粗糙集理论与应用研究综述

王国胤[1)] 姚一豫[2)] 于 洪[1),2)]

[1)](重庆邮电大学计算机科学与技术研究所 重庆 400065)

[2)](加拿大里贾纳大学计算机科学系 里贾纳 加拿大 S4S 0A2)

摘 要 在阐释粗糙集理论基本体系结构的基础上,从多个角度探讨粗糙集模型的研究思路,分析粗糙集理论与模糊集、证据理论、粒计算、形式概念分析、知识空间等其它理论之间的联系,介绍国内外关于粗糙集理论研究的主要方向和发展状况,讨论当前粗糙集理论研究的热点研究领域以及将来需要重点研究的主要问题.

关键词 粗糙集;模糊集;粒计算;形式概念分析;知识空间;智能信息处理

中图法分类号 TP18 **DOI 号**: 10.3724/SP.J.1016.2009.01229

A Survey on Rough Set Theory and Applications

WANG Guo-Yin[1)] YAO Yi-Yu[2)] YU Hong[1),2)]

[1)](*Institute of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing* 400065)

[2)](*Department of Computer Science, University of Regina, Regina, Saskatchewan, Canada* S4S 0A2)

Abstract This paper presents a framework for a systematic study of the rough set theory. Various views and interpretations of the theory and different approaches to study the theory are discussed. The relationships between the rough sets and other theories, such as fuzzy sets, evidence theory, granular computing, formal concept analysis, knowledge spaces, etc., are examined. The paper also reviews recent theoretic studies and applications of rough sets and points out future research directions.

Keywords rough sets, fuzzy sets, granular computing, formal concept analysis, knowledge spaces, intelligent information processing

1 引 言

智能信息处理是当前信息科学理论和应用研究中的一个热点领域.由于计算机科学与技术的发展,特别是计算机网络的发展,每日每时为人们提供了大量的信息.信息量的不断增长,对信息分析工具的要求也越来越高,人们希望自动地从数据中获取其潜在的知识.特别是近 20 年间,知识发现(规则提取、数据挖掘、机器学习)受到人工智能学界的广泛重视,知识发现的各种不同方法应运而生.

粗糙集(Rough Set,也称 Rough 集、粗集)理论是 Pawlak 教授于 1982 年提出的一种能够定量分析处理不精确、不一致、不完整信息与知识的数学工具[1].粗糙集理论最初的原型来源于比较简单的信息模型,它的基本思想是通过关系数据库分类归纳形成概念和规则,通过等价关系的分类以及分类对于目标的近似实现知识发现.

收稿日期:2009-03-29.本课题得到国家自然科学基金(60573068,60773113)、重庆市自然科学基金(2008BA2017)、重庆市杰出青年科学基金(2008BA2041)和重庆市教委科研基金(KJ080510)资助.**王国胤**,男,1970 年生,博士,教授,中国计算机学会理事,主要研究领域包括粗糙集、粒计算、神经网络、机器学习、数据挖掘、知识技术等.E-mail: wanggy@cqupt.edu.cn.**姚一豫**,男,1962 年生,博士,教授,主要研究领域包括 Rough 集、粒计算、信息检索、Web 智能、认知信息学和数据挖掘等.**于 洪**,女,1972 年生,博士,副教授,主要研究方向包括粗糙集、智能信息处理等.E-mail: yuhong@cqupt.edu.cn.

由于粗糙集理论思想新颖、方法独特，粗糙集理论已成为一种重要的智能信息处理技术[2-4]，该理论已经在机器学习与知识发现、数据挖掘、决策支持与分析等方面得到广泛应用. 目前，有 3 个有关粗糙集的系列国际会议，即 RSCTC、RSFDGrC 和 RSKT. 中国学者在这方面也取得了很大的成果，从 2001 年开始每年召开中国粗糙集与软计算学术会议；RSFDGRC2003、IEEE GrC2005、RSKT2006、IFKT2008、RSKT2008、IEEE GrC2008 等一系列国际学术会议在中国召开.

粗糙集理论与应用的核心基础是从近似空间导出的一对近似算子，即上近似算子和下近似算子（又称上、下近似集）. 经典 Pawlak 模型中的不分明关系是一种等价关系，要求很高，限制了粗糙集模型的应用. 因此，如何推广定义近似算子成为了粗糙集理论研究的一个重点.

目前，常见的关于推广粗糙集理论的研究方法有两种，即构造化方法和公理化方法. 构造化方法是以论域上的二元关系、划分、覆盖、邻域系统、布尔子代数等作为基本要素，进而定义粗糙近似算子，从而导出粗糙集代数系统. 公理化方法的基本要素是一对满足某些公理的一元集合算子，近似算子的某些公理能保证有一些特殊类型的二元关系的存在；反过来，由二元关系通过构造性方法导出的近似算子一定满足某些公理.

事实上，有两种形式来描述粗糙集，一个是从集合的观点来进行，一个是从算子的观点来进行. 那么，从不同观点采用不同的研究方法就得到粗糙集的各种扩展模型. 扩展模型的研究以及基于其上的应用研究已经成为新的研究热点.

粗糙集理论与其它处理不确定和不精确问题理论的最显著的区别是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，所以对问题的不确定性的描述或处理可以说是比较客观的，由于这个理论未能包含处理不精确或不确定原始数据的机制，所以这个理论与概率论、模糊数学和证据理论等其它处理不确定或不精确问题的理论有很强的互补性. 因此，研究粗糙集理论和其它理论的关系也是粗糙集理论研究的重点之一.

基于粗糙集理论的应用研究主要集中在属性约简、规则获取、基于粗糙集的计算智能算法研究等方面. 由于属性约简是一个 NP-Hard 问题，许多学者进行了系统的研究. 基于粗糙集的约简理论发展为数据挖掘提供了许多有效的新方法. 比如，针对不同的信息系统（协调的和不协调的、完备的和不完备的），结合信息论、概念格、群体智能算法技术等都有了相应的研究成果.

基于粗糙集理论的应用也涌现在各行各业. 许多学者将粗糙集理论应用到了工业控制[5-8]、医学卫生及生物科学[9-11]、交通运输[12-14]、农业科学[15-16]、环境科学与环境保护管理[17]、安全科学[18]、社会科学[19]、航空、航天和军事等领域[20-21].

粗糙集理论发展二十余年来，无论在理论研究还是应用研究上都取得了很多成果. 从认知科学的角度讲，我们如果要学习一个新的学科，就必须建立它的系统体系结构，同时学习思维及计算方法，这样我们就能从已知的结果推到未知的结果. 本文将在总结已有的这些研究成果的基础上，帮助读者建立起一个这样的系统体系结构，同时指出进一步的研究方向. 我们将这个理论目前的研究状况介绍给信息科学工作者，希望进一步推动并促进我国在这一领域的研究工作.

本文第 2 节介绍粗糙集理论基础；第 3 节介绍粗糙集模型研究，将从构造化方法和公理化方法、面向集合的观点和面向算子的观点来阐述；第 4 节将探讨粗糙集理论和证据理论、模糊集、形式概念分析、知识空间等的关系；第 5 节是基于粗糙集的研究以及应用. 最后是总结和展望.

2 粗糙集理论基础

本节在回顾粗糙集基础概念的基础上，说明常见的两种研究粗糙集的方法：构造化方法和公理化方法. 并且，从集合观点和算子观点来解释粗糙集.

2.1 概念、可定义集

为了对知识进行描述，首先需要知道什么是概念. 从经典的角度来看，每个概念都包含其内涵和外延. 为了给出概念内涵和外延的具体描述，我们考虑一个简单的知识表达系统，即信息表. 信息表就是一组对象的集合，对象通过一组属性来描述. 表 1 就是一个信息表的例子.

表 1 信息表实例

个体编号	头疼	肌肉疼	体温	流感
x_1	是	是	正常	否
x_2	是	是	高	是
x_3	是	是	很高	是
x_4	否	是	正常	否
x_5	否	否	高	否
x_6	否	是	很高	是

信息表 M 可以形式化地表达为四元组 $M=(U,At,\{V_a\mid a\in At\},\{I_a\mid a\in At\})$. 表 1 中，$U=\{x_1,x_2,\cdots,x_6\}$是有限非空对象的集合，也称为论域，$At=\{$头疼，肌肉疼，体温，流感$\}$是有限非空的属性集合. V_a表示属性 $a\in At$ 的属性值的范围，即属性 a 的值域，$I_a:U\rightarrow V_a$是一个信息函数. 如果$A\subseteq At$，则 $I_A(x)$表示 U 中对象 x 在属性 A 上的属性值.

为了形式化地定义概念的内涵，可以采用决策逻辑语言[22]来分析信息表. 我们定义和讨论的决策逻辑语言$\mathcal{L}$由原子公式组成，公式是一种(属性，数据)对，用命题联词：与、或、非等通过标准的方法构成复合公式. 公式是用来描述论域中对象的工具，可以用来描述论域中具有某些性质的对象的子集. 例如在原子公式中，有序对(头疼，是)解释为在属性“$a=$头疼”上值为“$v=$是”的所有对象的描述.

当 ϕ 为信息表M 中的一个公式时，集合$m(\phi)=\{x\in U,x\models_M\phi\}$称为 M 中公式 ϕ 的含义. 含义$m(\phi)$的自变量是语言的公式，其值是信息表中对象集合的子集. $m(\phi)$就是那些具有公式 ϕ 的性质的对象的全体. 换句话说，公式 ϕ 可以描述对象子集 $m(\phi)$. 这样，就建立起了公式 ϕ 和论域 U 的子集之间的关系.

利用决策逻辑语言$\mathcal{L}$，可以给出概念的形式描述：信息表 M 中的概念就是$(\phi,m(\phi))$，其中 $\phi\in\mathcal{L}$. 概念$(\phi,m(\phi))$的内涵是 ϕ，表示 M 中对对象子集 $m(\phi)$的描述；概念$(\phi,m(\phi))$的外延是$m(\phi)$，其含义是满足公式 ϕ 的所有对象的全体.

在粗糙集理论的很多应用中，经常考虑的只是一个属性子集 $A\subseteq At$，即在决策逻辑语言中只考虑 A 中的属性. 我们用符号$\mathcal{L}(A)$表示由属性子集 A 定义的语言. 将前面讨论中出现的$\mathcal{L}$用$\mathcal{L}(A)$来代替，相应的结论也都成立.

考虑属性子集 $A\subseteq At$ 及其相应的语言$\mathcal{L}(A)$，可定义集的形式化定义[23]如下.

定义 1.　在信息表 M 中，如果称子集 $X\subseteq U$ 是可被属性子集 $A\subseteq At$ 定义的，当且仅当在语言$\mathcal{L}(A)$中存在一个公式 ϕ 使得 $X=m(\phi)$. 否则，X 称为不可定义的.

值得注意的是，这里谈到的可定义，是指在属性子集 A 上是可定义的.

例如，表 1 中，我们考虑属性子集 $A=\{$头疼，肌肉疼$\}$，子集 $X_1=\{x_1,x_2,x_3\}\subseteq U$，公式 ϕ_1：(头疼＝是)∧(肌肉疼＝是). 那么在语言$\mathcal{L}(A)$中，显然有 $X_1=m(\phi_1)$，子集$\{x_1,x_2,x_3\}$是可定义集. 而且，子集 $X_2=\{x_4,x_6\}$、$X_3=\{x_5\}$也是可定义集，满足的公式分别是 ϕ_2：(头疼＝否)∧(肌肉疼＝是)；ϕ_3：(头疼＝否)∧(肌肉疼＝否).

根据定义 1，可定义集的全体表示为

$$Def(U,\mathcal{L}(A))=\{m(\phi)\mid\phi\in\mathcal{L}(A)\}.$$

如果概念的外延能用逻辑公式简洁地表达，那它就是一个可定义的概念；从这个角度讲，概念的外延就是可定义集.

类似地，由语言$\mathcal{L}(A)$定义的概念集合表示为

$$DefCon(U,\mathcal{L}(A))=\{(\phi,m(\phi))\mid\phi\in\mathcal{L}(A)\}.$$

很明显，如果两个对象 x_i,x_j是等价的，那么他们在语言$\mathcal{L}(A)$中由相同的公式描述，或者说他们在 A 上的各个属性值相同. 刚才得到的可定义集就是属性集合 A 上的等价关系$E(A)$在论域U 上产生的划分，记为 $U/E(A)=\{[x]_{E(A)}\mid x\in U\}$，$[x]_{E(A)}$是由关系 $E(A)$确定的等价类，同一个等价类中的对象是不可分辨的，所以，有时我们也称等价关系为不可分辨关系. 上例中，$U/E(A)=\{\{x_1,x_2,x_3\},\{x_4,x_6\},\{x_5\}\}$.

等价类$[x]_{E(A)}$的并显然是可定义集. 比如，考虑子集 $X=\{x_1,x_2,x_3,x_4,x_6\}$，它是等价类的并，描述这个子集的公式为 $\phi=\phi_1\vee\phi_2$.

2.2　近似空间

语言$\mathcal{L}(A)$的所有可定义集正好构造成一个 σ 代数$\sigma(U/E(A))$，即

$$Def(U,\mathcal{L}(A))=\sigma(U/E(A)).$$

序对 $apr=(U,E(A))$称为一个 Pawlak 近似空间，简称近似空间. 所以，也可以将语言$\mathcal{L}(A)$的所有可定义集记为 $Def(U,\mathcal{L}(A))=Def(apr)$.

通过 $U/E(A)$，可以构造一个 σ 代数，即 $\sigma(U/E(A))$，它包含空集$\varnothing$和等价关系 $E(A)$构成的等价类及其并，并且在交、并和补运算上是封闭的. 那么，Pawlak 近似空间也唯一确定了一个拓扑空间$(U,\sigma(U/E(A)))$.

2.3　上下近似

针对不可定义集，显然不可能构造一个公式来精确描述，只能通过上下界逼近的方式来刻画，这就是粗糙集理论中的上下近似算子.

定义 2.　设 $E(A)$是信息表 M 上的等价关系，$X\subseteq U$，上下近似算子$\overline{apr}_{E(A)}$，$\underline{apr}_{E(A)}$(下文我们采用缩写形式$\overline{apr}$，$\underline{apr}$)定义为

$$\begin{aligned}\overline{apr}(X)&=\bigcup\{Y\mid Y\in\sigma(U/E(A)),Y\cap X\neq\varnothing\}\\&=\bigcap\{Y\mid Y\in Def(U,\mathcal{L}(A),X\subseteq Y\};\end{aligned}$$

$$\underline{apr}(X)=\bigcup\{Y|Y\in\sigma(U/E(A)),Y\subseteq X\}$$
$$=\bigcup\{Y|Y\in Def(U,\mathcal{L}(A),Y\subseteq X\}.$$

上近似$\overline{apr}(X)$是包含 X 的最小可定义集，下近似$\underline{apr}(X)$是包含在 X 中的最大可定义集.

根据定义 2，可定义集显然有相同的上下近似. 刚才我们在可定义的基础上构造了一对近似算子. 也就是说，只有当对象不可定义时，才会用上下近似的方法来描述.

考虑子集 $X\subseteq U$，论域空间将被分成 3 个区域：

(1) 集合 X 的正域：

$$POS(X)=\underline{apr}(X);$$

(2) 集合 X 的负域：

$$NEG(X)=POS(\sim X)=U-\overline{apr}(X);$$

(3) 集合 X 的边界域：

$$BND(X)=\overline{apr}(X)-\underline{apr}(X).$$

如果 $BND(X)$是空集，则称集合 X 关于关系 $E(A)$是清晰的(crisp)；反之，如果 $BND(X)$不是空集，则称集合 X 为关于关系 $E(A)$粗糙的(rough).

2.4 粗糙集

Pawlak[1,22] 定义由等价关系确定的等价类 $[x]_{E(A)}$的集合就组成了 P1-粗糙集集合(P1-Rough Set，PRS1). 显然，P1-粗糙集集合是子集集合，即 $PRS1=\{[x]_{E(A)}\mid X\subseteq 2^U\}$.

借助上下近似的描述，也可以给出和 $PRS1$ 等价的关于粗糙集的另外一种定义，称为 P2-粗糙集集合. 即 $PRS2=\{\langle X_1,X_2\rangle\}=\{\langle\underline{apr}(X),\overline{apr}(X)\rangle\}$.

$PRS1$ 和 $PRS2$ 通称为 Pawlak 粗糙集.

有很多形式来描述粗糙集，比如 Iwinski[24] 在布尔代数 2^U的子代数上来描述粗糙集，他定义粗糙集为一对可定义集；Yao[25] 在论域 U 的子集空间 2^U 上描述粗糙集，定义粗糙集为一个闭区间集合. 从代数学意义上讲，Iwinski-粗糙集代数和区间代数是等价的，Iwinski-粗糙集代数系统可以看作为一个上下界是可定义集的区间代数系统.

同时，Pawlak、Skowron、Wong 和 Yao[26-27] 等也提出用论域 U 的元素 x 来描述粗糙集. 一种简单有效解释粗糙集的方法就是应用 3 个成员(隶属度)函数来描述，即粗糙隶属度函数 $\mu_X(x)$、强隶属度函数 $\mu_{\underline{apr}(X)}(x)$和弱隶属度函数 $\mu_{\overline{apr}(X)}(x)$.

关于粗糙隶属度函数的定义，有多种形式[28-29]. 比如，文献[28]采用三值逻辑$[0,\alpha,1]$定义的粗糙隶属度函数建立起了粗糙集和模糊集的关系，用模糊集合来描述 Iwinski-粗糙集集合，并给出了它们之间的关系. 换句话说，采用此种隶属度函数来描述粗糙集，建立起了粗糙集和模糊集的关系.

基于文献[28]的隶属度函数，许多学者研究并提出了各种概率模型：模糊集意义下的 α 截集、经典 Pawlak 粗糙集、可变精度粗糙集(VPRS)、决策粗糙集理论等.

以上对粗糙集的解释都是从集合的观点进行的；还存在另外一种观点，即从算子的观点来解释粗糙集. 在面向算子的观点中，上下近似被看作是论域幂集空间 2^U上的一对一元算子 $\boldsymbol{L}$ 和 $\boldsymbol{H}$. 也就是说，粗糙集理论中研究的系统$(2^U,\sim,\cap,\cup,\boldsymbol{L},\boldsymbol{H})$是标准集合系统$(2^U,\sim,\cap,\cup)$附加了两个近似算子的扩展. 文献[30]对各种算子的定义及其相互关系作了探讨.

2.5 理论研究方法

经典粗糙集理论的基本思想是基于等价关系的粒化与近似的数据分析方法. 粗糙集理论与应用的核心基础是从近似空间导出的一对近似算子，即上近似算子和下近似算子(又称上、下近似集). 目前，主要有两种研究方法来定义近似算子：构造化方法和公理化方法.

2.1 节中我们采用构造化方法回顾了经典的 Pawlak 粗糙集模型. 构造化方法的主要思路就是通过直接使用二元关系的概念来定义粗糙集的近似算子，从而导出粗糙集代数系统$(2^U,\cap,\cup,\sim,\underline{apr},\overline{apr})$. 构造化方法所研究的问题往往来源于实际，所建立的模型有很强的应用价值，其主要缺点是不易深刻体现近似算子的公理(代数)性质. 所以，也有许多学者从公理化的角度来研究粗糙集.

公理化方法也称为代数方法，有时也称为算子方法，这种方法不像构造化方法中是以二元关系为基本要素的，它的基本要素是一对满足某些公理的一元近似算子 $\boldsymbol{L},\boldsymbol{H}:2^U\rightarrow 2^U$，即粗糙代数系统$(2^U,\sim,\cap,\cup,\boldsymbol{L},\boldsymbol{H})$中近似算子 $\boldsymbol{L}$ 和 $\boldsymbol{H}$ 是事先给定的. 然后再去找二元关系使得由该二元关系及其生成的近似空间按构造化方法导出的近似算子恰好就是给定的由公理化方法定义的集合算子. 近似算子的某些特殊公理能保证有一些特殊类型的二元关系存在，使这些关系能够通过构造方法产生给定的算子；反之，由二元关系通过构造方法导出的近似算子一定满足某些公理，使这些公理通过代数方法产生给定的二元关系.

用公理化方法研究粗糙集最早的是 Lin 和 Liu[31]. 由于 Lin 和 Liu 给出的公理组不独立，文献[32]对其进行了改进. Yao 与 Thiele 从不同的角度

分别对于一般关系下的经典粗糙集近似算子作了比较完整的公理化刻画. 文献[33]刻画了一般关系下各种经典粗糙近似算子的独立性公理集.

目前，关于粗糙集理论的公理化研究，已经取得了进一步的成果[34-36]. 关于公理化的研究主要从公理组的极小化及独立性两方面展开研究工作. 近年来，许多学者也展开了关于模糊粗糙近似算子、粗糙模糊近似算子、直觉模糊粗糙近似算子的构造性定义及其公理集的研究；其中，关于公理集的最小化问题、独立性问题还有待进一步的研究.

目前，关于构造化方法研究粗糙集理论的成果相对多些，比如我们后面的很多介绍都是采用构造方法进行的.

3　粗糙集模型扩展

粗糙集模型扩展是粗糙集理论研究的一个重要方向. 结合其它理论方法与技术，大量的研究成果涌现. 本节通过举例说明如何进行模型扩展，希望起到抛砖引玉的作用.

本节首先介绍上下近似的扩展定义，然后从粗糙集基于元素、粒、子系统、论域、关系、集合或近似空间等方向介绍相关的一些扩展模型. 这些方面又常常不是单独出现的，还可以交叉、综合.

3.1　上下近似的扩展定义

如果二元关系 R 是等价关系 $[x]_R$，在近似空间 (U,R) 上就得到 Pawlak 的基于元素的定义[37]；如果 R 是等价关系 $[x]_R$，在近似空间 $(U,\sigma(U/R))$ 上则有基于粒的定义[37]；如果二元关系 R 是子集（子系统），在近似空间 $(U,\sigma(U/R))$ 上则有基于子系统的定义[37].

基于元素（element based）的定义：

$$\overline{apr}(X)=\{x\mid x\in U,[x]_R\cap X\neq\varnothing\}$$
$$=\{x\mid x\in U,\exists y\in U[xRy,y\in X]\},$$
$$\underline{apr}(X)=\{x\mid x\in U,[x]_R\subseteq X\}$$
$$=\{x\mid x\in U,\forall y\in U[xRy\Rightarrow y\in X]\}.$$

基于粒（granule based）的定义：

$$\overline{apr}(X)=\bigcup\{[x]_R\mid[x]_R\in\sigma(U/R),[x]_R\cap X\neq\varnothing\},$$
$$\underline{apr}(X)=\bigcup\{[x]_R\mid[x]_R\in\sigma(U/R),[x]_R\subseteq X\}.$$

基于子系统（subsystem based）的定义：

$$\overline{apr}(X)=\bigcap\{Y\mid Y\in\sigma(U/R),X\subseteq Y\},$$
$$\underline{apr}(X)=\bigcup\{Y\mid Y\in\sigma(U/R),Y\subseteq X\}.$$

上述 3 种定义分别从元素、粒和子系统的角度对等价关系进行了阐述. 这 3 种等价的定义给出了粗糙集理论中上下近似的不同表达方式.

在基于元素的定义中，如果一个元素 x 的所有等价元素（也就是它的等价类）都在集合 X 中，则 x 在 X 的下近似 $\underline{apr}(X)$ 中；如果至少有一个 x 的等价元素在 X 中，则这个元素 x 在 X 的上近似 $\overline{apr}(X)$ 中. 在基于粒的定义中，所有包含于 X 的等价类的并组成下近似 $\underline{apr}(X)$，所有和 X 交集不为空的等价类的并组成上近似 $\overline{apr}(X)$. 在基于子系统的定义中，下近似 $\underline{apr}(X)$ 就是包含于 X 的在子系统 $\sigma(U/R)$ 中的那些最大可定义集，上近似 $\overline{apr}(X)$ 就是那些包含 X 的在子系统 $\sigma(U/R)$ 中的最小可定义集.

有了上下近似的定义，就很容易得到粗糙集理论中的其它概念的定义，比如边界域、正域、负域等. 这 3 种定义为结合其它理论扩展粗糙集模型建立起了联系.

3.2　基于元素的扩展模型

设 $R\subseteq U\times U$ 是论域上的一个任意二元关系，则其定义了一个扩展的近似空间 $apr=(U,R)$.

从集合的观点来看，利用非等价关系显然可以扩展 3.1 节中基于元素的粗糙集定义. 例如，将元素 x 所在的等价类 $[x]_R$ 看成是 x 的一个邻域，从而得到基于邻域的粗糙集模型[38]. 邻域关系 $R_s(x)$ 只要求满足自反性，不要求一定满足对称性或者传递性. 那么在基于元素的定义中，将等价类 $[x]_R$ 用非等价关系 $R_s(x)$ 代替，就得到基于非等价关系 $R_s(x)$ 的粗糙集模型：

$$\overline{apr}(X)=\{x\mid x\in U,R_s(x)\cap X\neq\varnothing\},$$
$$\underline{apr}(X)=\{x\mid x\in U,R_s(x)\subseteq X\}.$$

经典粗糙集模型是邻域模型中 $R_s(x)$ 为等价关系时的特例.

同理，读者也可以定义其它的非等价关系来扩展粗糙集理论. 比如，为了处理不完备信息系统，已有的多种扩展模型：容差关系、相似关系、量化容差关系、限制容差关系和特征关系等等[39-42]都是利用各种非等价关系来扩展基于元素的粗糙集定义而得到的.

从算子的观点来看，粗糙集模型中的近似算子可以和模态逻辑中的必然性算子和可能性算子相联系起来[43]. 在模态逻辑的公理化系统中，如果必然性算子□用下近似算子 $\boldsymbol{L}$ 来代替，可能性算子◇用上近似算子 $\boldsymbol{H}$ 来代替，非联结符 ¬ 用集合补运算 ～ 代替，合取联结符 ∧ 用集合交运算 ∩ 代替，析取联结符 ∨ 用集合并运算 ∪ 代替，蕴涵 → 用集合包含 ⊆ 代

替,那么得到的公理化系统就是一个粗糙集代数系统.比如,已经有文献基于模态逻辑提出分级模态粗糙集模型和概率模态粗糙集模型等[43-45].

关于模态逻辑的研究成果已经很多,如果将模态逻辑中的研究方法与研究成果移植到粗糙集理论研究中来,或者结合粗糙集理论来研究模态逻辑都将是新的研究方向.

3.3 基于粒的扩展模型

基于粒的粗糙集定义是从等价类(划分)的角度出发来讨论的.在基于 Pawlak 经典粗糙集的粒计算模型中,划分就是一个基本粒.显然,如果扩展划分的概念,就可以得到基于粒的扩展粗糙集模型;同时,这也为粒计算的模型研究指明了新的研究方向.

Zakowski 在文献[46]中将划分扩展到了覆盖.设 C 是论域 U 上的子集族,如果 C 中的所有子集都不空,且 $\bigcup C=U$,则称 C 是 U 的覆盖.那么在基于粒的定义中,用 C 的子集代替 $[x]_R$,覆盖 C 代替 $\sigma(U/R)$,就得到一对覆盖上下近似算子:

$$\overline{apr}\,X=\bigcup\{Y\mid Y\in C,Y\cap X\neq\varnothing\};$$

$$\underline{apr}\,X=\bigcup\{Y\mid Y\in C,Y\subseteq X\}.$$

文献[36,47-50]都对覆盖广义粗糙集理论进行了深入的研究.文献[36]从不同的粒度结构来扩展粗糙集模型,从简单的单层结构到复杂的层次结构以及如何利用非等价关系来扩展粒度概念作了探讨.文献[47]结合形式概念分析来研究覆盖广义粗糙集模型,文献[48]讨论了基于覆盖的约简问题,文献[49-50]研究了各种不同类型的扩展覆盖粗糙集模型及其关系等.

另一方面,基于这些扩展模型的应用研究也得到了发展[51-52],比如覆盖粗糙集扩展模型应用于词计算、社会科学、软件水印与软件混淆、泛逻辑中等相关的成果已经开始出现.

目前,基于粒的粗糙集理论扩展主要结合了覆盖和形式概念分析理论,如何结合其它粒计算工具扩展粗糙集理论将是未来的研究方向之一.

3.4 基于子系统的扩展模型

在标准的基于子系统的粗糙集模型中,定义上下近似用到的是相同的子系统.如果要扩展基于子系统的定义,我们需要两个子系统,一个用于定义上近似,一个用于定义下近似[53].

例如,在拓扑空间$(U,O(U))$中,子集族 $O(U)\subseteq 2^U$ 称为 U 上的开集;开集的补集称为闭集,记为 $\boldsymbol{C}(U)=\{\neg X\mid X\in O(U)\}$.开集包含$\varnothing$和 U,并且在并运算和有限次交运算下是封闭的;闭集包含$\varnothing$和 U,在交运算和有限次并运算下封闭.那么,在 3.1 节基于子系统的下近似定义中用 $O(U)$ 代替 $\sigma(U/R)$,上近似定义中用 $\boldsymbol{C}(U)$ 代替 $\sigma(U/R)$ 就得到扩展了的模型:

$$\overline{apr}(X)=\bigcap\{Y\mid Y\in\boldsymbol{C}(U),X\subseteq Y\};$$

$$\underline{apr}(X)=\bigcup\{Y\mid Y\in O(U),Y\subseteq X\}.$$

在这种情况中,上下近似算子事实上就是拓扑中的内部算子和闭包算子.经典粗糙集模型是拓扑空间中开集和闭集相等时的特例.

同理,也可以结合拓扑、闭系统、布尔代数、格、偏序等来扩展粗糙集理论;或者,从其它理论出发来探讨他们同粗糙集理论的关系.设计合适的子系统是研究的关键.例如,文献[54]在完全分配格上定义了一般化的粗糙近似算子,文献[55]在完备的原子布尔格上定义了粗糙近似并且研究了粗糙近似的结构,文献[56]在布尔代数的基础上讨论了约简问题.相关的应用也出现在数据挖掘、信息检索[57]等领域.

3.5 双论域模型

在经典 Pawlak 粗糙集模型中考虑的论域通常只有一个 U,我们也可以从论域上来推广粗糙集理论.

文献[58]第一次将粗糙集模型推广到了两个不同但相关联的论域上.设 U,V 是两个论域,元素 $u\in U$ 和 $v\in V$ 是相容的,记为 $u\mathcal{C}U$.不失一般性,假定针对每个 $u\in U$,都有一个 $v\in V$ 存在,使得他们是相关联的,反之亦然.那么 U,V 之间的相容关系 $\mathcal{C}$ 就可以用一个多值映射 $\gamma:U\rightarrow 2^V$ 来定义,即 $\gamma(u)=\{v\in V\mid u\mathcal{C}v\}$.为了扩展粗糙集模型,文献[58]定义了如下一对上下近似:

$$\overline{apr}_{\mathcal{C}}(X)=\{u\in U\mid\gamma(u)\cap X\neq\varnothing\};$$

$$\underline{apr}_{\mathcal{C}}(X)=\{u\in U\mid\gamma(u)\subseteq X\}.$$

上述定义也可以看作是 3.1 节基于元素定义的模型推广,这时的二元关系就变成为两个论域笛卡尔乘积的一个子集.

文献[59]进一步研究推广了基于双论域的粗糙集模型,并将其应用于不确定性推理中.文献[59-61]结合模糊集将该理论推广到多个论域,其特点是论域 U 中的模糊集 X 的上下近似是由另一个论域 V 中的元素来表达的.

如何结合其它理论从多论域的角度来研究粗糙集理论还有待进一步的工作.

3.6 概率模型

根据是否使用了统计信息,粗糙集模型扩展大致可以分为两类:一类是经典的代数粗糙集模型,另

一类是概率型的粗糙集模型. 前述各种模型都是基于代数粗糙集模型的扩展. 概率型的粗糙集模型在粗糙隶属度函数的基础上来讨论.

对于任意 $X\subseteq U$，粗糙隶属度函数定义为

$$\mu_X(x)=\frac{|X\cap[x]_R|}{|[x]_R|}.$$

文献[27]提出了决策粗糙集理论，得到了 X 的上下近似定义：

$$\begin{aligned}\overline{apr}_{\beta,\delta}(X)&=POS_{\beta,\delta}(X)\cup BND_{\beta,\delta}(X)\\&=\{x\mid x\in U,P(X\mid r(x))>\delta\},\\\underline{apr}_{\beta,\delta}(X)&=POS_{\beta,\delta}(X)\\&=\{x\mid x\in U,P(X\mid r(x))\geq\beta\}.\end{aligned}$$

进一步分析会发现许多概率型粗糙集模型可以由决策粗糙集导出，它们均可以视为决策粗糙集的特例.

比如，当 $\beta=1,\delta=0$ 时，概率函数取

$$P(X\mid r(x))=\frac{|X\cap r(x)|}{|r(x)|},$$

那么 $\underline{apr}_{1,0}(X)$ 和 $\overline{apr}_{1,0}(X)$ 将表示为

$$\overline{apr}_{1,0}(X)=\{x\mid x\in U,r(x)\cap X\neq\varnothing\},$$

$$\underline{apr}_{1,0}(X)=\{x\mid x\in U,r(x)\subseteq X\}.$$

显然，如果 $r(x)$ 是等价关系，这个模型就是经典 Pawlak 粗糙集模型.

文献[61]指出经典代数粗糙集模型的约简理论不再适用于概率型粗糙集模型，由此提出了决策粗糙集约简所需保持不变的若干特征，并系统阐述了决策粗糙集约简理论. 决策粗糙集理论在网络支持系统、属性选择和信息过滤中得到了应用[62-65].

更多的基于概率的模型[66-69]得到发展，比如：0.5-概率粗糙集模型、可变精度粗糙集模型(VPRS)、参数化粗糙集模型和贝叶斯粗糙集模型等.

目前，概率型粗糙集的有关研究主要有 3 个重点：(1) 概率型上下近似集和正、负、边界区域特征；(2) 概率型规则的语义解释；(3) 概率型粗糙集属性约简理论.

4　粗糙集与其它不确定信息处理理论的联系

随着对粗糙集理论研究的不断深入，与其它数学分支的联系也更加紧密. 粗糙集理论研究不但需要以这些理论作为基础，同时也相应地推动这些理论的发展. 本节探讨粗糙集理论与其它不确定信息处理理论的关系.

4.1　粗糙集和证据理论

粗糙集理论与 Dempster-Shafer 证据理论(D-S 证据理论)在处理不确定性的问题方面，产生和研究的方法是不同的，但却有某种相容性，粗糙集理论是为开发规则的机器自动生成而提出的，而 D-S 证据理论主要用于证据推理.

D-S 证据理论用一对信任函数和似然函数在给定证据下对假设进行估计和评价. 利用基本概率分配函数 $m:2^U\to[0,1]$，信任函数和似然函数定义为

$$Bel(X)=\sum_{A\subseteq X}m(A),\quad Pl(X)=\sum_{A\cap X\neq\varnothing}m(A).$$

在证据理论中，基本概率分配函数(即 mass 函数)一般是由主观给出的可信度，所以又被称为可信度分配函数.

观察证据理论中信任函数和似然函数的定义形式，可以发现其和粗糙集理论中的上下近似有相似之处. 文献[70-71]讨论了二者的关系，得出以下结论：在一个随机近似空间 $apr=(U,R,(\sigma(M),P))$ 上，对于任意 $X\subseteq U$，定义

$$Bel(X)=P(\underline{apr}(X)),\quad Pl(X)=P(\overline{apr}(X)),$$

则 $Bel(X)$ 和 $Pl(X)$ 是 U 上一对对偶的信任函数和似然函数，其对应的 mass 函数为

$$m(X)=\begin{cases}P(X), & X\in U/R\\0, & \text{其它}\end{cases}.$$

在这种情况下，信任结构中的焦元是两两不相交的，是论域上的划分，此时信任函数退化为内概率. $P(\overline{apr}(X))$ 和 $P(\underline{apr}(X))$ 分别为 X 的上、下近似随机质量：

$$P(\overline{apr}(X))=\frac{|\overline{apr}(X)|}{|U|},$$

$$P(\underline{apr}(X))=\frac{|\underline{apr}(X)|}{|U|}.$$

可见，粗糙集理论中的下近似和上近似的概率恰好分别是信任函数和似然函数[70]，然而生成信任函数和似然函数的基本概率分配函数方法是不同的，前者来自于系统中数据本身，比较客观，而后者往往来自于专家的经验，带有很强的主观性.

文献[72-73]等进一步讨论了粗糙集理论与 D-S 证据理论的关系，二者有很强的互补性. 从代数学意义上来讲，任何具有概率特性的信任函数都可以有相应的扩展粗糙集模型，比如，在串行代数、区间代数等结构上已有一些研究成果出现.

4.2　粗糙集和模糊集

模糊集和粗糙集理论在处理不确定性和不精确

性问题方面都推广了经典集合论，两个理论的比较和融合一直是人们感兴趣的话题. 鉴于粗糙集理论和模糊集理论之间的互补性，文献[74-79]提出了模糊粗糙集和粗糙模糊集的概念.

设(U,R)是近似空间，R是论域U上的等价关系，F是一个模糊集，则$\forall F\in\mathcal{F}(U)$，模糊集$F$在空间$(U,R)$上的上下近似$\overline{apr}_R(F)$、$\underline{apr}_R(F)$是一对模糊集：

$$\mu_{\overline{apr}_R F}(x)=\sup\{\min[\mu_F(y),\mu_R(x,y)]\mid y\in U\},$$

$$\mu_{\underline{apr}_R F}(x)=\inf\{\max[\mu_F(y),1-\mu_R(x,y)]\mid y\in U\} \tag{R-F}$$

$(\underline{apr}_R(F),\overline{apr}_R(F))$称为模糊集$F$在$U$上的粗糙模糊集. Rough-Fuzzy 模型将近似对象由 crisp 集X扩展为了模糊集F.

如果进一步将等价关系R扩展为模糊相似关系$\mathcal{R}$，则有模糊近似空间$(U,\mathcal{R})$. $\forall F\in\mathcal{F}(U)$，模糊集$F$在空间$(U,\mathcal{R})$上的上下近似$\overline{apr}_{\mathcal{R}}(F)$、$\underline{apr}_{\mathcal{R}}(F)$是一对模糊集：

$$\mu_{\overline{apr}_{\mathcal{R}} F}(x)=\sup\{\min[\mu_F(y),\mu_{\mathcal{R}}(x,y)]\mid y\in U\},$$

$$\mu_{\underline{apr}_{\mathcal{R}} F}(x)=\inf\{\max[\mu_F(y),1-\mu_{\mathcal{R}}(x,y)]\mid y\in U\} \tag{F-R}$$

$(\underline{apr}_{\mathcal{R}}(F),\overline{apr}_{\mathcal{R}}(F))$称为模糊集$F$在$U$上的模糊粗糙集. Fuzzy-Rough 模型不仅将近似对象从 crisp 集X扩展到了模糊集F，而且将论域上的等价关系R扩展到了模糊相似关系$\mathcal{R}$.

从式(R-F)、(F-R)显而易见，粗糙模糊集是模糊粗糙集的特例. 而经典粗糙集又是粗糙模糊集中近似对象为 crisp 集时的特例，为了更好地理解这个结论，我们来看粗糙集的另一个定义形式.

如前所述，粗糙集是通过上下近似来描述的，上下近似有时候也称为粗糙集的弱强成员函数. 用μ_X，μ_R来表示X和R的成员函数，注意到集合描述的特征函数表示(即$\mu_X(y)=1$，若$y\in X$；否则为0)，则基于元素的粗糙集定义可改写为

$$\mu_{\overline{apr}_R(X)}(x)=\sup\{\mu_X(y)\mid y\in U,(x,y)\in R\},$$

$$\mu_{\underline{apr}_R(X)}(x)=\inf\{\mu_X(y)\mid y\in U,(x,y)\in R\}.$$

其含义是直观的，只有任意与x具有等价关系R的y都在集合X中(即$\mu_X(y)=1$)，才有$\mu_{\underline{apr}_R X}(x)=1$(即$x\in\underline{apr}_R X$). 另一式理解类似. 我们还可以进一步改写上式得到

$$\mu_{\overline{apr}_R(X)}(x)=\sup\{\min[\mu_X(y),\mu_R(x,y)]\mid y\in U\},$$

$$\mu_{\underline{apr}_R(X)}(x)=\inf\{\max[\mu_X(y),1-\mu_R(x,y)]\mid y\in U\} \tag{R}$$

事实上，上式也是一对模糊集. 观察式(R)、(R-F)，显然经典粗糙集是粗糙模糊集在近似对象为 crisp 集X时的特例.

既然粗糙集、模糊粗糙集和粗糙模糊集是模糊集，那它们一定可以用α-截集的形式来表达，文献[75]讨论了该工作. 另外，文献[76]结合概率粗糙集模型进行了推广；文献[77]讨论了应用模糊粗糙集于属性约简和决策规则的获取. 文献[78-81]等讨论了各种扩展模型应用于信息检索、文字识别、股票价格预测等.

4.3 粗糙集和形式概念分析

形式概念分析是一种从形式背景建立概念格来进行数据分析的方法. 形式背景(U,V,R)由对象集U和属性集V的一个二元关系R给出. 在一个形式背景中，我们可以构造(对象，属性)对，称为形式概念，记为(A,B)，其中$A\subseteq U$，$B\subseteq V$. 形式概念中的对象A称为外延，即$extent(A,B)=A$；形式概念中的属性B称为内涵，即$intent(A,B)=B$. 所有的形式概念组成一个完全格，称为概念格，记为$L(U,V,R)$，或者简记为L.

形式概念的外延可看成是对象的可定义集，从某种意义上来说，这一点和粗糙集的可定义集定义不同. 事实上，形式概念的外延是一个依据内涵得到的不可区分对象集. 根据内涵的属性集，外延中的所有对象都是不可区分的；外延中的所有对象具有内涵中的所有属性. 所有的外延，即对象集可以被认为是一个不同的可定义集系统. 任意的一个对象集可能不是形式概念的外延. 当对象集不是形式概念的外延时，我们认为它是不可定义集. 因此，在形式概念分析中，提出一种不同的可定义性.

所有的外延形成对象论域幂集上的一个子系统，这个子系统对交运算是封闭的. 任何形式概念的外延可看成是一个可定义集. 借助粗糙集上下近似算子，对任意一个对象集$X\subseteq U$，我们可以用概念的外延从上下逼近来描述. 扩展3.1节基于子系统的定义，即集合运算符$\cap$和$\cup$分别用格理论中算子$\wedge$和$\vee$代替，子系统$\sigma(U/R)$用格L代替，对象的可定义集就是形式概念的外延，那么就得到了在近似空间$apr=(U,L)$上基于格理论的关于对象子集X的上下近似定义[82]：

$$\begin{aligned}\overline{lapr}X&=extern(\wedge\{(A,B)\mid(A,B)\in L,X\subseteq A\})\\&=\cap\{A\mid(A,B)\in L,X\subseteq A\},\\\underline{lapr}X&=extern(\vee\{(A,B)\mid(A,B)\in L,A\subseteq X\})\\&=(\cup\{A\mid(A,B)\in L,A\subseteq X\})^{**}.\end{aligned}$$

运用近似算子的子系统定义，由于子系统只对交运算封闭，上近似由一个概念的外延组成而下近似由一组概念的内涵来描述，有可能 X 的下近似不是 X 的子集. 所以，上述算子是一种逼近描述. 粗糙集理论的目的是预测，形式概念分析的目的是描述[83].

目前，基于形式概念分析和粗糙集理论的结合研究已经形成一个新的研究热点. 研究成果体现在以下几方面：(1) 在形式概念分析中提出不同的近似算子[84]；(2) 将粗糙集的结果引入形式概念分析中研究形式概念分析的约简问题[85]；(3) 通过模式-类型算子描述概念和层次概念组织(即概念格)可应用到粗糙集分析中[86]. 此外，形式概念分析中的依赖空间问题也得到了研究[87].

4.4　粗糙集和知识空间

粗糙集理论和知识空间理论都是研究知识结构的理论；但他们用于解决不同的实际问题. 粗糙集主要研究如何对数据进行分析及知识发现；而知识空间[88-89]着重对问题集进行分析，从而对个体知识状态进行评估. 如何将知识空间和粗糙集理论结合正在成为一个新的研究方向.

粗糙集理论和知识空间都在一个有限的论域集以及一些论域集的子集上进行讨论，可记为$(U,\mathcal{K})$，其中$\mathcal{K}\subseteq 2^U$. 在粗糙集中，U 中的元素称为对象，$\mathcal{K}$中的元素称为可定义集；对不可定义集，我们必须通过一对可定义集合分别从上下逼近来表示. 在知识空间中，U 是一组问题集，而$\mathcal{K}$中的元素 K 称为个体的知识状态，$\mathcal{K}$ 称为知识结构. 某个个体的知识状态 K 由问题间的依赖关系或者不同个体掌握不同的问题集决定. 利用 surmise 关系 P[88]（关系 P 满足传递性和自反性），知识结构可以定义为

$$\mathcal{K}=\{K\mid(\forall q,q'\in Q,qPq',q'\in K)\Rightarrow q\in K\}.$$

此定义中，知识结构$\mathcal{K}$既包含空集$\varnothing$也包含问题集 U，并且在集合交运算和集合并运算下封闭. 由此，定义了一个近似空间 $apr=(U,\mathcal{K})$. 那么，在近似空间$(U,\mathcal{K})$上，针对问题子集 $X\subseteq U$，在 3.1 节基于子系统的上下近似定义的基础上有以下扩展定义[90]：

$$\overline{apr}(X)=\{\bigcap\{K\in\mathcal{K}\mid X\subseteq K\}\},$$
$$\underline{apr}(X)=\{\bigcup\{K\in\mathcal{K}\mid K\subseteq X\}\}.$$

在这个定义形式中，知识结构在补运算中不封闭，也就是说，扩展模型不满足经典粗糙集理论中的对偶性质.

读者也可以尝试其它形式的扩展，比如知识结构$\mathcal{K}$只在并运算下封闭，通过分析$\mathcal{K}$中的知识状态，找出最有效的解决问题的捷径.

虽然粗糙集和知识空间研究对象不同，但从粒计算的角度来看，它们都可看成由一些基本粒通过不同的方式构造粒结构的过程. $\mathcal{K}$其实是对知识从不同大小的粒度进行多层次的描述. 关于这方面的研究成果尚不多，希望能有更多的学者参与这个问题的研究.

4.5　粗糙集和粒计算

粒计算是一门飞速发展的新学科. 它融合了粗糙集、模糊集及人工智能等多种理论的研究成果. 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(Granular Compuing，GrC)模型[91-92]. 粗糙集理论已经成为研究粒计算的重要工具.

基于粗糙集模型的粒计算，它的粒是一个划分，是一个特别的粒计算结构. 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 $\sigma(U/R)$代替 U/R，它是布尔代数$(2^U,\sim,\cap,\cup)$的一个子代数，则$(U,\sigma(U/R))$构成一个拓扑空间.

近些年，基于粗糙集理论来研究粒计算的工作尤为突出. 文献[93]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. 文献[94]综述了关于 RNC 模型的主要研究线索. 文献[95]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，文献[45]对粒逻辑进行了探讨.

文献[96]结合粗糙集邻域系统对粒计算进行了详细的研究，为数据挖掘提供了新的方法和视角. 两个覆盖生成相同覆盖广义粗集的判别条件[97]、覆盖粒计算模型的不确定性度量[98]、基于集合论覆盖原理的粒计算模型[99]等也得到了研究.

文献[100-101]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.

结合粗糙集理论的粒计算方法已经在机器学习、数据分析、数据挖掘、规则提取、智能数据处理和粒逻辑等[92]方面取得了一定的应用.

5 基于粗糙集的应用研究

目前,基于粗糙集理论的应用研究主要集中在知识获取、基于粗糙集的计算智能算法研究等方面.这些研究成果成功应用在许多领域,有的已经获得了商业价值.

5.1 知识获取

知识获取是发现存在于数据库中有效的、新颖的、具有潜在效用的乃至最终可理解的模式的非平凡过程.粗糙集理论可支持知识获取的多个步骤[102],如数据预处理、属性约简、规则生成、数据依赖关系获取等.

传统的基于粗糙集理论的数据预处理过程通常包括决策表补齐和决策表离散化.关于这方面的研究工作已经取得一些成果[103-104].

属性约简就是保持信息系统分类能力不变的情况下,约去不必要的属性.属性约简在某些应用领域,又叫数据约简、特征提取、知识约简等.如何求属性约简是约简理论研究的一个重要方面.

基于粗糙集的知识约简理论发展为数据挖掘提供了许多有效的新方法.针对协调决策表,现已提出了求属性约简的许多算法[105-109],如数据分析法、基于信息熵的属性约简算法、动态约简算法、增量式算法、可辨识矩阵算法等.

随着粗糙集理论研究的不断深入,许多学者进一步在等价关系下,讨论了不协调决策表的多种约简[110-114],如广义决策约简、可能性约简、动态约简、分布约简、最大分布约简、β 约简、熵约简及近似熵约简等.

文献[105,115]从信息论的角度进一步研究了属性约简问题;并且修正了以前学术界认为基于代数观和基于信息观的粗糙集理论是等价的观点,得到了一系列有益的结论;文献[116]进一步提出了针对协调决策表和不协调决策表的核属性的不同计算方法.

同时,讨论的信息系统的形式也越来越多,如连续值信息系统、区间值信息系统、模糊值信息系统、集值信息系统等,并且相应系统的约简理论也得到了发展.

另一方面,随着概念格、偏序集等理论与粗糙集理论的结合,基于概念格的约简方法、广义协调决策形式背景知识约简方法、偏序关系下的决策形式背景规则提取与属性约简、对象概念格的属性约简方法、基于用户偏好的属性约简、属性序下的快速约简算法、权值约简、基于群体智能算法的属性方法等新方法[117-127]也大量涌现.比如,文献[125]结合高斯消去法通过矩阵运算直接得到属性约简,为属性约简研究提供了新思路.文献[126]对各种约简研究方法作了总结:从算法结构的层次来说,常见的约简策略有 3 种:删除法、增加法、加删法;而各个约简方法的不同体现在各自的启发式策略的不同上.

5.2 知识的不确定性度量

随着粗糙集理论的研究深入,一种新的不确定性——粗糙性正逐渐被人们认识和接受.至今,人们已经研究分析了 3 种不同的不确定性:随机性,即随机现象的不确定;模糊性,即模糊概念的不确定性;粗糙性,即信息系统中知识和概念的不确定性.

处理知识的不确定性的方法往往用香农(Shannon)信息熵来刻画,知识的粗糙性与信息熵的关系比较密切,知识的粗糙性实质上是其所含信息多少的更深层次的刻画.不少学者结合信息论做了研究工作:运用 Shannon 熵对粗糙集理论中的规则进行度量[128]、基于信息熵的知识约简算法[105-107]、度量粗糙集和粗糙分类的模糊性[129]、不完备系统中的熵度量[130]等.

信息熵和知识粒度从两个不同的角度研究了信息系统的不确定性度量.信息系统的信息熵越大,系统的不确定性越大;而信息系统的知识粒度越大,系统的不确定性越大.所以,结合粒计算来研究不确定性度量正在成为新的研究热点[130-131].

寻求合适的度量来刻画知识的不确定性是粗糙集理论研究的一个重要方向.在粗糙集理论与其它处理模糊性或不确定性方法的理论研究中,主要集中在它与概率统计、模糊数学、D-S 证据理论和信息论等的相互渗透与补充.例如,文献[132]从信息熵的角度提出了一种粗糙集不确定性的模糊度度量方法.

5.3 面向领域的数据驱动的数据挖掘

简而言之,数据挖掘的目的就是从数据中挖掘出知识.在机器学习的许多方法中,我们往往依赖于一些先验知识,比如:贝叶斯概率方法依赖于先验概率;模糊集理论依赖于成员隶属度函数;多专家决策

系统依赖于专家的权值属性. 毫无疑问地，依靠这些先验知识的帮助我们成功地解决了许多问题. 但是，有些领域的先验知识很难获得，比如网络入侵检测；另外，像外太空探索等新兴问题，要获得其先验知识也是很困难的. 因此，如何建立根据问题已有的信息，而不依赖于先验知识获得问题解的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支撑. 为此，有学者提出领域（用户）驱动（user-driven）的数据挖掘模型[133-134]、数据驱动（data-driven）的数据挖掘模型[135-137]等，取得了一些初步研究成果.

在这些工作的基础上，Wang[138]将数据驱动和领域（用户）驱动的思想相结合，提出在保证知识的不变性的前提下，如何研究面向应用领域的数据驱动的数据挖掘理论与技术方法，应该是数据挖掘研究领域下一步的发展方向. 由此提出了面向应用领域的数据驱动的数据挖掘模型，即 3DM（Domain-oriented Data-driven Data Mining）.

考虑到粗糙集理论不需要领域先验知识可以进行学习的特点，一些学者基于粗糙集理论研究了面向应用领域的数据驱动的数据挖掘问题. 比如，通过分析知识的不确定性特征[137]，在基于粗糙集的自主式缺省规则知识获取[135]、数据驱动的自主式决策树学习[136]、数据驱动的自主式粒计算[139]、数据驱动的概念格学习[140]研究中取得了很好的结果，初步证实了 3DM 的可行性和有效性.

Ohsuga[141]认为知识发现就是一个翻译的过程，是一个从无符号到有符号表达方式的转变，并从谓词逻辑学的角度讨论了知识的表达问题. 文献[141]与 3DM 的研究，均说明了数据挖掘是一个知识表达形式的转换过程，即知识从人们难以理解和运用的数据表示形态转换到人类易于理解和运用的知识表示形态，如产生式规则、决策树等. 3DM 更进一步探讨了数据挖掘过程中的各种知识来源，如数据、先验知识、用户兴趣、领域问题约束等，及其在整个知识学习过程中的关系。

在 3DM 的研究工作中，还存在一些待解决的研究问题，如：(1) 领域先验知识的编码模式；(2) 用户兴趣模式和领域制约模式的编码模式；(3) 用户交互控制信息的编码模式；(4) 增量式的面向应用领域的数据驱动的数据挖掘模型等.

5.4　海量数据挖掘

为了使粗糙集应用于海量数据处理，国内外学者从以下几方面进行了研究：

(1) 利用粗糙集理论对海量数据集进行合理分割. 文献[142]结合粗糙集给出了最佳分割的定义.

(2) 结合其它技术，研究增量式学习方法，处理粗糙集理论本身难以实现知识的积累、动态增长的问题. 基于粗糙集的关于属性最小约简的增量式算法、增量式更新概念格算法、增量式知识获取算法、结合决策树的增量式规则获取、增量式属性约简等[143-148]得到了研究.

如何提高增量式学习方法的效率并将其应用到相关领域将是下一步的工作重点.

(3) 粗糙集理论对不完备海量数据的处理，目前的方法基本上都是基于对等价关系的泛化来解决的. 比如容差关系[39]、相似关系[40]、限制容差关系[41]、利用模糊聚类处理不完备信息系统[149]、两步规则提取法处理不完备信息系统[150]. 文献[151]针对不完备信息系统中的“数据补齐法”和“模型扩展法”进行了系统比较分析，提出了所有可能的 8 类扩展模型，详细分析了已有扩展模型的优缺点和相互关系，同时对不满足自反性的几类可能的扩展模型作了讨论和分析.

粗糙集理论在不完备信息系统中的应用，是将其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.

(4) 利用数据库技术使粗糙集理论的相关算法直接处理驻留磁盘的海量数据. 文献[152]提出类分布链表，通过多步生成的方法解决了内存对海量数据限制的问题.

(5) 与快速排序结合，降低海量数据挖掘的处理复杂度. 文献[120,153]提出的新方法能有效提高海量数据处理的效率.

粒计算包含着将复杂问题划分为一系列更容易处理的、更小的子问题，从而降低全局计算代价的思想. 因此，研究基于粗糙集理论的粒计算模型也将是解决此类复杂问题的重要方向之一.

5.5　其它应用领域

从应用的领域来看，基于粗糙集理论的应用除了我们上文提到的信息科学等方面，还遍及其它许多领域. 许多学者将粗糙集理论成功应用到了工业控制、医学卫生及生物科学、交通运输、农业科学、环境科学与环境保护管理、安全科学、社会科学、航空、航天和军事等领域. 比如：电厂气温过热控制[5]、虚拟现实的可视化[6]、对原棉纱线强度和纤维性能之间的知识规则提取[7]、手写体识别[8]、胸部 X 线数字图像滤波增强[9]、湖泊生态系统健康评定指数法

的评价[10]、医疗图像处理[11]、遥感数据处理[12]、综合分类器设计与实现[13]、铁路行车调度指挥[14]、食品安全综合评价[15]、昆虫总科阶元分类[16]、泥石流危险度区划指标选取[17]、网络故障诊断[18]、上市公司违规行为预警[19]、武器系统灰色关联评估[20]和航空控制[21]等等.

6 总结与展望

虽然粗糙集理论从提出至今只有二十几年的发展历史,但取得的研究成果是令人瞩目的.在基于数据的决策与分析、机器学习、模式识别等计算机领域的成功应用,逐渐被人们所重视.

本文首先介绍了该理论的一些基础知识,然后介绍了如何从构造化方法、公理化方法来扩展粗糙集模型以及该理论和证据理论、模糊集、粒计算、形式概念分析、知识空间等的关系.

事实上,关于任何一门新学科的研究工作通常包括理论和应用两方面.

从理论研究来说,关于粗糙集理论的研究工作可以从粗糙集模型扩展来进行.研究者可以从构造化方法、公理化方法采用基于集合的观点或是基于算子的观点来扩展粗糙集理论.我们也可以结合模糊集、证据理论、粒计算、形式概念分析、知识空间等理论来研究粗糙集模型.从而,从理论上推动该学科的发展.

例如,可以从基于元素、基于粒和基于子系统的观念,结合概率论、逻辑学等其它理论来扩展定义粗糙集中的最基本概念——上下近似算子,进而解决相关领域的问题.比如,对等价关系的各种扩展定义就带来了各种处理不完备信息系统的方法.

另一方面,有了各种粗糙集模型,研究者可以在相关模型下进行应用研究.比如,知识约简及规则获取、不确定性信息处理、不完备性信息处理、海量数据挖掘等.

粗糙集理论研究目前正成为信息科学中的一个热点,我们将这个理论目前的研究状况归纳总结并介绍给读者,希望我国更多的感兴趣的研究同行更多地了解这个理论的研究工作,促进这一理论及其相关研究在我国的发展.

参 考 文 献

[1] Pawlak Z. Rough set. International Journal of Computer and Information Sciences, 1982, 11: 341-356

[2] Chan C C, Grzymala Busse J W, Ziarko W P. Rough sets and current trends in computing//Proceedings of the 6th International Conference, RSCTC 2008. Akron, Ohio, USA, 2008

[3] An A, Stefanowski J, Ramanna S, Butz C J, Pedrycz W, Wang G Y et al. Rough sets, fuzzy sets, data mining and granular computing//Proceedings of the 11th International Conference, RSFDGrC 2007. Toronto, Canada, 2007

[4] Wang G Y, Li T R, Grzymala-Busse J, Miao D Q, Skowron A, Yao Y Y. Rough sets and knowledge technology//Proceedings of the RSKT 2008. Chengdu, China, 2008

[5] Xie K M, Chen Z H, Xie G, Lin T Y. BGrC for superheated steam temperature system modeling in power plant//Proceedings of the 2006 IEEE International Conference on Granular Computing. Atlanta, USA, 2006: 708-711

[6] Valdes J J, Romero E, Gonzalez R. Data and knowledge visualization with virtual reality spaces, neural networks and rough sets: Application to geophysical prospecting neural networks//Proceedings of the IJCNN 2007. Orlando, Florida, USA, 2007: 160-165

[7] Ni Yong-Cheng, Yang Jian-Guo, Lv Zhi-Jun. Raw cotton yarn tenacity's rule extraction based on Rough Set theory. Progress in Textile Science & Technology, 2006, 6: 65-66 (in Chinese)
(倪永成,杨建国,吕志军.基于 Rough Set 理论对原棉纱线强度的规则提取.纺织科技进展,2006,6:65-66)

[8] Nguyen T T. Adaptive classifier construction: An approach to handwritten digit recognition//Proceedings of the RSCTC 2002. Malvern, PA, USA, 2002: 578-585

[9] Chen Zhen-Cheng, Zhang Feng, Jiang Da-Zong, Ni Li-Li, Wang Hong-Yan. The filtering method for X-ray digital image of chest based on multi-resolution and rough set. Chinese Journal of Biomedical Engineering, 2004, 23(6): 486-489(in Chinese)
(陈真诚,张锋,蒋大宗,倪利莉,王红艳.利用多分辨率分析的胸部 X 线数字图像粗糙集滤波增强.中国生物医学工程学报,2004,23(6):486-489)

[10] Pang Fa-Hu, Pang Zhen-Ling, Du Rui-Qing. Assessment on rough-set theory for lake ecosystem health index. Journal of Biomathematics, 2008, 23(2): 337-344(in Chinese)
(庞发虎,庞振凌,杜瑞卿.粗糙集理论对湖泊生态系统健康评定指数法的评价.生物数学学报,2008,23(2):337-344)

[11] Hirano S, Tsumoto S. Segmentation of medical images based on approximations in Rough set theory//Proceedings of the RSCTC2002, 2002: 554-563

[12] Hu Sheng-Wu, Guo Zeng-Zhang, Wang Xin-Zhou, Tao Ben-Zao. Fuzzy uncertainty of remote sensing data and the processing method based on Rough set. China Railway Science, 2006, 27(2): 132-136(in Chinese)
(胡圣武,郭增长,王新洲,陶本藻.论遥感数据的模糊不确定性及基于 Rough 集的处理方法.中国铁道科学,2006,27(2):132-136)

[13] Zhu You-Chan, Xiong Wei, Jing Yong-Wen, Gao Ya-Bin. Design and realization of integrated classifier based on Rough Set. Journal on Communications, 2006, 27(z1): 63-67(in Chinese)
(朱有产,熊伟,静永文,高亚彬. 基于 Rough Set 理论的综合分类器设计与实现. 通信学报,2006,27(z1):63-67)

[14] Wang Ming-Hui. Study on the application of rough set in railway dispatching system. China Railway Science, 2004, 25(4): 103-107(in Chinese)
(王明慧. 粗糙集理论在铁路行车调度指挥系统中应用的研究. 中国铁道科学,2004,25(4):103-107)

[15] Li Wei-Xiang, Cheng Ming, Li Ban-Yi. Extended dominance rough set theory's application in food safety evaluation. Food Research and Development, 2008, 29(2): 152-156(in Chinese)
(李为相,程明,李帮义. 粗集理论在食品安全综合评价中的应用. 食品研究与开发,2008,29(2):152-156)

[16] Du Rui-Qing, Chu Xue-Ying, Wang Qing-Lin, Zhao Qiu-Hong, Pang Fa-Hu. Application of a rough-set neural network to superfamily level in insect taxonomy. Journal of China Agricultural University, 2007, 12(1): 33-38(in Chinese)
(杜瑞卿,褚学英,王庆林,赵秋红,庞发虎. 粗糙集神经网络在昆虫总科阶元分类学上的应用. 中国农业大学学报,2007,12(1):33-38)

[17] Kuang Le-Hong, Xu Lin-Rong, Liu Bao-Chen, Yao Jing-Cheng. A new method for choosing zonation indicators of mudflow danger degrees based on the rough set theory. Journal of Geomechanics, 2006, 12(2): 236-242(in Chinese)
(匡乐红,徐林荣,刘宝琛,姚京成. 基于粗糙集原理的泥石流危险度区划指标选取方法. 地质力学学报,2006,12(2):236-242)

[18] Peng Y Q, Liu G Q, Lin T, Geng H S. Application of rough set theory in network fault diagnosis//Proceedings of the Information Technology and Applications, ICITA 2005. Sydney, Australia, 2005, 2: 556-559

[19] Xue Feng, Ke Kong-Ling. The early warning of listed company's illegal behaviors based on the integration of rough set and genetic algorithm. Soft Science, 2008, 22(4): 49-53(in Chinese)
(薛锋,柯孔林. 粗糙集理论和遗传算法集成的上市公司违规行为预警研究. 软科学,2008,22(4):49-53)

[20] Hu Fang, Huang Jian-Guo, Chu Fu-Zhao. Grey relation evaluation model of weapon system based on rough set. Acta Armamentarii, 2008, 29(2): 253-256(in Chinese)
(胡方,黄建国,褚福照. 基于粗糙集的武器系统灰色关联评估模型. 兵工学报,2008,29(2):253-256)

[21] Wojcik Z M. Detecting spots for NASA space programs using rough sets//Proceedings of the 2nd International Conference on Rough Sets and Current Trends in Computing, RSCTC'2000. Canada, 2000: 531-537

[22] Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning About Data. Boston: Kluwer Academic Publishers, 1991

[23] Yao Y Y. A note on definability and approximations//LNCS Transactions on Rough Sets VII, LNCS 4400. Springer-Verlag, 2007: 274-282

[24] Iwinski T B. Algebraic approach to rough sets. Bulletin of the Polish Academy of Sciences Mathematics, 1987, 35: 673-683

[25] Yao Y Y. Two views of the theory of rough sets in finite universes. International Journal of Approximation Reasoning, 1996, 15(4): 291-317

[26] Pawlak Z, Skowron A. Rough membership functions//Zadeh L A, Kacprzyk J eds. Fuzzy Logic for the Management of Uncertainty. New York: John Wiley & Sons, 1994: 251-271

[27] Yao Y Y, Wong S K M. A decision theoretic framework for approximating concepts. International Journal of Man-Machine Studies, 1992, 37: 793-809

[28] Yao Y Y. Generalized rough set models//Polkowski L, Skowron A eds. Rough Sets in Knowledge Discovery. Heidelberg: Physica-Verlag, 1998: 286-318

[29] Pawlak Z. Rough sets and fuzzy sets. Fuzzy Sets and Systems, 1985, 17: 99-102

[30] Yao Y Y. Constructive and algebraic methods of the theory of rough sets. Information Sciences, 1998, 109(1-4): 21-47

[31] Lin T Y, Liu Q. Rough approximate operators: Axiomatic rough set theory//Ziarko W P ed. Rough Sets, Fuzzy Sets and Knowledge Discovery. London: Springer-Verlag, 1994: 256-260

[32] Zhu Feng, He Hua-Can. The axiomatization of the rough set. Chinese Journal of Computers, 2000, 23(3): 330-333(in Chinese)
(祝峰,何华灿. 粗集的公理化. 计算机学报,2000,23(3):330-333)

[33] Sun Hui, Liu Da-You, Li Wen. The minimization of axiom groups of rough set. Chinese Journal of Computers, 2002, 25(2): 201-209(in Chinese)
(孙辉,刘大有,李文. 粗集公理组的极小化. 计算机学报,2002,25(2):201-209)

[34] Wu W Z, Mi J S, Zhang W X. Generalized fuzzy rough sets. Information Sciences, 2003, 151: 263-282

[35] Liu G L, Zhu W. The algebraic structures of generalized rough set theory. Information Sciences, 2008, 178(21): 4105-4113

[36] Zhang Wen-Xiu, Yao Yi-Yu, Liang Yi. Rough Set and Concept Lattice. Xi'an: Xi'an Jiaotong University Press, 2006(in Chinese)
(张文修,姚一豫,梁怡. 粗糙集与概念格. 西安:西安交通大学出版社,2006)

[37] Yao Y Y. On generalizing rough set theory//Proceedings of the 9th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing (RSFDGrC 2003), 2003: 44-51

[38] Yao Y Y. Relational interpretations of neighborhood operators and rough set approximation operators. Information Sciences, 1998, 111(1-4): 239-259

[39] Kryszkiewicz M. Rough set approach to incomplete information systems. Information Sciences, 1998, 112: 39-49

[40] Stefanowski J, Tsoukias A. Incomplete information tables and rough classification. Computational Intelligence, 2001, 17(3): 545-566

[41] Wang Guo-Yin. Extension of Rough set under incomplete information systems. Journal of Computer Research and Development, 2002, 39(10): 1238-1243(in Chinese)
(王国胤. Rough 集理论在不完备信息系统中的扩充. 计算机研究与发展, 2002, 39(10): 1238-1243)

[42] Grzymala-Busse J W. A rough set approach to data with missing attribute values//Proceedings of the 1st International Conference of Rough Sets and Knowledge Technology (RSKT2006). Chongqing, China, 2006: 58-67

[43] Yao Y Y, Lin T Y. Generalization of rough sets using modal logic. Intelligent Automation and Soft Computing, an International Journal, 1996, 2(2): 103-120

[44] Lin T Y, Liu Q, Zuo X L. Models for first order rough logic applications to data mining//Proceedings of the 1996 Asian, Fuzzy Systems Symposium. Soft Computing in Intelligent Systems and Information Processing, 1996: 152-157

[45] Liu Qing. Rough Sets and Rough Reasoning. Beijing: Science Press, 2001(in Chinese)
(刘清. Rough 集及 Rough 推理. 北京: 科学出版社, 2001)

[46] Zakowski W. Approximation in the space(u,π). Demonstratio Mathematica, 1983, 16: 761-769

[47] Bonikowski Z et al. Extensions and intentions in the rough set theory. Information Sciences, 1998, 107: 149-167

[48] Zhu W, Wang F Y. Reduction and axiomization of covering generalized rough sets. Information Sciences, 2003, 152: 217-230

[49] Tsang E C C, Chen D G, Lee J W T et al. On the upper approximations of covering generalized rough sets//Proceedings of the 3rd International Conference Machine Learning and Cybernetics. Shanghai, China, 2004: 4200-4203

[50] Zhu W. Topological approaches to covering rough sets. Information Sciences, 2007, 177: 1499-1508

[51] Wang F Y. On the abstraction of conventional dynamic systems: From numerical analysis to linguistic analysis. Information Sciences, 2005, 171(1-3): 233-259

[52] Zhu W, Clark Thomborson, Wang F Y. A survey of software watermarking//Proceedings of the IEEE ISI 2005. Atlanta, USA, 2005: 454-458

[53] Yao Y Y. On generalizing Pawlak approximation operators//Proceedings of the 1st International Conference on Rough Sets and Current Trends in Computing (RSCTC'98). Warsaw, Poland, 1998: 298-307

[54] Chen D G, Zhang W X, Yeung D, Tsang E C C. Rough approximations on a complete completely distributive lattice with applications to generalized rough sets. Information Sciences, 2006, 176: 1829-1848

[55] Järvinen J. On the structure of rough approximations. Fundamenta Informaticae, 2002, 53: 135-153

[56] Jaoua A, Elloumi S. Galois connection, formal concept and Galois lattice in real binary relation. Systems & Software, 2002, 60(2): 149-163

[57] Harms S K, Deogum J S. Sequential association rule mining with time lags. Journal of Intelligent Information Systems, 2004, 22(1): 7-22

[58] Yao Y Y, Wong S K M, Lin T Y. A review of rough set models//Lin T Y, Cercone N eds. Rough Sets and Data Mining: Analysis for Imprecise Data. Boston: Kluwer Academic Publishers, 1997: 47-75

[59] Pei D W, Xu Z B. Rough set models on two universities. International Journal of General Systems, 2004, 33: 569-581

[60] Mi J S, Zhang W X. An axiomatic characterization of a fuzzy generalization of rough sets. Information Sciences, 2004, 160: 235-249

[61] Yao Y Y. Decision-theoretic rough set models//Proceedings of the 2nd International Conference on Rough Sets and Knowledge Technology (RSKT 2007). Toronto, Canada, 2007: 1-12

[62] Yao J T, Zhang M. Feature selection with adjustable criteria//LNCS Transactions on Rough sets VII, LNAI 3641. Springer-Verlag, 2005: 204-213

[63] Wong S K M, Ziarko W. Comparison of the probabilistic approximate classification and the fuzzy set model. Fuzzy Sets and Systems, 1987, 21: 357-362

[64] Zhao Wen-Qing, Zhu Yong-Li, Gao Wei. An information filtering model based on decision-theoretic rough set theory. Computer Engineering and Applications, 2007, 43(7): 185-187(in Chinese)
(赵文清, 朱永利, 高伟. 一个基于决策粗糙集理论的信息过滤模型. 计算机工程与应用, 2007, 43(7): 185-187)

[65] Yao J T, Herbert J P. Web-based support systems with rough set analysis//Proceedings of the Rough Sets and Emerging Intelligent Systems Paradigms (RSEISP'07). Warsaw, Poland, 2007: 360-370

[66] Pawlak Z, Skowron A. Rough sets: Some extensions. Information Sciences, 2007, 177: 28-40

[67] Greco S, Matarazzo B, Slowinski R. Rough membership and Bayesian confirmation measures for parameterized rough sets//LNCS Transactions on Rough sets VII, LNAI 3641. Springer-Verlag, 2005: 314-324

[68] Slezak D. Rough sets and Bayes factor//LNCS Transactions on Rough sets VII, LNAI 3400. Springer-Verlag, 2005: 202-229

[69] Slezak D, Ziarko W. Attribute reduction in the Bayesian version of variable precision rough set model. Electronic Notes in Theoretical Computer Science, 2003, 82: 263-273

[70] Fagin R, Halpern J Y. Uncertainty, belief, and probability. Computational Intelligence, 1991, 7: 160-173

[71] Skowron A. The relationship between the rough set theory and evidence theory. Bulletin of the Polish Academy of Sciences Mathematics, 1989, 37: 87-90

[72] Yao Y Y, Lingras P J. Interpretations of belief functions in the theory of rough sets. Information Sciences, 1998, 104(1-2): 81-106

[73] Wu W Z, Leung Y, Zhang W X. Connections between rough set theory and Dempster-Shafer theory of evidence. International Journal of General Systems, 2002, 31(4): 405-430

[74] Dubois D, Prade H. Rough fuzzy set and fuzzy rough sets. International Journal of General Systems, 1990, 17: 191-209

[75] Yao Y Y. Combination of rough and fuzzy sets based on α-level sets//Rough Sets and Data Mining: Analysis for Imprecise Data. Kluwer Academic Publishers, 1997: 301-321

[76] Srinivasan P, Ruiz M, Kraft D H, Chen J H. Vocabulary mining for information retrieval: Rough sets and fuzzy sets. Information Processing and Management, 2001, 37: 15-38

[77] Jensen R, Shen Q. Fuzzy-rough attribute reduction with application to Web categorization. Fuzzy Sets and Systems, 2004, 141: 469-485

[78] Dubois D. On ignorance and contradiction considered as truth-values. Logic Journal of the IGPL, 2008, 16(2): 195-216

[79] Intan R, Mukaidono M. Generalized fuzzy rough sets by conditional probability relations. International Journal of Pattern Recognition and Artificial Intelligence, 2002, 16(7): 865-881

[80] Wang Y F. Mining stock price using fuzzy rough set system. Expert System with Applications, 2003, 24: 13-23

[81] Kasemsiri W, Kimpan C. Printed thai character recognition using fuzzy-rough sets//Proceedings of the IEEE Region 10 International Conference on Electrical and Electronic Technology (volume 1), 2001: 326-330

[82] Yao Y Y, Chen Y H. Rough set approximations in formal concept analysis. Journal of Transactions on Rough Sets, V, LNCS 4100, 2006: 285-305

[83] Yao Y Y. A comparative study of formal concept analysis and rough set theory in data analysis//Proceedings of the 4th International Conference on Rough Sets and Current Trends in Computing (RSCTC 2004). Uppsala, Sweden, 2004: 59-68

[84] Wolski M. Galois connections and data analysis. Fundamenta Informaticae, 2004, 60: 401-415

[85] Li H R, Zhang W X, Wang H. Classification and reduction of attributes in concept lattices//Proceedings of the 2006 IEEE International Conference on Granular Computing (GrC'06). Georgia State University, Atlanta, USA, 2006: 142-147

[86] Yao Y Y. Concept lattices in rough set theory//Proceedings of the 23rd International Meeting of the North American Fuzzy Information Processing Society (NAFIPS 2004). Banff, Alberta, Canada, 2004: 796-801

[87] Ma J M, Zhang W X, Wang X. Dependence space of concept lattice based on rough set//Proceedings the 2006 IEEE International Conference on Granular Computing (GrC'06). Georgia State University, Atlanta, USA, 2006: 200-204

[88] Doignon J P, Falmagne J C. Spaces for the assessment of knowledge. International Journal of Man-Machine Studies, 1985, 23: 175-196

[89] Duntsch I, Gediga G. A note on the correspondences among entail relations, rough set dependencies, and logical consequence. Mathematical Psychology, 2001, 43: 393-401

[90] Xu F F, Yao Y Y, Miao D Q. Rough set approximations in formal concept analysis and knowledge spaces//Proceedings of the 17th International Symposium on Methodologies for Intelligent Systems (ISMIS'2008), 2008: 319-328

[91] Miao Duo-Qian, Wang Guo-Yin, Liu Qing, Lin Zao-Yang, Yao Yi-Yu. Granular Computing: Past, Present, Future. Science Press, 2007(in Chinese)
(苗夺谦,王国胤,刘清,林早阳,姚一豫. 粒计算:过去、现在与展望. 北京:科学出版社,2007)

[92] Wang Guo-Yin, Zhang Qing-Hua, Hu Jun. An overview of granular computing. CAAI Transactions on Intelligent Systems, 2007, 2(6): 8-26(in Chinese)
(王国胤,张清华,胡军. 粒计算研究综述. 智能系统学报,2007,2(6):8-26)

[93] Polkowski L, Skowron A. Rough-Neuro computing//Proceedings of the 2nd International Conference on Rough Sets and Current Trends in Computing (RSCTC'2000). Banff, Canada, 2005: 25-32

[94] Peters J F, Szczuka M S. Rough neurocomputing: A survey of basic models of neurocomputation//Proceedings of the 3rd International Conference on RSCTC'2002. Malvern, PA, USA, 2002: 308-315

[95] Yao J T, Yao Y Y. Induction of classification rules by granular computing//Proceedings of the 3rd International Conference on RSCTC'2002. Malvern, PA, USA, Springer, 2002: 331-338

[96] Yao Y Y. A partition model of granular computing. LNCS Transactions on Rough Sets, 2004, 1: 232-253

[97] Zhu W, Wang F Y. Reduction and axiomization of covering generalized rough sets. Information Sciences, 2003, 152: 217-230

[98] Hu J, Wang G Y, Zhang Q H. Uncertainty measure of covering generated rough set//Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IATW'06), 2006: 498-504

[99] Ma J M, Zhang W X, Li T J. A covering model of granular computing//Proceeding of the 4th International Conference on Machine Learning and Cybernetics (ICMLC 2005). Guangzhou, China, 2005: 1625-1630

[100] Wang G Y, Hu F, Huang H, Wu Y. A granular computing model based on tolerance relation. The Journal of China Universities of Posts and Telecommunications, 2005, 12(3): 86-90

[101] Hu Feng, Huang Hai, Wang Guo-Yin, Wu Yu. Granular computing in incomplete systems. Journal of Chinese Computer Systems, 2005, 26(8): 1335-1339(in Chinese)
(胡峰，黄海，王国胤，吴渝. 不完备信息系统的粒计算方法. 小型微型计算机系统，2005，26(8)：1335-1339)

[102] Wang Guo-Yin. Rough Set Theory and Knowledge Discovery. Xi'an: Xi'an Jiaotong University Press, 2001(in Chinese)
(王国胤. Rough 集理论与知识获取. 西安：西安交通大学出版社，2001)

[103] Chen C Y, Li Z G, Qiao A Y, Wen A P. Study on discretization in rough set based on genetic algorithm//Proceedings of the 2nd International Conference on Machine Learning and Cybernetics. Xi'an, 2003: 1430-1434

[104] Hou Li-Juan, Wang Guo-Yin, Wu Yu, Nie Neng. Discretization in rough set theory. Computer Science, 2000, 27(12): 89-94(in Chinese)
(侯利娟，王国胤，吴渝，聂能. 粗糙集理论中的离散化问题. 计算机科学，2000，27(12)：89-94)

[105] Wang Gong-Yin, Yu Hong, Yang Da-Chun. Decision table reduction based on conditional information entropy. Chinese Journal of Computers, 2002, 25(7): 759-766(in Chinese)
(王国胤，于洪，杨大春. 基于条件信息熵的决策表约简. 计算机学报，2002，25(7)：759-766)

[106] Wang Jue, Wang Ren, Miao Duo-Qian et al. Data enriching based on Rough Set theory. Chinese Journal of Computers, 1998, 21(5): 393-400(in Chinese)
(王珏，王任，苗夺谦等. 基于 Rough Set 理论的"数据浓缩". 计算机学报，1998，21(5)：393-400)

[107] Yu H, Wang G Y, Yang D C, Wu Z F. Knowledge reduction algorithms based on rough set and conditional information entropy//Proceedings of SPIE: Data Mining and Knowledge Discovery: Theory, Tool, and Technology IV. Orlando, USA, 2002, 4730: 422-431

[108] Wei L, Li H R, Zhang W X. Knowledge reduction based on the equivalence relations defined on attribute set and its power set. Information Sciences, 2007, 177: 3178-3185

[109] Hu Q H, Zhao H, Xie Z X, Yu D R. Consistency based attribute reduction//Proceedings of the PAKDD 2007. Nanjing, China, 2007: 96-107

[110] Kryzkiewicz M. Comparative study of alternative types of knowledge reduction in inconsistent systems. International Journal of Intelligent Systems, 2001, 16: 105-120

[111] Deng D Y, Huang H K, Li X J. Comparison of various types of reductions in inconsistent systems. Acta Electronica Sinica, 2007, 35: 252-255

[112] Zhang Wen-Xiu, Mi Ju-Sheng, Wu Wei-Zhi. Knowledge reductions in inconsistent information systems. Chinese Journal of Computers, 2003, 26(1): 12-18(in Chinese)
(张文修，米据生，吴伟志. 不协调目标信息系统的知识约简. 计算机学报，2003，26(1)：12-18)

[113] Slezak D. Approximate entropy reducts. Fundamenta Informatica, 2002, 53: 365-390

[114] Chen D G, Wang C Z, Hu Q H. A new approach to attribute reduction of consistent and inconsistent covering decision systems with covering rough sets. Information Sciences, 2007, 177: 3500-3518

[115] Wang G Y, Zhao J, An J J, Wu Y. A comparative study of algebra viewpoint and information viewpoint in attribute reduction. Fundamenta Informaticae, 2005, 68(3): 289-301

[116] Wang Guo-Yin. Calculation methods for core attributes of decision table. Chinese Journal of Computers, 2003, 26(5): 611-615(in Chinese)
(王国胤. 决策表核属性的计算方法. 计算机学报，2003，26(5)：611-615)

[117] Zhang Wen-Xiu, Wei Ling, Qi Jian-Jun. Attribute reduction based on concept lattice. Science in China (Series E), 2005, 35(6): 628-639(in Chinese)
(张文修，魏玲，祁建军. 概念格的属性约简理论与方法. 中国科学 E 辑：信息科学，2005，35(6)：628-639)

[118] Harms S K, Depgum J S. Sequential association rule mining with time lags. Journal of Intelligent Information Systems, 2004, 22(1): 7-22

[119] Liang H, Wang J, Yao Y Y. User-oriented feature selection for machine learning. The Computer Journal, 2007, 50(4): 421-434

[120] Hu Feng, Wang Guo-Yin. Quick reduction algorithm based on attribute order. Chinese Journal of Computers, 2007, 30(8): 1429-1435(in Chinese)
(胡峰，王国胤. 属性序下的快速约简算法. 计算机学报，2007，30(8)：1429-1435)

[121] Jiang Yuan-Chun, Liu Ye-Zheng. An attribute reduction method based on ant colony optimization//Proceedings of the 6th World Congress on Intelligent Control and Automation (WCICA 2006). Dalian, China, 2006: 3542-3546(in Chinese)
(姜元春，刘业政. 基于蚁群优化算法的属性约简方法//第 6 届全球智能控制与自动化大会. 中国，大连，2006：3542-3546)

[122] Ye D Y, Chen Z J, Liao J K. A new algorithm for minimum attribute reduction based on binary particle swarm optimization with vaccination//Proceedings of the PAKDD07, Nanjing, China, 2007: 1029-1036

[123] Xu C, Min F. Weighted reduction for decision tables//Proceedings of the FSKD. Xi'an, China, 2006: 246-255

[124] Hu J, Wang G Y, Fu A. Knowledge reduction of covering approximation space//Proceedings of the 6th IEEE International Conference on Cognitive Informatics (ICCI'07). Lake Tahoe, California, USA, 2007: 140-144

[125] Yao Y Y, Zhao Y. Attribute reduction in decision-theoretic rough set models. Information Sciences, 2008, 178(17): 3356-3373

[126] Yao Y Y, Zhao Y, Wang J. On reduct construction algorithms//Proceedings of the 1st International Conference on Rough Sets and Knowledge Technology (RSKT'06). Chongqing, China, 2006: 297-304

[127] Yu H, Wang G Y, Lan F K. Solving the attribute reduction problem with ant colony optimization//Chan C C, Grzymala-Busse J W, Ziarko W P eds. Proceedings of the 6th International Conference on Rough Sets and Current Trends in Computing (RSCTC 2008). Akron, Ohio, USA, 2008: 242-251

[128] Düntsch I, Gediga G. Uncertainty measure of rough set prediction. Artificial Intelligence, 1998, 106: 109-137

[129] Liang Ji-Ye, Li De-Yu. Uncertainty and Knowledge Acquisition in Information System. Beijing: Science Press, 2005 (in Chinese)
(梁吉业,李德玉. 信息系统中的不确定性与知识获取. 北京:科学出版社,2005)

[130] Liang J Y, Shi Z Z, Li D Y, Wireman M J. The information entropy, rough entropy and knowledge granulation in incomplete information systems. International Journal of General Systems, 2006, 34(1): 641-654

[131] Zhang Ling, Zhang Bo. Theory of fuzzy quotient space (methods of fuzzy granular computing). Journal of Software, 2003, 14(4): 770-776(in Chinese)
(张铃,张钹. 模糊商空间理论(模糊粒度计算方法). 软件学报,2003,14(4):770-776)

[132] Wang Guo-Yin, Zhang Qing-Hua. Uncertainty of rough sets in different knowledge granularities. Chinese Journal of Computers, 2008, 31(9): 1588-1598(in Chinese)
(王国胤,张清华. 不同知识粒度下粗糙集的不确定性研究 计算机学报,2008,31(9):1588-1598)

[133] Yao Y Y, Zhao Y, Wang J, Han S. A model of machine learning based on user preference of attributes//Proceedings of the 5th International Conference on Rough Sets and Current Trends in Computing (RSCTC 2006). Kobe, Japan, 2006: 587-596

[134] Kuntz P, Guillet F, Lehn R, Briand H. A user-driven process for mining association rules//Zighed A D A, Komorowski J, Zytkow J M eds. PKDD 2000. LNCS (LNAI) 1910. Springer, Heidelberg, 2000: 483-489

[135] Wang Guo-Ying, He Xiao. A self-learning model under uncertain condition. Journal of Software, 2003, 14(6): 1096-1102(in Chinese)
(王国胤,何晓. 一种不确定性条件下的自主式知识学习模型. 软件学报,2003,14(6):1096-1102)

[136] Yin D S, Wang G Y, Wu Y. Data-driven decision tree learning algorithm based on rough set theory//Proceedings of the 2005 International Conference on Active Media Technology (AMT2005). Takamatsu, Kagawa, Japan, 2005: 579-584

[137] Zhao J, Wang G Y. A data-driven knowledge acquisition method based on system uncertainty//Proceedings of the 4th IEEE International Conference on Cognitive Informatics. Irvine, USA, 2005: 267-275

[138] Wang G Y. Domain-oriented data-driven data mining (3DM): Simulation of human knowledge understanding//Proceedings of the WImBI 2006. Beijing, China, 2006: 278-290

[139] Gan Q, Wang G, Hu J. A self-learning model based on granular computing//Proceedings of the 2006 IEEE International Conference on Granular Computing(IEEE GrC2006). Atlanta, USA, 2006: 530-533

[140] Wang Yan, Wang Guo-Yin, Deng Wei-Bin. Data-driven knowledge acquisition based on concept lattice. Pattern Recognition and Artificial Intelligence, 2007, 20(5): 636-642 (in Chinese)
(王燕,王国胤,邓维斌. 基于概念格的数据自主式不确定知识获取. 模式识别与人工智能,2007,20(5):636-642)

[141] Ohsuga S. Knowledge discovery as translation//Studies in Computational Intelligence(SCI), 2005, 6: 3-19

[142] Qin Zhen-Ren. Study on data mining based on Rough Set theory[M. S. dissertation]. Chongqing University of Posts and Telecommunications, 2004(in Chinese)
(覃政仁. 基于 Rough Set 的海量数据挖掘算法研究. 重庆邮电大学,重庆,2004)

[143] Liu Zong-Tian. An incremental arithmetic for the smallest reduction of attributes. Acta Electronica Sinica, 1999, 27 (11): 96-98(in Chinese)
(刘宗田. 属性最小约简的增量式算法. 电子学报,1999,27(11):96-98)

[144] Godin R, Missaoui R, Alaui H. Incremental concept formation algorithms based on galoies (Concept) lattics. Computational Intelligence, 1995, 11(2): 246-267

[145] Wang Zhi-Hai, Hu Ke-Yue et al. General and incremental algorithms of rule extraction based on concept lattice. Chinese Journal of Computers, 1999, 22(1): 66-70(in Chinese)
(王志海,胡可云等. 概念格上规则提取的一般算法与渐进式算法. 计算机学报,1999,22(1):66-70)

[146] Zheng Z, Wang G Y. RRIA: A rough set and rule tree based incremental knowledge acquisition algorithm. Fundamenta Informaticae, 2004, 59(2-3): 299-313

[147] Hu F, Wang G Y, Huang H, Wu Y. Incremental attribute reduction based on elementary sets//Proceedings of the RSFDGrC2005. Regina, Canada. Part 1, 2005: 185-193

[148] Yu H, Yang D C. An incremental rule acquisition algorithm based on rough set. The Journal of China University of Posts and Telecommunications, 2005, 12(1): 47-52

[149] Zhang Q H, Wang G Y, Hu J, Liu X Q. Incomplete information systems processing based on fuzzy-clustering//Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT 2006 Workshops)(WI-IATW'06). Hong Kong, China, 2006: 486-489

[150] Li H X, Yao Y Y, Zhou X Z, Huang B. Two-phase rule induction from incomplete data//Wang G et al. eds. RSKT 2008. Chengdu, China. LNAI 5009. Berlin Heidelberg: Springer-Verlag, 2008: 47-54

[151] Wang G Y, Guan L H, Hu F. Rough set extensions in incomplete information systems. Frontiers of Electrical and Electronic Engineering in China, 2008, 3(4): 399-405

[152] Qin Z R, Wang G Y, Wu Y, Xue X R. A scalable rough set knowledge reduction algorithm//Rough Sets and Current Trends in Computing. LNAI 3006. Springer-Verlag, 2004: 445-454

[153] Liu Shao-Hui, Sheng Qiu-Jian, Wu Bin, Shi Zhong-Zhi, Hu Fei. Research on efficient algorithms for Rough set methods. Chinese Journal of Computers, 2003, 26(5): 524-529(in Chinese)
(刘少辉，盛秋戬，吴斌，史忠植，胡斐. Rough 集高效算法的研究. 计算机学报，2003，26(5)：524-529)

WANG Guo-Yin, born in 1970, Ph. D., professor. His research interests include rough set, granular computing, neural networks, machine learning, data mining, knowledge technology and cognitive computing, etc.

YAO Yi-Yu, born in 1962, Ph. D., professor. His research interests include information retrieval, granular computing, rough sets, fuzzy sets, data mining, Web intelligence, and intelligent information systems, etc.

YU Hong, born in 1972, Ph. D., associate professor. Her research interests include Rough set, intelligence information processing and Web intelligence, etc.

Background

The advances of computer technology have led to an unprecedented power in data collection, processing, management and storage. A grand challenge is how to transform data into information, information into knowledge, and knowledge into wisdom. For this purpose, many theories, models, methods and tools have been proposed and extensively studied. Rough set theory, proposed in the early 1980s, is a mathematical theory and tool for dealing with imprecise, incomplete, uncertain, and vague information. It has been applied successfully in many fields such as decision analysis, machine learning, data mining, knowledge discovery, pattern recognition, etc.

This paper provides a review of the main theoretical developments and applications of rough set theory. The authors introduce a framework for a systematic study of the rough set theory, discuss various views of the theory and different approaches to study the theory. The authors summarize the relationships between the rough set theory and other theories of data analysis, such as fuzzy sets, evidence theory, granular computing, formal concept analysis, knowledge spaces, etc, and point out possible future research and development trends in rough set theory.

This work is partially supported by the National Natural Science Foundation of China (Nos. 60573068 and 60773113), Natural Science Foundation of Chongqing of China (Nos. 2008BA2017 and 2008BA2041), Science & Technology Research Program of the Municipal Education Committee of Chongqing of China (No. KJ080510).