[49-50]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.李道国等人^[51]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.杜伟林等人^[52]根据概念格^[53]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.Yager^[54]探讨了基于粒计算的学习方法和应用.Lin^[55]在2006年粒计算国际会议上提出了新的研究思路“infrastructures for AEngineering”.同时,Bargiela和Pedrycz^[56]也从各个侧面对粒计算的根源和实质进行了详细的探讨和总结.Yager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^[57].

1 粒计算的基本组成

粒计算的基本组成主要包括3部分:粒子、粒层

和粒结构.

1.1 粒子

粒子是构成粒计算模型的最基本元素^[58-59],是粒计算模型的原语.一个粒可以被解释为许多小颗粒构成的一个大个体,现实生活中,粒子无处不在,如在地图上观察洲、国家、海洋、大陆和山脉等是一些粗的粒子(大的粒子),观察省、市、区等是一些中等的粒子,而观察街道、饭店、机场等是一些相对较小的粒子.一个粒子可以被同时看作是由内部属性描述的个体元素的集合,以及由它的外部属性所描述的整体.一个粒子的存在仅仅在一个特定的环境中才有意义.一个粒子的元素可以是粒子,一个粒子也可以是另外一个粒子的元素.而衡量粒子“大小”的概念是粒度,一般来讲,对粒子进行“量化”时用粒度来反映粒化的程度^[59].

1.2 粒层

按照某个实际需求的粒化准则得到的所有粒子的全体构成一个粒层,是对问题空间的一种抽象化描述.根据某种关系或算子,问题空间产生相应的粒子.同一层的粒子内部往往具有相同的某种性质或功能.由于粒化的程度不同,导致同一问题空间会产生不同的粒层.粒层的内部结构是指在该粒层上的各个粒子组成的论域的结构,即粒子之间的相互关系.在问题求解中,选择最合适的粒层对于问题求解尤为关键,因为,在不同粒层求解同一问题的复杂度往往不同.在高一级粒层上的粒子能够分解成为下一级粒层上的多个粒子(增加一些属性),在低一级粒层上的多个粒子可以合并成高一级粒层上的粒子(忽略一些属性).粒计算模型的主要目标是能够在不同粒层上进行问题求解,且不同粒层上的解能够相互转化.

1.3 粒结构

一个粒化准则对应一个粒层,不同的粒化准则对应多个粒层,它反应了人们从不同角度、不同侧面来观察问题、理解问题、求解问题.所有粒层之间的相互联系构成一个关系结构,称为粒结构^[20].粒结构给出了一个系统或者问题的结构化描述.通过从系统思维、复杂系统理论和层次结构理论(技术)中得到的启发至少需要确定一个粒结构网^[20]中3个层次的结构:粒子的内部结构、粒子集结构和粒子网的层次结构.粒子集的集体结构可以看作是全部层次结构中一个层次或者一个粒度视图中的结构.它本身可以看作是粒的内部连接网络.对于同一个系统或者同一个问题,许多解释和描述可能是同时存在的.所以,粒结构需要被模型化为多种层次结构,

以及在一个层次结构中的不同层次.虽然一个粒子在某个粒层上被视为一个整体,但粒子内部元素(子粒子)的结构在问题求解时也很重要,因为它能提供粒子更为详细的特性.而在同一层上的粒子之间也具有某种特殊的结构,它们可能是相互独立,或者部分包含.如果同一粒层上的粒子之间的独立性越好,可能问题求解后合并起来越方便;反之,如果粒子之间的相关性越好,则问题求解后的合并工作相对越繁杂.粒子网的层次结构是对整个问题空间的概括,它的复杂性在一定程度上决定了问题求解的复杂程度.

2 粒计算的基本问题

粒计算中存在2个最基本的问题,即粒化和粒的计算.问题空间的粒化是指将问题空间分解为许多子空间,或是基于有用的信息和知识将问题空间中的个体聚集成不同的类,这些类称之为粒.粒中的元素可以理解为对应概念的实例.可以把粒计算和概念生成、知识发现和数据挖掘联系起来,因为概念生成的目的之一是对具有某些概念的粒的表示、特征化、描述和解释,而知识发现和数据挖掘就是在粒之间建立关联和因果等联系.

2.1 粒化

粒化是问题求解空间的一个构造性过程,它可以简单理解为在给定粒化准则下得到一个粒层的过程,是粒计算基础单元的构建,包括粒子、粒视图、粒网和层次结构.在不同的粒化准则下就得到多个粒层,进而得到粒层的网络结构.通常的粒化方法有自顶而下通过分解粗粒子得到细粒子的方法,和自底向上将细粒子通过合并得到粗粒子的方法.粒化过程是粒计算的必要过程.问题空间的粒化过程主要涉及粒化准则、粒化算法(方法)、粒子和粒结构的表示(描述)以及粒子和粒结构的定性(定量)描述等问题^[59].粒化准则主要是语义方面的问题,解决为什么2个对象能放进同一个粒子内的问题.它是根据实际问题求解的具体需求和具体精度要求得到的.粒化准则的一个基本要求是忽略掉那些无关紧要的细节,从而达到降低问题求解复杂度的目的.粒化方法面对实际问题,回答如何对问题空间进行粒化,采用什么算法或工具实现粒层的构造,它属于算法方面的问题.如在粗糙集理论中,如何对对象集进行划分产生粒层,如何高效实现属性的约简等问题.粒子的结构描述主要是用粒化方法得到的粒子,如何用形式化的语言表述出来,以便后面进行计算.例如在粗糙集理论模型中,粒子的表示可能是一个子集.而

在概念格理论中,粒子的表述就是一个概念,它包括概念的外延(一个对象子集)和内涵(一个属性子集) 2 部分. 粒结构的描述往往形式多样,在商空间理论模型中,粒结构是一种分层递阶的结构,在概念格模型中,粒结构是一种 Hasse 图. 粒子和粒结构的定性、定量描述主要指粒子和粒结构的大小(主要是指粒度的结果)和复杂性度量. 当前,成功的粒化方法往往都是以将解空间形成划分空间为主要的目标,这样便于将子空间上的解合成原问题空间的解,商空间理论就是这样一个成功的实例. 当然,如果用某种粒化方法形成的解空间不是划分(如覆盖),这将增加合成的复杂度.

2.2 粒的计算

以粒子为运算对象进行问题的求解或推理,是狭义的粒计算. 粒计算可以通过系统访问粒结构来解决问题,包括在层次结构中向上和向下 2 个方向的交互,以及在同一层次内的移动,主要分为 2 种^[59]:同一粒层上粒子之间相互转换和推理,不同粒层上粒子之间的转换或推理. 不同粒层之间的联系可以由映射来表示,在不同粒层上同一问题以不同的粒度、不同的细节表示,粒层之间的映射就建立了同一问题的不同细节描述之间的关系. 商空间理论模型就是通过自然投影建立了分层递阶的商空间链式结构. 粒计算的主要特点是同一问题的解可以在不同粒层之间自由转化. 正是基于这一点,人们才能用粒计算方法高效地实现复杂问题的求解. 模糊商空间上的分层递阶结构可以通过模糊等价关系的截关系建立相应的转化联系;粗糙集理论中的划分粒度可以通过属性的增加或删减来控制;而概念格理论模型中的概念粒子的相互转化可以通过改变概念的内涵来实现. 这些转化虽然方式不同,但一个共同的特点是在转化的过程中,问题求解的重要性质必须能在不同粒层上表现出来,这也是评价粒化方法好坏的一个重要指标. 如果在粒化后粒层之间的相互转化过程中,某些重要属性不能体现出来,这不但不利于问题的求解,反而会导致问题求解过程发散,从而增加问题求解的复杂度. 商空间理论模型中的“保真”和“保假”原理使得粒化后形成的商空间具有“保序”性,使得问题求解的搜索空间大大减少,复杂度由相乘变为相加.

粒计算的 2 个基本问题中,粒化是关键,它直接决定粒计算的成功与否. 因此,粒化方法是人们研究的热点问题. 目前,粒化方法很多,如基于等价关系的划分产生粒子^[17],基于模糊集产生模糊信息粒^[1],基于模糊等价关系截集产生分层递阶粒空

间^[35],基于概念格产生概念信息粒和概念知识粒^[60],基于邻域系统产生邻域粒子^[3]等等.

总之,粒计算是一个多准则学科,它从许多领域中获得其基本的思想、准则和方法,是基于不同层次粒度和细节的问题求解的一般性理论. 在粒计算的“大伞”下进行统一的研究,可以发现不同学科之间原理的关联,它与具体的学科研究是相互独立的^[59]. 一旦掌握了粒计算中的结构化思维和结构化问题求解的抽象思想,就可以很容易地在任何领域中运用.

3 粒计算的主要模型与理论方法

3.1 词计算模型

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^[61]. 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,但可以用一些程度词语,如不很可能、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. 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. 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.

词计算(computing with words)是用词语代替数进行计算及推理的方法^[62]. 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 2 个方面的概念. 狭义的模糊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

基础^[63]. 模糊逻辑在词计算中起中心作用, 它可以近似地被认为与词计算相同^[62]. 在词计算中存在 2 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.

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^[64-65] (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.

Zadeh 指出^[64], 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^[66]编写了词计算一书. 广义词计算理论的研究工作, 中国刚刚起步, 李征等人^[67-68]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, 是对人类智能更高层次的模拟. 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, 从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精确的^[69-72]. Wang 利用自然语言知识和信息, 建立以词计算为基础的语言动力学系统

(linguistic dynamic systems, LDS), 并通过融合几个不同领域的概念和方法^[37], 提出基于词计算的语言动力学系统的计算理论框架, 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.

总之, 词计算理论和方法对于复杂信息系统的模糊推理和控制非常重要, 但由于自身的局限性, 它必须和其他理论体系相结合, 才能更有效地处理复杂信息.

3.2 粗糙集模型

一个对象属于某个集合的程度随着属性粒度的不同而不同, 为了更好地刻画集合边界的模糊性, 波兰学者 Pawlak^[73]在 20 世纪 80 年代提出了粗糙集理论, 其本质思想是利用不可分辨关系 (等价关系) 来建立论域的一个划分, 得到不区分的等价类 (即不同属性粒度下的概念粒), 从而建立一个近似空间 (由不同大小的概念粒形成). 在近似空间上, 用 2 个精确的集合 (上近似集和下近似集) 来逼近一个边界模糊的集合. 如果近似空间的粒度较粗, 被近似的集合的边界域较宽, 而如果近似空间的粒度较细, 被近似集合的边界域较窄.

给定集合 X 上的一个划分等价于在 X 上给定一个等价关系 R . X/R 表示 U 上由 R 导出的所有等价类, $[x]_R$ 表示包含元素 x 的等价类, 其中 $x \in U$. Pawlak 称之为在论域上给定了一个知识基 (X, R) , 然后讨论一个一般的概念 X (U 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 (X, R) 中的集合的并来表示的集合, 借用拓扑中的内核和闭包的概念, 引入下近似和上近似的概念: $R_-(X) = \{x \in U \mid [x]_R \subseteq X\}$ 和 $R_+(X) = \{x \in U \mid [x]_R \cap X \neq \emptyset\}$. 当 $R_-(X) = R_+(X)$ 时, 就称 X 为粗糙集, 从而创立了“粗糙集理论”. 粗糙集理论是一种软计算方法. 软计算 (soft computing) 概念是由模糊集创始人 Zadeh 提出的^[61-65]. 传统的计算方法即所谓硬计算, 使用精确、固定和不变的算法来表达和解决问题, 而软计算的指导原则是利用所允许的不精确、不确定性和部分真实性得到易于处理、鲁棒性强和成本较低的解决方案, 以便更好地与现实系统相协调.

粗糙集理论的研究, 已经经历了 20 多年的时间, 无论是在系统理论、计算模型的建立和应用系统的研制开发上, 都已经取得了很多成果, 也建立了一套较为完善的粗糙集理论体系^[74-75]. 目前粗糙集理

论已是处理模糊、不精确和不完备问题的重要数学工具.它在机器学习、知识获取、决策分析、数据库的知识发现、专家系统、决策支持系统、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不少成果,业已成为粒计算研究的主要工具之一.

经典的粗糙集理论主要是针对完备信息系统的,即处理对象的所有属性值都是已知的.为了使粗糙集理论适用于对不完备信息系统的处理,目前有 2 种主要途径:一是补齐不完备的数据,二是扩充经典的 Rough 集模型.至少可以从 3 个方面扩展粗糙集理论^[63]:1) 等价关系的泛化;2) 基本知识粒度的构造和知识的表示方法的拓广;3) 粗糙集的代数方法.

等价关系的泛化问题实质是将满足等价关系的 3 个条件(自反、对称、传递)根据实际问题进行组合,得到不同的二元关系,再根据这些二元关系得到不同的模型.如 Kryszkiewicz^[76]提出的基于容差关系的扩充粗糙集模型,Stefanowski 等人^[77]提出的基于非对称相似关系的扩充粗糙集模型和基于量化容差关系的扩充粗糙集模型.王国胤^[78]分析了前面 2 种扩充模型的不足,提出了基于限制容差关系的粗糙集模型,并发现:容差关系和非对称相似关系是对不可分辨关系扩充的 2 个极端,即容差关系的条件太松,非对称相似关系的条件太紧,而限制容差关系介于二者之间.张清华等人^[79]根据不完备信息的特点,利用模糊聚类的思想,将非等价关系转化为等价关系,从而用经典的粗糙集模型来处理不完备的信息系统,这种方法的优势在于可以得到变精度的正域以及上下近似.总之,粗糙集理论在不完备信息系统中的应用,是将粗糙集理论进一步推向实用的关键之一,因为现实数据可能在一定程度上是不完备的.

基本知识粒度的构造和知识表示方法的拓广实质是将粗糙集的商集扩展成一个拓扑空间,以此保证运算的封闭性,即用 (U/R) 代替 U/R ,它是布尔代数 $(2^U, \sim, \cap, \cup)$ 的一个子代数, $(U, (U/R))$ 构成一个拓扑空间.

张文修等人^[80]详细讨论了一般关系下的粗糙集模型,粗糙集代数的公理化方法以及粗糙集系统的代数结构等问题,对变精度的粗糙集模型、概率粗糙集模型、模糊粗糙集模型和随机集粗糙集模型等进行了系统的阐述,进一步推广了经典的粗糙集理论.

近期,基于 Rough 集理论来研究粒计算的工作尤为突出^[81].Pawlak^[82]在不分明关系和 Rough 隶

属函数的基础上,利用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. Polkowski 和 Skowron 等人^[83]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 Rough 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. Skowron^[84]在文献[83]的基础上,进一步完善了基于粗糙集的神经计算方法,并在一个参数化的近似空间上,讨论了信息粒的语法、语义、分解和合成问题,给出了粒语言的概念,提出了在分布式系统中关于信息粒结构的模式.但是,他没有提出一套行之有效的粒计算系统,也未涉及分布式环境下基于粒计算的多 Agent 推理中的冲突问题,对信息粒结构模式中的参数也没有提出有效的优化技术. Peters 等人^[85]使用不分明关系将实数划分成多个子区间,将一个全域划分成若干个网格单元,每个网格单元被视为一个粒,提出了 2 个信息粒之间的邻近关系和包含关系的度量,但其提出的方法只能局限于单个传感器的样本数据. Peters 等人^[86]综述了关于 RNC 模型的主要研究线索. Lin^[3-4]基于二元关系提出了邻域 Rough 系统,建立了粒计算模型,并使用 Rough 集中的近似空间作为信息粒结构,定义了粒隶属函数,从而提出了粒 Fuzzy 集,并得出了一些重要的性质^[87]. Yao 等^[88]利用粗糙集粒计算模型来学习分类规则,用粒网格来表示学习所得的分类知识,提出了粒之间关联性的度量公式,通过搜索粒来归纳分类规则,给出了构造粒网格的算法.在研究 Rough 推理的基础上,刘清等人^[28-30]对粒逻辑进行了探讨.

3.3 商空间理论模型

张钹和张铃在研究问题求解时,提出了商空间理论^[11-12],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^[11].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^[31,34,89-94]。

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 $([X], [F], [T])$,称之为对应于 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、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2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2个较粗粒度的商空间中是真的,则在一定条件下在其合成的商空间中对应的问题也是真的。这2个原理在商空间模型的推理中起到了很重要的作用。设在2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2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 $1 - \alpha$ 。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,张铃和张铃^[35]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。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和处理信息。他们提出扩展模糊商空间理论的途径,即可从3个方向推广商空间理论成为模糊商空间理论:

- 1) 研究的论域是模糊空间 X ;
- 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^[31,34,89-95],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、难于准确求解的问题,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”、“保假”2个原理,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相加,避免了计算复杂度高的困难,使得看似难于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。

3.4 其他相关粒计算模型

词计算模型、粗糙集模型和商空间模型是3个主要的粒计算模型。在这3个模型的基础上,人们提出了很多新的模型,如基于划分的粒计算模型,基于覆盖的粒计算模型,基于概念格的粒计算模型和基于相容关系的粒计算模型等。

3.4.1 基于划分的粒计算模型

Yao^[17]在讨论了粒计算的基本原理和基本问题的基础上,从语义和算法2个方面研究了粒计算方法中粒子的构建、描述和表达,以及利用粒子进行计算和推理的规则等问题,提出了基于集合论的划分粒计算模型。该模型对一个有限集进行划分得到相应的粒子,这些粒子互不相交,通过子集的包含关系,不同粒度上的粒子之间形成了格的层次结构。他构建了2个算子:Zooming-in和Zooming-out。利用这2个算子,不同粒层之间的粒子可以相互转化。

3.4.2 基于覆盖的粒计算模型

Lin以邻域系统为工具,研究了二元关系下的粒计算模型^[3-10],对粒计算的结构、表示和应用进行了系统的诠释。他研究的粒计算模型是一个典型的覆盖模型。Zhu^[96-97]等人从覆盖约简这个概念出发,讨论了2个覆盖生成相同覆盖广义粗集的判别条件,解决了覆盖的冗余问题,并设计了计算覆盖约简的算法,建立了覆盖下近似运算的公理化体系和上近似运算公理化体系。胡军等人^[98]研究了覆盖粒计算模型的不确定度量。马建敏等人^[99]提出了基于集合论覆盖原理的粒计算模型,该模型是基于一个有限集合上的一个自反二元关系,并利用Zooming-in和Zooming-out 2个算子来实现不同粒层上粒子

的相互转化。

3.4.3 基于容差关系的粒计算模型

粗糙集能有效地分析和处理不精确、不一致和不完整等各种不完备信息,并能从中揭示潜在的规律,近年来在机器学习、数据挖掘等多个领域得到了广泛应用。在对粗糙集理论的研究中,对不完备信息系统的传统处理方法是先进行补齐,然后再用粗糙集的方法来进行处理,由于补齐的过程导致了原始系统信息的变化,所得到的结果不一定反映原始系统的真实情况,因而有必要扩充不完备信息系统的理论与方法,直接从不完备信息系统中获取知识。王国胤等人^[38-39]以容差关系为基础,提出了不完备信息系统的粒计算方法,使用属性值上的容差关系给出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。郑征^[46-47,59]等人根据人类具有依据具体的任务特性把相关数据和知识泛化或者例化成不同程度、不同大小的粒的能力,以及进一步依据这些粒和粒之间的关系进行问题求解的能力提出了相容粒度空间模型。

3.4.4 基于概念格的粒计算模型

概念格,也称形式概念分析,最早由 Wille 于 1982 年提出^[100],它几乎与粗糙集理论一起被提出,并都为数据分析提供了有效的研究方法。它是根据二元关系提出的一种概念层次结构。从数据集中生成概念格的过程,实际上是一种概念聚类的过程,它的每个节点被称为一个概念,概念的外延表示为属于这个概念的所有对象的集合,而内涵则表示为所有对象所共有的属性的集合。概念格在本质上描述了对对象和属性之间的联系,表明了概念之间的泛化和例化关系,它的 Hasse 图则实现了对数据的可视化,作为数据分析和知识处理的形式化工具。概念格理论已被广泛地应用于软件工程、知识工程、数据挖掘、信息检索等领域^[101]。仇国芳^[102-103]等人在概念格的基础上,提出概念知识格,以及概念信息粒格,讨论了概念信息粒之间的蕴含关系,以及由概念信息粒生成的不确定规则的方法,扩充了概念格的研究内容。

张文修等人^[104]认为:知识是人类认知的成果和结晶,包括经验知识和理论知识。经验知识可以理解为从大量现象中归纳的共同特征,数据库知识发现即是这种思维方式的程序化;而理论知识是对事务因果的探索,是对经验知识的理性概括。但任何知识都不是孤立存在的,都存在于相互关联的知识系统中。对于知识的深层次认识有 2 种方法:一是细

分,即知识的分解;二是泛化,即知识的综合。对于确定的知识,没有必要进行泛化和细化,但对于不确定性知识,利用外延内涵算子和内涵外延算子可以将知识进行泛化和例化。结合包含度理论在不同的泛化和例化知识系统上发现知识,这种方法与粒计算的思路完全一致。因此,知识系统的泛化和例化是粒计算的一种特殊形式,它与 Yao^[17]提出的基于划分的粒计算模型(即采用逻辑决策语言(DL-语言)来描述集合的粒(用满足公式 元素的集合来定义等价类 $m(\cdot)$),构建粒世界的逻辑框架,并应用于构建粒网络和分析了相关规则、例外规则和特殊规则)非常相似。曲开社等人^[105]在概念信息粒格上建立了 3 个偏序集:G 偏序集、M 偏序集和 GM 偏序集,并将包含度的概念引入到这 3 个偏序集上,证实了概念信息系统中的内涵、外延和蕴涵规则均可归结为偏序集上的序表示及包含度表示。近几年来,人们进一步对概念知识粒和概念信息粒的转化进行了研究^[102-103],加快了概念格粒计算模型的发展。

随着粒计算研究工作的飞速发展,粒计算模型的种类层出不穷,如基于神经网络的粒计算模型^[49-50]、基于进化计算的粒计算模型^[106]等等。限于篇幅,不再一一列举。

4 粒计算模型之间的关系

基于模糊集合论的词计算模型、基于粗糙集理论的粒计算模型和基于商空间理论的粒计算模型都是描述人类能按不同粒度处理事物的能力的模型。商空间理论模型和粗糙集理论模型认为概念粒子可以用子集来表示,不同粒度下的粒子用不同大小的子集来描述,所有的粒子都通过等价关系获得划分产生。而词计算模型认为表示知识的粒子用一个“词”(即模糊子集)来表示,不同程度的知识可以通过不同的程度词进行刻画。从研究对象上看,3 种模型都是在有限的集合论域上研究,只是各自的侧重点不同。

粗糙集理论的研究对象是由一个多值属性集合描述的一个对象集合,对于每个对象及其属性都有一个值作为其描述符号。对象、属性和描述符是表述决策问题的 3 个基本要素。关于对象可得到的信息不一定足以划分其成员类型,这种不精确性导致了对象的不可分辨性。给定对象间的一个等价关系,即导致由等价类构成的近似空间的不分明关系,粗糙集就用不分明对象类形成的上近似和下近似来描述。它能够解决重要的分类问题,所有冗余对象和属性的约简包含属性的最小子集,能够很好地近似分

类,得到可以接受的分类质量.另外,它还可以用决策规则集合的形式表示最重要属性和特定分类之间的所有重要关系^[75].所以,粗糙集研究的是离散的对象集,针对属性值的差异对对象进行分类,从而形成集合的上、下近似,获得相应的规则,而对象之间没有结构关系或拓扑关系.

商空间理论的着重点不同,它不是只针对给定的商空间(知识基)来讨论知识的表达问题,而是在所有可能的商空间中,找出最合适的商空间,利用从不同商空间(从不同角度)观察同一问题,以便得到对问题不同角度的理解,最终综合成对问题总的理解(解).它的求解过程是在“由所有商空间组成的半序格”中转换的过程,故可看成是宏观的粒计算.而粗糙集理论是在给定的商空间中运动,故可看成是微观的粒计算^[107].商空间理论是讨论不同粒度空间之间的表示、转换和相互依存等问题,与粗糙集理论相同,都是利用等价类来描述“粒子”,用“粒度”来描述概念,但侧重点有所不同.粗糙集理论研究在给定的空间(知识基)上不同概念的表示、转换和相互依存问题,其论域是点集,元素之间没有拓扑关系,故是无拓扑结构情况的商空间特例.商空间理论就是将问题在不同的粒度世界与数学上集合论的商集概念统一起来表示对象模型的方法,即以商集作为不同粒度世界的数学模型的方法.问题的不同粒度表示对应于不同的等价关系,也就是不同的粒度,是对论域进行不同的划分.因此,划分就是构成不同粒度世界的方法,这与粗糙集理论相同.但商空间理论方法提出的商空间描述法与状态空间、问题规约等方法相比具有更强的表达能力.它不仅可以描述论域中的元素、元素之间的不同结构(关系),而且可以定义多种不同的属性函数或运算.它利用拓扑空间和偏序(半序)空间,为不同粒度世界提供了形式描述的手段.商空间模型主要论述的是,当人们面临一个很复杂的问题时,通常从比较“粗”的粒度层次来分析问题,再根据问题的需求粒化问题空间,在不同的商空间上寻求问题的解.常见的思路就是分类讨论的问题.例如二次方程求解,首先将问题粒化为3个商空间:判别式大于0、判别式等于0和判别式小于0.再利用相应的定理(即附加的信息),在相应的商空间上求解,直到问题得到确定解.

粗糙集理论根据属性形成划分.属性一旦确定,划分就确定下来.任何一个核属性都会对划分产生影响,即使影响并不大,也不能忽略它.但在用商空间理论求解实际问题时,粒度的划分可以是动态的,即先进行一次分类,在这个粒度上进行推理与分析,

得到一定的粗略或梗概的性质,根据实际的需求,再进一步分类,直至问题的解决.一般说来,商空间理论中粒度的选择与划分,与问题所属的领域的具体知识密切相关,它依赖一个有序的拓扑空间或半序空间.如当问题空间是半序结构时,通常希望所构造的商空间也是半序空间,但一般情况下商空间未必是半序空间,此时通过忽略一些次要的因素(可能这个因素在粗糙集方法中是核属性),适当地选取粒度,通过商空间的合并法和分解法,使得到的商空间成为半序空间.这样有利于快速寻找到问题的解(可能不是精确解).

基于模糊集的词计算理论与商空间理论、粗糙集理论稍有不同,词计算模型主要讨论粒度的表示问题,即当人类进行各种思考和推理时,都离不开粒度,这些粒度一般就是用语言、词来表示,然后利用模糊逻辑进行词计算的推理和计算.这种方法与人们的主观因素有密切的关系,对于处理复杂的人文系统非常有效.

商空间理论、粗糙集理论是“精确”的粒计算方法,而词计算理论是模糊的粒计算方法.将词计算模型的基本方法(模糊数学)应用到粗糙集模型中就形成了模糊粗糙集和粗糙模糊集理论.模糊集和粗糙集理论在处理不确定性和不精确性问题方面推广了经典集合论,这2个理论的比较和融合是人们感兴趣的话题^[108].模糊粗糙集理论模型的建立和发展,成为粗糙集理论推广的主要方向之一.从Dubois等人^[109]提出模糊粗糙集理论,到后来的各种广义模糊粗糙集理论、公理化的模糊粗糙集理论^[110-111],使该理论的发展达到了一个相对完善的状态.张铃等人^[35]将模糊集理论应用到商空间理论得到模糊商空间理论,利用模糊商空间理论建立了从大量的事物(数据)中获取具有粒度结构(层次)的知识的方法,然后对各商空间提取相应的知识,从而实现了从模糊信息粒结构到分层递阶结构再到具有粒度结构的知识的相互转化.毛军军等人^[112]从商空间理论和信息粒度原理角度出发,引入模糊商空间形成的分层递阶结构,将Fuzzy聚类分析应用于实例.与粗糙集理论相比^[113],商空间理论对分类界线不确定(或是模糊的情况)也可以求解.

虽然商空间理论、词计算理论和粗糙集理论等粒计算模型从解决问题的初衷和解决问题的目标不尽相同,各有特色,但是三者都有一个共同的特点,就是在处理实际复杂问题时,不一定去追求问题的完美解或精确解,而是根据实际需求,得到近似的较优解.如果将三者结合起来,充分发挥它们各自的优

势,将形成更加有效的粒计算方法和理论.目前,模糊粗糙集、粗糙模糊集和模糊商空间对于处理复杂问题都表现出比单一模型更好的优势,得到较好的发展.但如何将粗糙集与商空间有效地结合起来,还没有进一步的研究成果.

基于划分的粒计算模型是粗糙集理论模型的一种特例.它将概念知识分为内涵和外延,在外延上利用集合论上的交并等运算来合成粒子或分解粒子,在内涵上采用属性的多少来控制粒子的大小.基于覆盖的粒计算模型是比基于划分粒计算模型更广的一种粒计算模型,是广义粗糙集理论模型的一种特例.这 2 个模型都是通过某种二元关系(等价关系或非等价关系)产生邻域,再通过 Zooming-in 和 Zooming-out 2 个算子(或其他类似的算子)来实现不同粒层上粒子之间的相互转化.基于容差关系的粒计算模型和基于相容关系的粒计算模型实质是同一个模型,都是基于自反和对称的二元关系,只是各自的侧重点不同.基于概念格的粒计算模型是一种层次结构模型,它的粒子之间通过包含关系形成偏序关系,构成一个完备格.在知识系统中,概念信息粒和概念知识粒之间的相互转化为概念的形成奠定了基础,也为概念知识的获取提供了方法.而结合神经网络形成的神经网络粒计算模型在知识获取方面表现出较大的优势^[49-50].

5 粒计算的应用研究

粒计算方法的应用越来越广泛,已经渗透到自然科学和社会科学的很多领域.这里作一个简单的介绍.

5.1 图形图像处理

图像分割是图像分析和处理技术的重要手段,是模式识别、图像处理、计算机视觉等领域的一个基础环节.一个图像分析系统是否成功,很大程度上依赖于分割的质量^[31,114].随着数学理论特别是应用数学理论的发展,人们借助新的数学理论,对图像分割问题进行研究,并提出了图像分割的许多方法.基于粗糙集的粒计算方法在适当降低精确度为代价的条件下可以解决这一问题. Zadeh 提出了模糊信息粒理论研究的一般框架后, Pedrycz 等人^[115]深入研究了信息粒间的结合问题,并结合数字化图像可粒化的特点,讨论了利用图像的内容信息和空间信息进行粒化的问题.后来, Hirota 和 Pedrycz^[116]用模糊关系来描述一幅静态灰度图像,提出了基于模糊关系计算的图像压缩方法,将模糊信息粒化思想运用到图像压缩领域. Nobuhara 等人^[117]提出了一种求

解模糊关系方程最大解的快速方法,减少了图像重构的运算量. 修保新等人^[118-119]给出了基于模糊信息粒化思想的图像边缘检测方法,并在基于图像模糊粒化思想进行图像插值的基础上,提出了具体的基于图像模糊粒化结构的插值方法. 刘仁金等人^[31,114,120]通过对粒度概念及粗糙集、模糊集等粒计算理论的比较研究以及对图像分割概念分析及对现有的图像分割方法系统分析,提出了图像分割中的商空间粒度原理和基于粒度合成技术原理的纹理图像分割算法,成功设计出纹理图像的分割算法.

5.2 海量数据挖掘

随着计算机及网络的日益普及,丰富的数据与贫乏的知识问题日见突出.不同领域的人们都期待着从这些数据中得到自己的答案,将信息变为知识.数据挖掘就是从大型数据库或数据仓库中储存的大量的、不完整的、有噪声的数据中发现潜在的、有价值的、有趣知识的过程.面对海量数据处理,人们提出了各式各样的挖掘方法.粒计算方法凭借自身的优势在数据挖掘中显示出较大的优势.

基于模糊集的词计算模型是较早提出的粒计算模型,在数据挖掘中有非常成功的应用.模糊集在数据挖掘中应用最多的是聚类分析.聚类分析是通过相似性将用户或数据记录进行分组,从而获得相应的规则.与分类不同,聚类是一种无导师的学习,它将论域中的所有对象划分为簇,同一簇中的对象具有很高的相似性,不同簇中的对象则很不相似,即差异很大^[48].将模糊集引入聚类分析的思想,是首先由 Bellman、Kalaba 和 Zadeh 等人于 1966 年提出的,后来有许多研究者对模糊聚类进行了研究,提出了基于摄动的模糊聚类方法等新的模糊聚类算法.近年来聚类方法层出不穷,文献[48]从信息粒度的角度剖析了聚类和分类技术,试图使用信息粒度原理的框架来统一聚类和分类.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.

粗糙集理论是一种研究不完整、不确定知识和数据的表达、学习、归纳的理论方法.由于它是建立在分类机制基础之上的方法,因而它在数据挖掘分类问题中的应用很多.粗糙集理论在分类问题中的应用大致可以分为 2 类^[121]:一是利用粗糙集理论的属性约简、值约简及核属性,直接从数据表中获取分类规则;二是将粗糙集理论与其他方法结合起来进行分类^[122-124].

商空间理论是研究不同粒度世界的数学工具.在面對数据仓库中海量数据挖掘时,可以针对要处

理的问题,在保证问题求解精度的前提下,选择合适的粒度空间,这样既可以得到所需的处理结果,又可以提高挖掘知识的效率。张昱等人^[91]利用商空间粒度理论中将原问题变成商空间层次上的问题进行描述的方法,引入对数据库和数据仓库不同粒度问题处理,从不同粒度去考察数据库,得出了比较满意的结果。

经典的 Apriori 算法是一种宽度优先算法,通过对数据库的多次扫描来发现所有的频繁项目集。文献[125]提出了一种基于粒计算的关联规则挖掘算法。该算法使用粒的“与”计算来求项目集的支持度,而不用多次遍历数据库,避免了模式的匹配运算,还利用信息粒的“或”计算来求概念层次树上每个节点的信息粒。这样能简单有效地由包含数据项为孩子的支持度得到包含其双亲的支持度。

5.3 复杂问题求解

面对复杂的、难于准确把握的问题,人们通常不是采用系统的、精确的方法去追求问题的最佳解,而是通过逐步尝试的办法达到有限的合理的目标,也就是取得所谓足够满意的解。人类就是采用这种概括的、由粗到细、不断求精的多粒度分析法,避免了计算复杂度高的困难,使得原来看似非多项式难解的问题迎刃而解^[95]。对于复杂问题的描述方法,关键在于不同粒度世界的描述问题。商空间模型给出了描述不同粒度世界的分层递阶方法,再通过合成技术将不同角度、不同层次上得到的信息合成得到原问题的解。近年来,针对实际问题,很多学者将商空间理论进行了推广和应用^[89-91,126-127]。

基于粒计算的应用还有很多,如在模式识别中,因为基于精确数学模型的经典模式识别方法虽然取得了一些成果,但面对复杂的系统,传统的方法遇到了挑战^[63]。实践证明,越是复杂的控制系统、难以建立精确数学模型的智能系统,越难以实现精确定量计算的信息系统,基于粒计算的模式识别方法的作用就越大。在智能控制中,基于模糊集合论的模糊控制理论,在20世纪70年代取得了辉煌的成就,到了80年代,基于粗糙集理论的粗糙控制论得到了快速的发展。近年来,基于模糊商空间的智能控制又得到很好的应用^[92-93]。在医疗诊断系统中,刘清等人^[30,128]依据血液粘稠度信息研制了一个诊断大出血的软件系统(DS-BVS),它是一个基于粒计算的专家推理系统。在人工神经网络中,Zhang等人^[49-50]对粒神经网络进行了探讨,并在快速知识发现中得到很好的应用。在知识获取中,胡军和王国胤等人^[129]利用二进制编码技术来构造粒子,提出了

一种新的属性约简方法。苗夺谦等人^[130]利用粒计算的思想,结合 Skowron 不分明矩阵,给出了粒计算方法在知识约简中的应用模型和实例。在语言动力学系统中,Wang^[37]给出了基于词计算的语言动力学系统的计算理论框架。在生物信息处理中,毛军军等人^[94]利用归一化距离度量和模糊等价关系,结合商空间理论,分析了生物信息序列之间比较的本质问题。

总之,粒计算的应用相当广泛,特别是在处理模糊的、不完整的、不精确的和不确定的海量数据挖掘等方面具有突出的优势,它的应用已经涉及到航空、生物、农业等相关领域,这里不再一一赘述。

6 粒计算研究的展望

粒计算的研究,虽然已经取得一定的成果,且在很多领域得到应用,并取得较好的效果。但是,作为一种正在兴起的智能计算方法,粒计算本身还有许多地方有待发展和完善,对粒计算的研究还存在很多有待进一步研究的问题。这里从以下几个方面进行讨论。

6.1 构建统一的粒计算模型

虽然词计算模型、粗糙集模型和商空间模型之间具有不同程度的联系,但处理实际问题时,人们往往根据问题的实际情况选择其中的某种模型求解。如对于信息表的约简、规则的提取、海量数据信息的挖掘问题,人们可能会选择粗糙集理论模型;对于具有偏序(半序)关系的对象集问题,人们可能会选择商空间理论模型;而对于模糊现象,人们可能会选择词计算模型。但人们很少将同一问题在不同模型下进行考虑。换句话说,如何去寻找解决实际问题的统一粒计算模型,是粒计算研究工作中可能长期存在的问题。这种统一模型的构建将可能深化粒计算的研究,便于人们探索综合利用各种理论模型的优点来解决复杂问题。

6.2 复杂问题空间的粒化

问题空间的粒化是进行粒计算的必要前提。面对具体的实际复杂问题,粒化的程度直接影响着问题求解的计算复杂度和效率。既要避免粒度过粗造成求解失败,也要避免粒度过细造成信息的冗余而导致求解效率低下。因此,对于具体的复杂问题,如何确定出粒化程度的标准是粒计算的关键问题。不同的问题,粒化程度的标准可能无法统一,但面临新的复杂问题,人们可能要尝试多种不同程度的粒化标准,快速寻找到实际问题的最优粒化程度是问题求解的关键。因此,如果能够建立一个求解最优粒化

程度的形式化数学模型,将是一个非常有意义的工作.如从粒化的复杂度、合成的复杂度以及在粒子上求解的复杂度等多方面建立一个求解最优粒化程度的数学模型.该模型的建立将有效地推进粒计算在各类复杂问题求解中的应用.另外,在粒化程度已知时,粒化的方法将直接决定粒化的效率.如粗糙集模型采用等价关系进行粒化,商空间模型利用等价关系建立分层递阶的问题空间模型,而词计算用模糊集得到粒子.尽管在粒计算的各种模型中,一些算法在某些具体实例上获得了成功,但是这些方法缺乏通用性.粒计算的粒化算法设计是一个复杂的问题,迄今为止,还没有一套较成熟的方法.如果能够借鉴现有的相关理论设计出粒计算的粒化算法,逐步形成一套算法体系.如用王国胤提出的自主式知识获取方法^[131-132],利用数据本身的特点来驱动问题空间的粒化,力争实现自主式的粒化过程,从而获得符合实际问题的粒化算法,可能是解决这一问题的有效方法.

6.3 粒层之间的转换

不同粒层次之间的转换是粒计算的一个主要问题.首先,不同粒层之间必须能够自由的转换,往返自如.其次,在能够转换的基础上,确保问题的主要性质不变.例如商空间理论中的“保真”、“保假”原理,是确保商空间理论能够解决复杂问题的关键.模糊集理论模型中截集具有某些性质的不变性.粗糙集理论模型中的核属性是属性约简的不变属性.在基于划分模型的粒计算模型和基于覆盖的粒计算模型中,不同层次之间通过 Zooming-in 和 Zooming-out 2 个算子来实现粒层之间的转换.但在这些粒计算模型中,粒层之间的转化方法各不相同,且在转换过程中,如何确保主要性质不变,以及属性不变的程度等问题没有一个统一的标准.当然,不同复杂问题,其方法不同,主要属性不变的要求程度也可能不同.但如果能够建立一个用于计算不同模型的“保性”程度的形式化数学模型,将有助于粒计算模型不同粒层之间的相互转化.虽然张钹和张铃从定性的角度提出了商空间理论中的“保真”和“保假”原理以及“弱保真”和“弱保假”原理,但如何从定量的角度去分析不同粒层之间相互转换后重要属性的不变性的研究工作,目前还没有研究人员对之进行讨论.如果能够从定量的角度去研究问题在不同粒层上表现出来的“保真”和“保假”程度,将有利于实现不同粒层之间的相互转化,有助于提高用粒计算方法求解复杂问题时的收敛速度.

6.4 高效的粒计算方法

粒计算方法的初衷就是求解复杂问题.而用传统的方法来解决实际的复杂问题,不管是时间复杂度,还是空间复杂度,往往都非常高.粒计算方法要得到广泛的推广和应用,其复杂度应该低于现有的传统方法.虽然目前的粒计算方法在计算效率上较传统方法有一定的优势,但它们处理问题的时间复杂度都大于或等于 $O(n^2)$.因此,研究高效粒计算方法是一个重要的课题.如果能获得解决问题的时间复杂度为 $O(n^{1.5})$ 或 $O(n \lg n)$,甚至更低的粒计算算法,将有助于提高用粒计算方法解决复杂问题的效率.

6.5 新的粒计算模型

词计算模型、粗糙集理论模型和商空间理论模型在处理复杂问题时表现出不同的优势,也表现出某些方面的不足.例如,粗糙集理论的优势在于对知识的表示,而商空间理论的特色在于对问题结构的讨论,而形式概念分析具有良好的层次结构等特色.如何将不同的粒计算模型有机地结合起来,发挥各自的优势,弥补各自的缺陷,从而寻求一个更有效的粒计算模型,是目前粒计算工作中的一个重要研究课题.张铃和张钹将商空间理论和模糊集理论结合起来得到的模糊商空间理论表现出良好的性能,它能够更好地反映人类处理不确定问题的若干特点^[135].将模糊集理论和粗糙集理论结合形成模糊粗糙集理论和粗糙模糊集理论的文献很多,并在处理不确定信息时表现出更好的性能^[108-111].但关于如何将粗糙集理论与商空间模型有机地融合起来,以及如何将粗糙集理论和形式概念分析 2 种智能数据分析方法融合起来,形成更有效的粒计算模型的研究工作很少.另外,关于论域空间上代数运算的讨论,在粗糙集理论和商空间理论中都不是很多,如果能将代数中的商代数理论移植到商空间理论中,问题求解的商空间理论的内涵将得到丰富.总之,发现更有效的粒计算模型将是研究者追求的目标.

6.6 动态粒计算模型

目前,绝大多数粒计算模型都是基于静态数据(信息)的处理.而很多实际问题的信息都是随着时间的变化而动态变化的.如网络数据、入侵检测数据和卫星检测数据等都是动态数据.如何构建基于动态数据处理的粒计算模型是一个具有广泛应用前景的课题.

6.7 自主式粒计算模型

在没有领域先验知识条件下的不确定知识的主动式学习是机器学习领域中新兴的热点问题.对于不确定性问题的处理,现在还没有一个能够摆脱先

验知识,完全由原始数据自主决策的理论方法.如果能够摆脱学习过程中对先验知识的依赖,由数据自主地完成知识的获取过程,无疑将对机器学习理论的推广起到重要的推动作用.王国胤等人^[131-132]通过研究决策表和决策规则的不确定性,建立基于粗糙集表示、度量和处理不确定性信息和知识的理论,并且结合 Skowron 缺省规则获取算法,提出一种不确定性条件下的数据自主式学习模型和方法,并通过仿真实验,验证了该自主式学习方法的有效性.这个研究作为知识的获取提供了一个新的研究方向.而面对复杂问题的求解,多数粒计算模型借助先验知识对问题空间进行粒化,如模糊集理论往往借助模糊关系的截集产生粒子(依赖隶属函数),而商空间理论和粗糙集理论模型借助等价关系产生粒子,但在不完备信息系统中这些粒计算模型无法借用等价关系,需要人为地增加辅助信息(如去掉等价关系的条件).这些模型从某种意义上讲都不是自主式的求解模型.因此,建立如何根据问题已有的信息,而不依靠先验知识获得问题解的粒计算模型具有非常重要的价值,可为真正的智能化数据挖掘提供理论依据.

6.8 粒计算方法的模糊化

在自然科学或社会科学研究中,存在着许多定义不很严格或者说具有模糊性的概念.各门学科,尤其是人文、社会学及其他“软科学”无法用精确的数学描述.随着计算机、控制论、系统科学的迅速发展,要使计算机能像人脑那样对复杂事物具有识别能力,就必须研究和处理模糊性问题.从现有的集合论发展看来,精确的康托集不能描述的问题已经逐渐转化为由 Zadeh 提出的模糊集来描述,而康托集是模糊集的特殊情况.因此,原有的基于康托集的很多理论可以推广到模糊集理论,而原来用康托集无法解决的问题,用模糊集可能会得到解决.因此,基于经典集合论的粒计算方法,通过一定的转化可以推广为基于模糊集的粒计算方法.如模糊商空间理论、模糊粗糙集理论、模糊区间估计理论都是典型的模糊化的粒计算方法,而模糊概念格理论、模糊 D-S 证据理论和模糊神经网络等则是近年来人们提出的新的粒计算模型.由于粒计算方法的初衷就是解决复杂问题,而实际的复杂问题往往无法用精确的数据来刻画.比如,研究一个通过电话听筒声音辨别来电是谁这样一个简单的语音识别系统,就无法用精确数学刻画,必须借助模糊数学.尽管普通的精确方法(如数学、化学等)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在某些人文系统中的应用出现异常和失败.为了求解巨大复杂的人文学科系统问题,如病虫测报、品种选育、图像识别、天气预报、地质地震、交通运输、医疗诊断、信息控制、人工智能等诸多领域,研究模糊化的粒计算方法是必然的趋势.因此,粒计算方法的模糊化是一个极具潜力的研究方向.通过对它的研究,可以清楚地揭示不同粒计算模型内在的本质特征.

6.9 粒计算模型应用的推广

粗糙集、模糊集和商空间理论各自的应用比较多见,但基于上述理论的粒计算模型的很多文章对粒计算的研究还停留在理论上,没有结合具体问题进行讨论.因此,一个当务之急的研究课题就是结合具体的问题来分析、讨论和应用粒计算模型的重要结论.图像处理、小波分析等是粒计算模型得以应用的很好领域.因为一幅图像的所有像素之间存在自然的结构,可以用拓扑或者闭包运算来描述这种结构,进而利用商空间理论解决数字图像的相关问题.虽然现在很多粒计算模型都应用在图像处理领域,但这方面的应用研究还应该进一步深入.张钹和张铃利用分层递阶的商空间结构提出网络推理模型,并把确定性推理和不确定性推理统一起来,成功应用在复杂问题的推理中.另外,商空间理论也在运动规划、时间规划和统计启发式搜索方面得到较广泛的应用.如何将现有的粒计算模型进行推广应用,逐渐形成求解复杂问题的有效粒计算模型方法,将有助于人们对复杂问题的求解.

粒计算方法的研究是近几年人工智能领域中崛起的一个新方向,它用可行的满意近似解替代精确解,改变了传统的计算观念,其主要思想是在不同的粒度层次上进行问题求解,在很大程度上体现了人类问题求解过程中的智能.虽然粒计算方法的应用已经逐渐得到人们的认可,但很多环节需要进一步的探讨和研究,如粒化标准、粒子的表述、合成的复杂度以及高效的实现等问题,这些工作都有待进一步解决.

7 结束语

粒计算是人工智能领域中的一种新理念和新方法,它覆盖了所有和粒度相关的理论、方法、技术和工具,主要用于对不确定、不精确、不完整信息的处理,对大规模海量数据的挖掘以及对复杂问题的求解.粒计算最本质的思想是通过合适粒度的选择,来寻找问题的一种较好的、近似的解决方案,从而降低

问题求解的复杂度. 本文首先简要回顾了粒计算的基本发展情况,着重介绍了粒计算理论的几个基本组成和基本问题,论述了粒计算的基本模型和方法,并探讨了这些模型之间的差异和相互关系,最后提出了粒计算的进一步研究方向和问题. 希望这些工作能够对粒计算的研究和推广提供借鉴和帮助.

参考文献:

- [1] ZADEH L A. Fuzzy sets and information granulation. advances in fuzzy set theory and applications[M]. Amsterdam: North-Holland Publishing, 1979.
- [2] HOBBS J R. Granularity[A]. Proceedings of the Ninth International Joint Conference on Artificial Intelligence [C]. Los Angeles, CA, 1985.
- [3] LIN T Y. Neighborhood systems and relational database [A]. Proceedings of CSC '88[C]. New York, 1988.
- [4] LIN T Y. Granular computing on binary relations I: data mining and neighborhood systems, II: rough set representations and belief functions, rough sets in knowledge discovery[M]. Physica-Verlag, 1998: 107 - 140.
- [5] LIN T Y. Data mining: granular computing approach methodologies for knowledge discovery and data mining [A]. Proceedings of PAKDD '99[C]. [S.l.], 1999.
- [6] LIN T Y. Granular computing: fuzzy logic and rough sets. Computing with Words in Information/ Intelligent Systems[M]. Physica-Verlag, 1999.
- [7] LIN T Y. Data mining and machine oriented modeling: a granular computing approach[J]. Journal of Applied Intelligence, 2000, 13(2): 113 - 124.
- [8] LIN T Y. Granular computing: structures, representations, applications and future directions [A]. The Proceedings of 9th International Conference, RSFDGrC 2003 [C]. [S.l.], 2003.
- [9] LIN T Y. Granular computing rough set perspective[J]. The Newsletter of the IEEE Computational Intelligence Society, 2005, 2(4): 1543 - 4281.
- [10] LIN T Y. Granular computing: a problem solving paradigm[A]. The Proceedings of the 2005 IEEE International Conference on Fuzzy Systems[C]. [S.l.], 2005.
- [11] 张钹, 张铃. 问题求解理论及应用[M]. 北京: 清华大学出版社, 1990.
- [12] ZHANG B, ZHANG L. Theory and applications of problem solving[M]. Amsterdam: North-Holland Publishing, 1992.
- [13] ZADEH L A. Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic[J]. Fuzzy Sets and Systems, 1997, 19: 111 - 127.
- [14] YAO Y Y. Relational interpretations of neighborhood operators and rough set approximation operators[J]. Information Sciences, 1998, 111: 239 - 259.
- [15] YAO Y Y. Rough sets, neighborhood systems, and granular computing[A]. Proceedings of the 1999 IEEE Canadian Conference on Electrical and Computer Engineering[C]. [S.l.], 1999.
- [16] YAO Y Y. Granular computing using neighborhood systems, advances in soft computing: engineering design and manufacturing[A]. The 3rd On-line World Conference on Soft Computing (WSC3) [C]. London, 1999.
- [17] YAO Y Y. A partition model of granular computing [J]. LNCS Transactions on Rough Sets, 2004(1): 232 - 253.
- [18] YAO Y Y. Stratified rough sets and granular computing [A]. Proceedings of the 18th International Conference of the North American Fuzzy Information Processing Society[C]. New York, 1999.
- [19] YAO Y Y. Information granulation and rough set approximation[J]. International Journal of Intelligent Systems, 2001, 16(1): 87 - 104.
- [20] YAO Y Y. Granular computing for data mining [A]. Proceedings of SPIE Conference on Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security[C]. Kissimmee, USA, 2006.
- [21] CHEN Y H, YAO Y Y. Multiview intelligent data analysis based on granular computing[A]. Proceedings of 2006 IEEE International Conference on Granular Computing[C]. [S.l.], 2006.
- [22] YAO Y Y. Three perspectives of granular computing [J]. Journal of Nanchang Institute of Technology, 2006, 25(2): 16 - 21.
- [23] SKOWRON A. Toward intelligent systems: calculi of information granules[J]. Bulletin of International Rough Set Society, 2001(5): 9 - 30.
- [24] SKOWRON A, STEPANIUK J. Towards discovery of information granules[A]. Proceedings of PKDD '99[C]. [S.l.], 1999.
- [25] SKOWRON A, STEPANIUK J. Information granules: towards foundations of granular computing[J]. International Journal of Intelligent Systems, 2001, 16: 57 - 85.
- [26] POLKOWSKI L, SKOWRON A. Towards adaptive calculus of granules[A]. Proceedings of 1998 IEEE International Conference on Fuzzy Systems[C]. [S.l.], 1998.
- [27] POLKOWSKI L, SKOWRON A. Rough sets in knowledge discovery[M]. Heidelberg: Physica-Verlag, 1998.
- [28] 刘清, 黄兆华. G逻辑及其归结推理[J]. 计算机学报, 2004, 27(7): 865 - 872.
- LIU Qing, HUANG Zhaohua. G logic and its resolution reasoning[J]. Chinese Journal of Computers, 2004,

- 27(7):865 - 872.
- [29]刘清,刘群.粒及粒计算在逻辑推理中的应用[J].计算机研究与发展,2004,41(4):546 - 551.
- LIU Qing, LIU Qun. Granules and applications of granular computing in logical reasoning[J]. Journal of Computer Research and Development, 2004, 41(4):546 - 551.
- [30]刘清. Rough集及 Rough推理[M]. 北京:科学出版社,2001.
- [31]刘仁金,黄贤武. 图像分割的商空间粒度原理[J]. 计算机学报,2005,28(10):1680 - 1685.
- LIU Renjin, HUANG Xianwu. The granular theorem of quotient space in image segmentation[J]. Chinese Journal of Computers, 2005, 28(10):1680 - 1685.
- [32]王国胤. Rough集理论在不完备信息系统中的扩充[J]. 计算机研究与发展,2002(10):11 - 15.
- WANG Guoyin. Extension of rough set under incomplete information systems[J]. Journal of Computer Research and Development, 2002(10):11 - 15.
- [33]伞冶,叶玉玲.粗糙集理论机器在智能系统中的应用[J]. 智能系统学报,2007,2(4):40 - 47.
- SAN Ye, YE Yuling. Rough sets theory and its application in the intelligent systems[J]. CAAI Transactions on Intelligent Systems, 2007, 2(4):40 - 47.
- [34]ZHANG L, ZHANG B. A quotient space approximation model of multi-resolution signal analysis[J]. J Comput Sci & Technol, 2005, 20(1):90 - 94.
- [35]张铃,张钹. 模糊商空间理论(模糊粒度计算方法)[J]. 软件学报,2003,14(4):770 - 776.
- ZHANG Ling, ZHANG Bo. Theory of fuzzy quotient space (methods of fuzzy granular computing) [J]. Journal of Software, 2003, 14(4):770 - 776.
- [36]苗夺谦,范世栋. 知识的粒度计算及其应用[J]. 系统工程理论与实践,2002,22(1):48 - 56.
- MIAO Duoqian, FAN Shidong. The calculation of knowledge granulation and its application[J]. Systems Engineering-Theory & Practice, 2002, 22(1):48 - 56.
- [37]王飞跃. 词计算和语言动力学系统的计算理论框架[J]. 模式识别与人工智能,2001,14(4):377 - 384.
- WANG Feiyue. Computing with words and a framework for computational linguistic dynamic systems[J]. Pattern Recognition and Artificial Intelligence, 2001, 14(4):377 - 384.
- [38]WANG G Y, HU F, HUANG H, et al. A granular computing model based on tolerance relation[J]. The Journal of China Universities of Posts and Telecommunications, 2005, 12(3):86 - 90.
- [39]胡峰,黄海,王国胤,等. 不完备信息系统的粒计算方法[J]. 小型微型计算机系统,2005,26(8):1335 - 1339.
- HU Feng, HUANG Hai, WANG Guoyin, et al. Granular computing in incomplete information systems[J]. Mini-micro Systems, 2005, 26(8):1335 - 1339.
- [40]邓蔚,王国胤,吴渝. 粒计算综述[J]. 计算机科学,2004,31(10):178 - 181.
- DENG Wei, WANG Guoyin, WU Yu. A survey of granular computing[J]. Computer Science, 2004, 31(10):178 - 181.
- [41]邓蔚,王国胤,吴渝. 粒计算对进化计算的一个改进[J]. 计算机科学,2004,31(10):209 - 211.
- DENG Wei, WANG Guoyin, WU Yu. An improved algorithm of evolutionary computation based on granular computing[J]. Computer Science, 2004, 31(10):209 - 211.
- [42]安久江,甘全,王国胤,等. 一种基于粒计算的规则知识获取算法[J]. 计算机科学,2005,32(8):137 - 140.
- AN Jiujiang, GAN Quan, WANG Guoyin, et al. An Algorithm for Rule and Knowledge Acquisition Based on Granular computing, Computer Science, 2005, 32(8A):137 - 140.
- [43]邓蔚,王国胤,吴渝. 基于近似分割的粒计算规则获取算法[A]. 中国人工智能学会第11届学术年会论文集[C]. 武汉,2005.
- DENG Wei, WANG Guoyin, WU Yu. A rule acquisition algorithm of granular computing based on approximate partition[A]. Proceedings of 2005 National Conference on Artificial Intelligence[C]. Wuhan, 2005.
- [44]AN J J, WANG G Y, WU Y, et al. A rule generation algorithm based on granular computing[A]. 2005 IEEE International Conference on Granular Computing[C]. Beijing, 2005.
- [45]ZHENG Z, HU H, SHI Z Z. Granulation based image texture recognition[J]. Lecture Notes in Computer Science, 2004, 3066:659 - 664.
- [46]ZHENG Z, HU H, SHI Z Z. Tolerance relation based information granular space[J]. Lecture Notes in Computer Science, 2005, 3641:682 - 691.
- [47]ZHENG Z, HU H, SHI Z Z. Tolerance granular space and its applications[A]. IEEE International Conference on Granular Computing[C]. Beijing, 2005.
- [48]卜东波,白硕,李国杰. 聚类/分类中的粒度原理[J]. 计算机学报,2002,25(8):810 - 816.
- BU Dongbo, BAI Shuo, LI Guojie. Principle of granularity in clustering and classification[J]. Chinese Journal of Computers, 2002, 25(8):810 - 816.
- [49]ZHANG Y Q, FRASER M D, GALIANO R A. Granular neural networks for numerical-linguistic data fusion and knowledge discovery[J]. IEEE Transactions on Neural Networks, 2000(11):658 - 667.
- [50]ZHANG Y Q. Constructive granular systems with uni-

- versal approximation and fast knowledge discovery[J]. IEEE Trans on Fuzzy Systems, 2005, 13(1): 48 - 57.
- [51] 李道国, 苗夺谦, 杜伟林. 粒度计算在人工神经网络中的应用[J]. 同济大学学报(自然科学版), 2006, 34(7): 960 - 964.
- LI Daoguo, MIAO Duoqian, DU Weilin. Application of granular computing to artificial neural network [J]. Journal of Tongji University(Natural Science), 2006, 34(7): 960 - 964.
- [52] 齐红. 基于形式概念分析的知识发现方法研究[D]. 吉林: 吉林大学, 2005.
- QI Hong. Knowledge discovery methods research based on formal concept analysis[D]. Jinlin: Jinlin University, 2005.
- [53] 杜伟林, 苗夺谦, 李道国, 等. 概念格与粒度划分的相关性分析[J]. 计算机科学, 2005, 32(1): 182 - 187.
- DU Weilin, MIAO Duoqian, LI Daoguo, et al. Analysis on relationship between concept lattice and granular partition lattice[J]. Computer Science, 2005, 32(1): 182 - 187.
- [54] YAGER R R. Some learning paradigms for granular computing[A]. 2006 IEEE International Conference on GrC[C]. Atlanta, USA, 2006.
- [55] LIN T Y. Granular computing 2: infrastructures for AI engineering[A]. 2006 IEEE International Conference on GrC[C]. Atlanta, USA, 2006.
- [56] BARGIELA A, PEDRYCZ W. The roots of granular computing[A]. Proceedings of the 2006 IEEE International Conference on Granular Computing (GrC '06) [C]. Atlanta, USA, 2006.
- [57] YAGER R R, FILEV D. Operations for granular computing: mixing words and numbers[A]. IEEE International Conference on Fuzzy Systems[C]. [S.l.], 1998.
- [58] 陈万里. 基于商空间理论和粗糙集理论的粒计算模型研究[D]. 安徽: 安徽大学, 2005.
- CHEN Wanli. Research on models of granular computing based on quotient space theory and rough set theory [D]. Anhui: Anhui University, 2005.
- [59] 郑征. 相容粒度空间模型及其应用研究[D]. 北京: 中国科学院研究生院, 2006.
- ZHENG Zheng. Tolerance granular space and its applications[D]. Beijing: Graduate School of the Chinese Academy of Sciences, 2005.
- [60] 仇国芳, 陈劲. 概念知识系统与概念信息粒格[J]. 工程数学学报, 2005, 22(6): 963 - 969.
- QIU Guofang, CHEN Jin. Concept knowledge system and concept information granular lattice [J]. Chinese Journal of Engineering Mathematics, 2005, 22(6): 963 - 969.
- [61] 阮达, 黄崇福. 模糊集与模糊信息粒理论[M]. 北京: 北京师范大学出版社, 2000.
- [62] ZADEH L A. Fuzzy logic = computing with words[J]. IEEE Trans on Fuzzy Systems, 1996(4): 103 - 111.
- [63] 李道国, 苗夺谦, 张东星, 等. 粒度计算研究综述[J]. 计算机科学, 2005, 32(9): 1 - 12.
- LI Daoguo, MIAO Duoqian, ZHANG Dongxing, et al. An overview of granular computing[J]. Computer Science, 2005, 32(9): 1 - 12.
- [64] ZADEH L A. Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic[J]. Fuzzy Sets and Systems, 1997, 19: 111 - 127.
- [65] ZADEH L A. Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems [J]. Soft Computing, 1998(2): 23 - 25.
- [66] WANG P. Computing with words[M]. New York: John Wiley & Sons, 2001.
- [67] 李征, 邵世煌. 基于信息粒化、语词计算模糊控制系统中的信息重组[J]. 东华大学学报, 2000, 2(3): 9 - 13.
- LI Zheng, SHAO Shihuang. Information retrieval in fuzzy control system based on information granulation and computing with words[J]. Journal of Donghua University, 2000, 2(3): 9 - 13.
- [68] 李征, 王维工. 基于信息粒化和语词计算的模糊控制器[J]. 辽宁工程技术大学学报(自然科学版), 2001, 20(5): 636 - 639.
- LI Zheng, WANG Weigong. Fuzzy controller based on information granulation and computing with words[J]. Journal of Liaoning Technical University (Natural Science Edition), 2001, 20(5): 636 - 639.
- [69] WANG F Y, TONG L Y. Linguistic dynamic systems and computing with words for complex systems [A]. Proceedings of IEEE International Conference on Systems, Man, and Cybernetics [C]. Nashville, USA, 2000.
- [70] WANG F Y, LIN Y T. Linguistic dynamic systems and computing with words for modeling and simulation of complex systems, in discrete event modeling and simulation technologies: a tapestry of systems and AI based theories and methodologies[M]. [S.l.], 1998.
- [71] WANG F Y. Computing with words and linguistic dynamic systems [M]. World Scientific Publishing Co, 2004.
- [72] 王飞跃. 词计算和语言动力学系统的基本问题 and 研究[J]. 自动化学报, 2005, 31(6): 844 - 852.
- WANG Feiyue. Fundamental issues in research of computing with words and linguistic dynamic systems[J]. Acta Automatica Sinica, 2005, 31(6): 844 - 852.
- [73] PAWLAK Z. Rough sets[J]. International Journal of

- Computer and Information Sciences, 1982(11):341 - 356.
- [74] PAWLAK Z. Rough sets: Theoretical Aspects of Reasoning about Data[M]. [S. l.] Kluwer Academic Publishers, Dordrecht, 1991.
- [75] 王国胤. Rough 集理论与知识获取[M]. 西安:西安交通大学出版社, 2001.
- [76] KRYSZKEIWICZ M. Rough set approach to incomplete information systems[J]. Information Science, 1998, 112: 39 - 49.
- [77] STEFANOWSKI J, TSOUKIAS A. On the extension of rough sets under incomplete information[A]. Proc of the 7th International Workshop on New Directions in Rough Sets, Data Mining, and Granular Soft Computing[C]. Berlin, 1999.
- [78] 王国胤. Rough 集理论在不完备信息系统中的扩充[J]. 计算机研究与发展, 2002, 39(10): 1238 - 1243.
- WANG Guoyin. Extension of rough set under incomplete information systems[J]. Journal of Computer Research and Development, 2002, 39(10): 1238 - 1243.
- [79] ZHANG Q H, WANG G Y, HU J, et al. Incomplete information systems processing based on fuzzy-clustering[A]. 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology [C]. Hong Kong, China, 2006.
- [80] 张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法[M]. 北京: 科学出版社, 2001.
- [81] 黄兆华, 邓毅雄. 粒计算及其应用的研究[J]. 华东交通大学学报, 2005, 22(5): 124 - 128.
- HUANG Zhao-hua, DENG Yi-xiong. An approach for granular computing and its application[J]. Journal of East China Jiaotong University, 2005, 22(5): 124 - 128.
- [82] PAWLAK Z. Granularity of knowledge, indiscernibility and rough sets [A]. Proceedings of 1998 IEEE International Conference on Fuzzy Systems [C]. [S. l.], 1998.
- [83] POLKOWSKI L, SKOWRON A. Rough-neuro computing[A]. Proceedings of the Second International Conference on Rough Sets and Current Trends in Computing [C]. Berlin: Springer-Verlag, 2005.
- [84] SKOWRON A. Toward intelligent systems: calculi of information granules [A]. Proceedings of International Workshop on Rough Set Theory and Granular Computing (RSTGC '2001) [C]. Matsue, Shimane, Japan, 2001.
- [85] PETERS J F, SKOWRON A, SURAJ Z, et al. Measures of inclusion and closeness of information granules: a rough set approach[A]. Proceedings of the Third International Conference on RSCTC '2002 [C]. Malvern, USA: Springer, 2002.
- [86] PETERS J F, SZCZUKA M S. Rough neurocomputing: a survey of basic models of neurocomputation[A]. Proceedings of the Third International Conference on RSCTC '2002 [C]. Malvern, USA: Springer, 2002.
- [87] LIN T Y. Granular fuzzy sets: a view from rough set and probability theories [J]. International Journal of Fuzzy Systems, 2001, 3(2): 373 - 381.
- [88] YAO J T, YAO Y Y. Induction of classification rules by granular computing[A]. Proceedings of the Third International Conference on RSCTC '2002 [C]. Malvern, USA: Springer, 2002.
- [89] 徐峰, 张铃, 王伦文. 基于商空间理论的模糊粒度计算方法[J]. 模式识别与人工智能, 2004, 17(4): 424 - 429.
- XU Feng, ZHANG Ling, WANG Lunwen. The approach of the fuzzy granular computing based on the theory of quotient space [J]. Pattern Recognition and Artificial Intelligence, 2004, 17(4): 424 - 429.
- [90] 徐峰, 张铃. 基于商空间的非均匀粒度聚类分析[J]. 计算机工程, 2005, 31(3): 26 - 28.
- XU Feng, ZHANG Ling. An analysis of uneven granules clustering based on quotient space [J]. Computer Engineering, 2005, 31(3): 26 - 28.
- [91] 张旻, 吴涛, 王伦文, 等. 商空间粒度计算理论在数据库和数据仓库中应用[J]. 计算机工程与应用, 2003, 39(17): 47 - 49.
- ZHANG Min, WU Tao, WANG Lunwen, et al. The application of granularity of the quotient space theory in database and data warehouse [J]. Computer Engineering and Applications, 2003, 39(17): 47 - 49.
- [92] 张持健, 李畅, 张铃. 商空间理论(粒度计算方法)实现高精度模糊控制[J]. 计算机工程与应用, 2004, 40(11): 37 - 39.
- ZHANG Chijian, LI Chang, ZHANG Ling. Realizing the high-precision fuzzy control based on the theory of quotient space (methods of granular computing) [J]. Computer Engineering and Applications, 2004, 40(11): 37 - 39.
- [93] 刘岩, 李友一, 陈占军, 等. 基于商空间理论的模糊控制在航空相机中的应用[J]. 南京航空航天大学学报, 2006, 38(7): 88 - 90.
- LIU Yan, LI Youyi, CHEN Zhanjun, et al. Application of fuzzy controller based on quotient space theory to aerial camera [J]. Journal of Nanjing University of Aeronautics & Astronautics, 2006, 38(7): 88 - 90.
- [94] 毛军军, 郑婷婷, 张铃. 基于商空间理论的生物序列比较模型[J]. 计算机工程与应用, 2004, 40(34): 15 - 17.
- MAO Junjun, ZHENG Tingting, ZHANG Ling. Biological sequence alignments based on quotient space [J]. Computer Engineering and Applications, 2004, 40(34):

- 15 - 17.
- [95]张燕平,张 铃,吴 涛.不同粒度世界的描述法——商空间法[J].计算机学报,2004,27(3):328-333.
ZHANG Yanping, ZHANG Ling, WU Tao. The representation of different granular worlds:a quotient space [J]. Chinese Journal of Computers, 2004, 27(3):328-333.
- [96]ZHU W, WANG F Y. Reduction and axiomization of Covering generalized rough sets [J]. Information Sciences, 2003,152, 217 - 230.
- [97]祝 峰,何华灿.粗集的公理化[J].计算机学报,2000,23(3):330-333.
ZHU Feng, HE Huacan. The axiomatization of the rough set [J]. Chinese Journal of computers, 2000,23(3):330-333.
- [98]HU J, WANG G Y, ZHANG Q H. Uncertainty measure of covering generated rough set [A]. 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology [C]. Hong Kong,2006.
- [99]MA J M, ZHANG W X, LI T J. A Covering model of granular computing [A]. Proceeding of the Fourth International Conference on Machine Learning and Cybernetics[C]. Guangzhou, China,2005.
- [100]WILLE R. Restructuring lattice theory: an approach based on hierarchies of concepts [M]. Reidel: Dordrecht-Boston, 1982.
- [101]胡可云,陆玉昌,石纯一.概念格及其应用进展[J].清华大学学报(自然科学版),2000,40(9):77-81.
HU Keyun, LU Yuchang, SHI Chunyi. Advances in concept lattice and its application[J]. Journal of Tsinghua University (Science and Technology), 2000, 40(9):77-81.
- [102]仇国芳,陈 劲.概念知识系统与概念信息粒格[J].工程数学学报,2005,22(6):963-969.
QIU Guofang, CHEN Jin. Concept knowledge system and concept information granular lattice [J]. Chinese Journal of Engineering Mathematics, 2005,22(6):963-969.
- [103]仇国芳,荆彦玲.粗糙集与概念格[M].西安:西安交通大学出版社,2006.
- [104]张文修,仇国芳.基于粗糙集的不确定性决策[M].北京:清华大学出版社,2005.
- [105]曲开社,翟岩慧.偏序集、包含度与形式概念分析[J].计算机学报,2006,29(2):219-226.
QU Kaishe, ZHAI Yanhui. Posets, inclusion degree theory and FCA [J]. Chinese Journal of Computers, 2006,29(2):219-226.
- [106]蒙祖强,蔡自兴.一种新的计算方法:粒度进化计算[J].计算机工程与应用,2006,42(1):5-8.
MENG Zuqiang, CAI Zixing. A new computing:granular evolutionary computing[J]. Computer Engineering and Applications, 2006,42(1):5-8.
- [107]张 铃.浅谈粒度计算[EB/OL].http://www.ahcit.com,2006-06-21.
- [108]黄正华,胡宝清.模糊粗糙集理论研究进展[J].模糊系统与数学,2005,19(4):125-134.
HUANG Zhenghua, HU Baoqing. The studies of fuzzy rough sets theory:a survey[J]. Fuzzy Systems and Mathematics, 2005,19(4):125-134.
- [109]DUBOIS D, PRADE H. Rough fuzzy sets and fuzzy rough sets[J]. International Journal of General Systems, 1990, 17:91-109.
- [110]GRECO S, MATARAZZO B, SLOWIFISKI R. Fuzzy similarity relation as a basis for rough approximations[A]. RSCTC '98[C]. Heidelberg: Springer-Verlag, 1998.
- [111]RADZIKOWSKA A M, KERRE E E. A comparative study of fuzzy rough sets[J]. Fuzzy Sets and Systems, 2002, 126:137-155.
- [112]毛军军,张 铃,许义生.基于商空间和信息粒度的Fuzzy聚类分析[J].运筹与管理,2004,13(4):25-29.
MAO Junjun, ZHANG Ling, XU Yisheng. Fuzzy clustering analysis based on quotient space and information granularity [J]. Operations Research and Management Science, 2004,13(4):25-29.
- [113]张燕平,张 铃,夏 莹.商空间理论与粗糙集的比较[J].微机发展,2004,14(10):21-24.
ZHANG Yanping, ZHANG Ling, XIA Ying. To compare the theory of quotient space with rough set [J]. Microcomputer Development, 2004,14(10):21-24.
- [114]刘仁金,黄贤武,孟 静,等.基于商空间的纹理图像分割[J].计算机应用,2004,24(7):37-39.
LIU Renjin, HUANG Xianwu, MENG Jing, et al. Texture image segmentation based on quotient space [J]. Computer Applications, 2004,24(7):37-39.
- [115]PEDRYCZ W, SMITH M H, BARGIELA A. A granular signature of data [A]. Int Conf NAFIPS '2000[C]. Atlanta, USA,2000.
- [116]HIROTA H, PEDRY W. Fuzzy relational compression [J]. IEEE Trans Syst, Man, Cybern, 1999, 29(3):407-415.
- [117]NOBUHARA H, PEDRYEZ W, HIROTA K. Fast solving method of fuzzy relational equation and its application to image compression/reconstruction[J]. IEEE Trans Fuzzy Syst, 2000,8(3):325-334.
- [118]修保新,吴孟达.图像模糊信息粒的适应性度量及其在边缘检测中的应用[J].电子学报,2004,32(2):274-277.
XIU Baoxin, WU Mengda. Adaptability measure to

- fuzzy information granule on image and its application to edge-detection[J]. Acta Electronica Sinica, 2004, 32(2): 274 - 277.
- [119] 修保新, 任双桥, 张维明. 基于模糊信息粒化理论的图像插值方法[J]. 国防科技大学学报, 2004, 26(3): 34 - 38.
- XIU Baoxin, REN Shuangqiao, ZHANG Weiming. An approach to image interpolation based on the theory of fuzzy information granulation[J]. Journal of National University of Defense Technology, 2004, 26(3): 34 - 38.
- [120] 刘仁金. 基于商空间的纹理图像分割研究[D]. 合肥: 安徽大学, 2005.
- LIU Renjin. Study of texture image segmentation based on the quotient space[D]. Hefei: Anhui University, 2005.
- [121] 张丽娟, 李舟军, 陈火旺. 粒度计算及其在数据挖掘中的应用[J]. 计算机科学, 2005, 32(12): 178 - 180.
- ZHANG Lijuan, LI Zhoujun, CHEN Huowang. Granular computing and its applications in data mining[J]. Computer Science, 2005, 32(12): 178 - 180.
- [122] WEI J M. Rough set based approach to selection of node[J]. International Journal of Computational Cognition, 2003, 1(2): 25 - 40.
- [123] 苗夺谦, 王珏. 基于粗糙集的多变量决策树构造方法[J]. 软件学报, 1997, 8(6): 425 - 431.
- MIAO Duoqian, WANG Jue. Rough sets based approach for multivariate decision tree construction[J]. Journal of Software, 1997, 8(6): 425 - 431.
- [124] JELONEK J. Rough set reduction of attributes and their domains for neural networks[J]. Inter J of Computational Intelligence, 1995, 11(2): 339 - 347.
- [125] 沈亚兰. 基于粒度计算的数据挖掘方法的研究[D]. 沈阳: 沈阳工业大学, 2006.
- SHEN Yalan. Research on methods of data mining based on granular computing[D]. Shenyang: Shenyang University of Technology, 2006.
- [126] 张持健. 商空间下模糊系统与模糊控制的问题求解[D]. 合肥: 安徽大学, 2005.
- ZHANG Chijian. Problem solving of fuzzy system & fuzzy control based on the theory of quotient space[D]. Hefei: Anhui University, 2005.
- [127] 张燕平. 基于商空间的构造性数据挖掘方法及应用[D]. 合肥: 安徽大学, 2003.
- ZHANG Yanping. A structural method of data mining based on quotient space and its applications[D]. Hefei: Anhui University, 2003.
- [128] LIU Q, FENG J, DENG D Y. Design and implement for diagnosis systems of hemorheology on blood viscosity syndrome based on GrC[A]. RSFDGrC '2003[C]. Berlin, 2003.
- [129] HU J, WANG G Y, ZHANG Q H, et al. Attribute reduction based on granular computing[A]. The Fifth International Conference on Rough Sets and Current Trends in Computing[C]. [S.l.], 2006.
- [130] WEI L, MIAO D Q. Application of granular computing in knowledge reduction[A]. First International Conference RSKT 2006[C]. [S.l.], 2006.
- [131] 王国胤, 何晓. 一种不确定性条件下的自主式知识学习模型[J]. 软件学报, 2003, 14(6): 1096 - 1102.
- WANG Guoyin, HE Xiao. A self-learning model under uncertain condition[J]. Journal of Software, 2003, 14(6): 1096 - 1102.
- [132] WANG G Y. Domain-oriented data-driven data mining based on rough set[A]. Proc Int Forum on Theory of GrC from Rough Set Perspective[C]. [S.l.], 2006.

作者简介:



王国胤,男,1970年生,教授,博士生导师,博士,上海交通大学、西安交通大学、西南交通大学、电子科技大学和西安电子科技大学兼职教授,国际粗糙集学会常务理事兼顾问委员会主席,IEEE高级会员,中国人工智能学会常务理事兼粗糙集与软计算专业委员会主任委员,中国计算机学会人工智能与模式识别专业委员会常务理事,重庆计算机学会副理事长,重庆市人工智能学会副理事长,《Transactions on Rough Sets》、《International Journal of Cognitive Informatics & Natural Intelligence》、《计算机学报》等6种学术期刊的编委。王国胤教授

组织召开了10个国际国内学术会议,并担任会议主席或共同主席,是10多个系列性国际学术会议的程序委员会委员,应邀在MSRAS2004、RSGrC2007、PReMI2007、IECT2007等国际会议,以及CRSSC2002、CAAI2005、CNNC2006等全国会议上作特邀报告。主持国家自然科学基金、国家863计划等20多项国家级、省部级科研项目。获重庆市自然科学二等奖、三等奖各1次。在国内外主要学术刊物和学术会议上发表200多篇,其中被国际权威检索刊物SCI、EI等收录130多篇(次),论著被他人引用1400多次。出版学术专著2部。

E-mail: wanggy@cqupt.edu.cn.



张清华,男,1974年生,博士研究生,主要研究方向为粒计算、模糊集理论、粗糙集理论。

E-mail: zhangqh@cqupt.edu.cn.



胡军,男,1977年生,博士研究生,主要研究方向为粒计算、粗糙集理论、知识获取等。

E-mail: hujun@cqupt.edu.cn.