粒计算的集合论描述

苗夺谦[1)]　徐菲菲[2)]　姚一豫[3)]　魏　莱[4)]

[1)](同济大学计算机科学与技术系　上海　201804)

[2)](上海电力学院计算机科学与技术系　上海　200090)

[3)](里贾纳大学计算机科学系　里贾纳　S4S0A2　加拿大)

[4)](上海海事大学计算机科学系　上海　200135)

摘　要　粒计算的形式化研究一直没有被仔细讨论.文中在集合论框架下,对粒计算做了系统研究,给出了粒度空间的三层模型(论域,基,粒结构).借用逻辑语言L判定粒的可定义性,将经典粗糙集通过此模型重新解释.根据模型中从基到粒结构不同的构造规则,引出并可约和交可约粒度空间的定义,分别讨论了不同粒度空间下覆盖、基和粒结构的关系,从而给出从覆盖求基的方法;进一步,利用子系统表示方法对扩展粗糙集以及一般的交可约与并可约空间的上下近似进行了研究,分析了现有的4种基于覆盖的粗糙集模型的合理性;研究了形式概念分析以及知识空间的粒度空间模型,给出这两种理论中上下近似的概念.

关键词　粒计算;粒度空间;粗糙集;形式概念分析;知识空间

中图法分类号　TP18　　**DOI号**:10.3724/SP.J.1016.2012.00351

Set-Theoretic Formulation of Granular Computing

MIAO Duo-Qian[1)]　XU Fei-Fei[2)]　YAO Yi-Yu[3)]　WEI Lai[4)]

[1)](*Department of Computer Science and Technology*, *Tongji University*, *Shanghai*　201804)

[2)](*Department of Computer Science and Technology*, *Shanghai University of Electric Power*, *Shanghai*　200090)

[3)](*Department of Computer Science*, *University of Regina*, *Regina*　S4S0A2, *Canada*)

[4)](*Department of Computer Science*, *Shanghai Maritime University*, *Shanghai*　200135)

Abstract　Granular computing has been widely studied in recent years. However, it still lacks of a formal framework which can cover all the affiliate models of granular computing, such as rough sets, formal concept analysis and knowledge spaces. In this paper, a three levels' model of granular spaces (the universe, the basis and the granular structure) in set-theoretic formulation is proposed. By using the definability defined by the logic language L, a three levels' model of granular spaces in Pawlak rough sets is established. In terms of the different construction rules from the basis to the granular structure, the definitions of the union reducible granular spaces and the intersection reducible granular spaces are given. We discuss the relationship among the covering, the basis, and the corresponding granular structure, and give the methods for seeking the basis from the covering in the two types of granular spaces respectively. Then we focus on set-theoretic setting and generalized rough set approximations with subsystem-based definition. The generalized approximations are also investigated in the union reducible and the intersection reducible granular spaces. Furthermore, the rationality of the four types of covering-based rough set models is examined. Finally, the three levels' models of granular spaces in formal concept analy-

收稿日期:2009-08-04;最终修改稿收到日期:2011-10-10.本课题得到国家自然科学基金(60970061)、上海电力学院人才引进基金(K2011-002)、上海市教育委员会科研创新项目(Z2011-080)、上海市优秀青年教师基金(B211058K)资助.**苗夺谦**,男,1964年生,教授,博士生导师,主要研究领域为粗糙集理论、模式识别、人工智能等.**徐菲菲**,女,1983年生,博士,讲师,主要研究方向为模糊粗糙集理论、数据挖掘、人工智能等.E-mail:xufeifei1983@hotmail.com.**姚一豫**,男,1962年生,教授,博士生导师,主要研究领域为粒计算、粗糙集、网络智能等.**魏　莱**,男,1980年生,博士,讲师,主要研究方向为流形学习、机器学习等.

sis and knowledge spaces are expressed. The generalized approximations are also explored in the two theories. It is shown that the set-framework of granular computing can help us to study the branch models of granular computing in a more formal view.

Keywords granular computing; granular spaces; rough sets; formal concept analysis; knowledge spaces

1 引 言

近年来粒计算[1]成为人工智能研究中的一个新热点,已引起了众多学者的关注[1-6].粒计算研究横跨了多门学科,具有多个分支模型,包括粗糙集理论[7]、形式概念分析[8-9]、知识空间[10-12]等.这些诸多的理论工具暗示了粒计算的广泛应用前景,但也恰恰说明了粒计算研究本身所具有的复杂性.事实上,尽管学者们就粒计算的基本理论和方法做了大量工作,但目前为止仍然缺乏一个统一的理论背景来分析和讨论上述多个分支模型,这显然不利于粒计算的进一步发展,在一定程度上也限制了粒计算的应用范围.

我们在对粗糙集、形式概念分析和知识空间的研究中发现,3 种理论都是建立在对"集合"的分析基础上的.研究对象都包括一个感兴趣的对象集合,即论域,一些具有背景意义的对象子集,以及由这些对象子集形成的集族.因此,我们尝试建立一个粒计算的集合论模型,在这一统一模型下,粗糙集、形式概念分析、知识空间可以看作是该统一模型在不同的诱导条件下的特例.

为了完成上述目的,我们从粒计算中的基本概念——粒的可定义性出发,给出粒结构的三层模型.然后用其来重新表示 Pawlak 经典粗糙集[7].继而根据该模型从基到粒结构不同的构造规则,做了一般性的推广,引出并可约粒度空间和交可约粒度空间的定义.在这两种粒度空间下,分别讨论了覆盖、基和对应的粒结构的关系,并给出了从覆盖求基的方法.然后,我们利用子系统的表示方法对扩展的粗糙集以及一般的交可约与并可约空间的上下近似进行了研究,并分析了 William Zhu 提出的 4 种基于覆盖的粗糙集模型[13-16]的合理性;最后,我们研究了形式概念分析以及知识空间的粒度空间模型,并给出了这两种理论中上下近似的概念.

2 粒度空间三层模型

在集合论框架下,一个集合可以看作一个粒.假设 U 是一个有限非空集,称为论域.所有在 U 上定义的粒可被解释为 2^U 的子集(2^U 是 U 的幂集).我们定义一个三元组

$$(U,S_0,S) \tag{1}$$

其中,U 是论域,$S_0\subseteq 2^U$ 是基本粒,$S\subseteq 2^U$ 是可定义粒族.S_0 是基本集族,即基本粒(也叫基).可定义粒族 S 是基本集的超集族或者子集族,称为粒结构,S 的子集即为可定义集.三元组(U,S_0,S)就称为粒度空间的三层模型.

如果 $S_0=\{\{x\}\mid x\in U\}$,$S_0=2^U$,那么模型是平凡的,因为它不能提供任何除论域外任何有用信息.因此为了让模型(1)有意义,我们假设一个基于集合论的粒计算三层模型,满足 3 个最基本要求:

(1) 每个所考虑的粒必须有一个明确的解释;

(2) 存在一些基本粒,其它粒可由基本粒构建;

(3) 所有粒可形成一个非平凡的层次结构.

换句话说,即要求粒和粒结构都要有意义.否则,我们从粒化过程中并不能获得有用的知识.当一个粒是可定义粒时,那么我们认为这个粒是有意义的.文献[17]中指出,粒是由一定的关系或谓词、函数产生的,但在这里我们先不确定该三层模型上所定义的特定关系,只给出一个粒度空间模型所应该满足的基本要求,通过赋予这个模型特定的关系,可以得到不同的粒度空间描述.

下面,我们将从粒的可定义性出发,讨论 Pawlak 粒结构.

3 Pawlak 粒结构

波兰数学家 Pawlak 提出的粗糙集理论[7]是粒计算研究中采用最为广泛的模型之一.在粗糙集理论中,等价类可以视为一个基本粒.任意等价类的并得到的类也可看成是一个粒.通过等价类的并运算,我们可以构建一个对交、并、补封闭的代数系统,称为 σ-代数.

一个粒是一个可定义集.粗糙集理论通过定义上下近似来描述一个不可定义集.这里,我们首先介绍可定义性的概念.然后,我们引出上下近似的定义.

需要说明的是，我们不再用等价关系或等价类定义上下近似，而是从可定义集导出粗糙集模型. 因此虽然形式上和 Pawlak 的上下近似不同，但在粗糙集理论框架下是等价的.

3.1　粒的可定义性

一个概念可被理解为由两部分组成：内涵和外延[18]. 内涵反映的是内在性质或者一些对象的共同属性. 概念的外延是能为这个概念提供具体实例的对象集. 为了形式化定义概念，Yao 采用 Orlowska[18] 和 Pawlak[7] 在分析信息表时所用的决策逻辑语言 L 来描述[19]. 这里的信息表可以描述为

$$M=(U,At,\{V_a \mid a\in At\},\{I_a \mid a\in At\}) \tag{2}$$

其中 U 是对象的有限非空集，At 是属性的有限非空集，V_a 是属性 $a\in At$ 的非空值集，$I_a:U\rightarrow V_a$ 是一个信息函数.

L 中的公式可由一组对应于一些基本概念的原子公式得到. 一个原子公式由 $a=v$ 表示，其中 $a\in At$，$v\in V_a$. 对每个原子公式 $(a=v)$，如果对象 x 满足 $I_a(x)=v$，记 $x\models(a=v)$. 如果不满足 $(a=v)$ 则写为 $\neg x\models(a=v)$. 给定原子公式，其它的公式可通过逻辑连接运算 $\neg$，$\wedge$，$\vee$，$\rightarrow$ 和 $\leftrightarrow$ 得到.

如果 ϕ 是一个公式，集合 $m(\phi)$ 定义为

$$m(\phi)=\{x\in U \mid x\models\phi\} \tag{3}$$

$m(\phi)$ 被称为公式 ϕ 在信息表 M 的涵义集. 公式 ϕ 的涵义集实际上是由公式 ϕ 表示的具有相同性质的所有对象集. 换句话说，ϕ 可被看作对象集 $m(\phi)$ 的描述. 因此，信息表 M 中的一个概念可表达为 $(\phi,m(\phi))$，其中 $\phi\in L$. ϕ 是一个概念的内涵，$m(\phi)$ 是概念的外延. 这样公式和 U 上的子集就建立起了联系. $m(\phi)$ 满足下述性质[7]：

(1) $m(\neg\phi)=(m(\phi))^c$，

(2) $m(\phi\wedge\psi)=m(\phi)\cap m(\psi)$，

(3) $m(\phi\vee\psi)=m(\phi)\cup m(\psi)$，

(4) $m(\phi\rightarrow\psi)=(m(\phi))^c\cup m(\psi)$，

(5) $m(\phi\equiv\psi)=(m(\phi)\cap m(\psi))\cup((m(\phi))^c\cap(m(\psi))^c)$，其中 $(m(\phi))^c=U-m(\phi)$ 是 $m(\phi)$ 的补. 在这样的数学表述中，我们可以在逻辑环境下讨论粒的内涵，同时在集合论下讨论粒的外延.

给定属性集 At 上的子集 $A\subseteq At$. $L(A)$ 是定义在 A 上的语言. 一个对象子集的可定义性可描述如下[3].

定义 1.　子集 $X\subseteq U$ 在信息表 $M=(U,At,\{V_a \mid a\in At\},\{I_a \mid a\in At\})$ 和属性集 $A\subseteq At$ 下是可定义的当且仅当在逻辑语言 $L(A)$ 中存在一个公式 ϕ 满足

$$X=m(\phi) \tag{4}$$

否则，它是不可定义的.

即一个集合是可定义的如果能找到一个逻辑公式能定义这个集合的元素. 一个可定义的粒由集对 $(\phi,m(\phi))$ 表示. 根据这个定义，所有可定义集组成的集族定义为

$$Def(L(A))=\{m(\phi) \mid \phi\in L(A)\} \tag{5}$$

3.2　Pawlak 粒结构

有了上述定义，我们可以在三元组 (U,S_0,S) 的基础上构造一个有意义的结构. 描述为

$$(U,\{m(\wedge_{a\in At}\, a=I_a(x))\},Def(L(A))) \tag{6}$$

其中，U 是论域，$m(\wedge_{a\in At}a=I_a(x))$ 是基本可定义粒，$Def(L(A))$ 是由基本可定义粒得到的可定义粒族.

我们称三元组 $(U,\{m(\wedge_{a\in At}a=I_a(x))\},Def(L(A)))$ 为 Pawlak 粒度空间. 根据 Pawlak 粗糙集模型，可定义粒族 $Def(L(A))$ 具有以下性质：

(1) $\varnothing,U\in Def(L(A))$；

(2) $X,Y\in Def(L(A))\Rightarrow X\cap Y\in Def(L(A))$；

(3) $X,Y\in Def(L(A))\Rightarrow X\cup Y\in Def(L(A))$；

(4) $X\in Def(L(A))\Rightarrow X^c\in Def(L(A))$.

性质(1)说明空集 $\varnothing$、论域 U 是可定义集. 性质(2)表示可定义粒族 $Def(L(A))$ 对交运算封闭. 性质(3)和(4)指出可定义粒族 $Def(L(A))$ 对并运算和补运算均封闭.

从上述讨论中可知，粒结构可由逻辑语言 $L(A,\{\neg,\wedge,\vee\},U)$ 描述. 为了让粒结构更加实用，我们可以考虑只用逻辑连接词的子集形成的逻辑语言. 作为特例，假设粒结构对交、并运算封闭但对补运算不封闭. 由于逻辑算子 $\wedge$ 和 $\vee$ 对应集合交和并，我们可以用逻辑语言 $L(A,\{\wedge,\vee\},U)$ 来描述这类粒结构. $L(A,\{\wedge\},U)$ 和 $L(A,\{\vee\},U)$ 是两个粒结构更为特殊的例子[4]. 前者对交运算封闭，后者对并运算封闭(包含 U 并且对交运算封闭的粒结构是闭式系统).

事实上，对 Pawlak 粗糙集模型用基于集合论的粒计算方法描述与用基于等价关系进行描述是等价的. 等价关系 E 产生论域上的一个划分，称为等价类，记为 U/E. 很显然，每个等价类的元素在 $L(A)$ 中满足相同的等式. 等价类 U/E 即 $\{m(\wedge_{a\in At}a=I_a(x))\}$. 通过 U/E，我们可以构造一个 σ-代数，$\sigma(U/E)$，包括空集 $\varnothing$、全集 U 和作为基的等价类 U/E，并且对交、并、补运算封闭. 利用逻辑语言

$L(A)$,σ-代数$\sigma(U/E)$确切地说是关于语言$L(A)$的可定义集族,即$Def(L(A,\{\neg,\wedge,\vee\},U))$.因此,Pawlak粒度空间也可记为$(U,U/E,\sigma(U/E))$.

3.3 Pawlak粗糙集上下近似

对一个不可定义集,是不可能找到一个逻辑公式来描述它的涵义集.为了刻画一个不可定义集,我们可以用两个可定义集分别从上和下对它进行近似.可定义集族是幂集的一个子系统.我们可以用基于子系统的定义对粗糙集上下近似进行描述.

定义2. 假设对象的一个子集$X\subseteq U$,我们定义一对上下近似:

$$\underline{apr}(X)=\bigcup\{Y|Y\in Def(L(A)),Y\subseteq X\},$$
$$\overline{apr}(X)=\bigcap\{Y|Y\in Def(L(A)),X\subseteq Y\} \quad (7)$$

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

由于可定义集族对交、并、补运算封闭,所以这个定义得到的上、下近似均是唯一的一个集合.采用基于等价关系的表述,近似空间记为

$$apr=(U,U/E,\sigma(U/E)).$$

定义3. 在近似空间$apr=(U,U/E,\sigma(U/E))$中,对象的一个子集$X\subseteq U$,

$$\underline{apr}(X)=\bigcup\{Y|Y\in\sigma(U/E),Y\subseteq X\},$$
$$\overline{apr}(X)=\bigcap\{Y|Y\in\sigma(U/E),X\subseteq Y\} \quad (8)$$

即$\underline{apr}(X)$是$\sigma(U/E)$中包含在X中的最大集合,$\overline{apr}(X)$是$\sigma(U/E)$中包含了X的最小集合.

文献[7]给出了另一种基于粒的定义.等价关系E可由一个从U到2^U的映射表达.更具体地说,该映射为

$$[x]_E=\{y\in U|xEy\} \quad (9)$$

子集$[x]_E$表示包含x的等价类.所有等价类组成的集族称为商集,记为$U/E=\{[x]_E|x\in U\}$.

定义4. 在近似空间$apr=(U,U/E,\sigma(U/E))$中,对任意一个对象的子集$X\subseteq U$,

$$\underline{apr}(X)=\bigcup\{[x]_E|[x]_E\in(U/E),[x]_E\subseteq X\},$$
$$\overline{apr}(X)=\bigcap\{[x]_E|[x]_E\in(U/E),[x]_E\cap X\neq\varnothing\} \quad (10)$$

即$\underline{apr}(X)$是包含在X中的等价类的并,$\overline{apr}(X)$是与X交不为空的所有等价类的并.

这3种定义对粗糙集上下近似提供了不同的解释,但三者相互等价.由定义可知,$\underline{apr}(X)\subseteq X\subseteq\overline{apr}(X)$,并且$\underline{apr}(X)$和$\overline{apr}(X)$均为可定义集.由此很容易得到:一个可定义集的下近似是其本身,上近似也为其本身[3].

定理1. 一个对象集$X\subseteq U$是一个可定义集当且仅当满足下列条件:

$$\underline{apr}(X)=\overline{apr}(X)=X \quad (11)$$

这个定理通常被用来判断一个论域中子集X的可定义性.它可被看成是可定义性的一个应用.

3.4 实 例

假设$U=\{a,b,c,d,e\}$.给定一个等价关系:$aEa,aEb,bEa,bEb,cEc,cEe,dEd,eEc,eEe$,由此我们可以得到一个划分$U/E=\{\{a,b\},\{c,e\},\{d\}\}$.我们通过任意等价类的并可以构建一个$\sigma$-代数:

$$\sigma(U/E)=\{\varnothing,\{a,b\},\{c,e\},\{d\},\{a,b,c,e\},\{a,b,d\},\{c,d,e\},U\}.$$

相应的粒度空间为

$$(U,U/E,\sigma(U/E)),$$

其中,$U=\{a,b,c,d,e\}$,$U/E=\{\{a,b\},\{c,e\},\{d\}\}$,$\sigma(U/E)=\{\varnothing,\{a,b\},\{c,e\},\{d\},\{a,b,c,e\},\{a,b,d\},\{c,d,e\},U\}$.

考虑对象子集$X=\{a,b,c,d\}$.它不能通过基本粒的并得到,即它是个不可定义粒.我们在σ-代数$\sigma(U/E)$中用一对子集分别从上下对它近似.从等式(8)可知:

$$\underline{apr}(X)=\bigcup\{Y|Y\in\sigma(U/E),Y\subseteq X\}$$
$$=\{a,b\}\cup\{d\}\cup\{a,b,d\}=\{a,b,d\},$$
$$\overline{apr}(X)=\bigcap\{Y|Y\in\sigma(U/E),X\subseteq Y\}$$
$$=\{a,b,c,d,e\}.$$

显而易见

$$\underline{apr}(X)=\{a,b,d\}\subseteq X\subseteq\overline{apr}(X)=\{a,b,c,d,e\}.$$

4 并可约粒度空间与交可约粒度空间

Pawlak粗糙集模型有很多扩展模型[20-27],多数是对等价关系进行推广.文献[25-26]将经典粗糙集拓展为基于相似关系的粗糙集.文献[27]进一步拓展为基于一般二元关系下的粗糙集.一般二元关系将经典粗糙集对论域的划分泛化为覆盖.对于基于覆盖的粗糙集我们也可以研究其粒度空间模型.

为了更一般地讨论,我们从基到粒结构的构造规则开始研究.一组基构造一个粒结构有两种对偶的构造规则.一种叫做并可约,另一种叫做交可约.它们分别对应不同的粒度空间.

4.1 并可约粒度空间

假设S_0是由覆盖C得到的基.我们说基S_0并可约如果其满足以下定义.

定义5. 假设C是一个覆盖.S_0是覆盖C的基.$S_0=\{S_{01},S_{02},\cdots,S_{0n}\}$.我们称$(U,S_0,S)$并可约

的，如果它满足

(1) $\bigcup_{i=1,2,\cdots,n} S_{0i}=U$；

(2) $S_{0j}\neq\bigcup S_{0i}, i,j=1,2,\cdots,n, i\neq j$.

式(1)表示基 S_0 中的所有基本粒的并是整个论域. 式(2)说明基 S_0 中任意一个基本粒都不能由其它基本粒的并得到. 我们对 S_0 中的子集用并运算，可以得到一个粒结构，记为 S. 我们称三元组(U,S_0,S)为并可约粒度空间. 从 S_0 构造 S 的过程是一个从“小粒”到“大粒”的过程.

需要说明的是，一个基本粒可能是另一个基本粒的子集(如 $U=\{a,b,c,d\}$，$\{\{a,b\},\{a,b,c\},\{b,d\}\}$可以是一个并可约粒度空间的基本粒族).

下面，我们介绍两种从覆盖 C 得到基 S_0 的方法.

方法 1(删除法 1).

假设在论域 U 上存在一个覆盖，记为 $C=\{C_1,C_2,\cdots,C_n\}$. 找到基的最简单的方法是逐个删除 C_j 如果 $C_j=\bigcup C_i, i,j=1,2,\cdots,n, i\neq j$，直到任意的 C_j 都不能由其它 C_i 通过并运算得到. 这个过程很容易被理解. 所有剩下的集合组成这个覆盖的基，即基本粒族. 很显然，剩下的集合的并覆盖整个论域 U 并且任意一个集合都不能由其它集合的并得到，满足并可约定义的两个条件.

采用算法描述可以表示如下：

```
1. 令 C={C1,C2,…,Cn};
2. m=0;
   while |n-m|>0
     m=n;
     for j=1 to n
       if Cj=∪Ci, i=1,2,…,n, i≠j
         C=C-{Cj};
         n=card(C);
         break;
       end
     end
   end
3. RETURN C.
```

方法 2(最小描述法).

另一种从覆盖得到并可约基本粒的方法是建立在覆盖粗糙集的一些概念基础上的.

定义 6. 令 C 是 U 上的一个覆盖，$x\in U$，

$$Md(x)=\{K\mid K\in C\wedge x\in K\wedge (\forall L\in C\wedge x\in L\wedge L\subseteq K\Rightarrow K=L)\} \tag{12}$$

称为 x 的最小描述.

对论域 U 上的任意元素 x，我们在一个覆盖下都能得到关于 x 的一个最小描述. x 的最小描述可能由包含 x 的多个集合组成. 因此，论域 U 上所有 x 的最小描述的子集族组成基 S_0，即基本粒族：

$$S_0=\{K\mid K\in Md(x)\wedge x\in U\} \tag{13}$$

显而易见，$Md(x)$的每个子集都包含 x，即所有 x 的最小描述覆盖了整个论域 U. 对每一个论域 U 上的元素 x，x 的最小描述中的一个子集都不能是其它 x 最小描述的子集的并，也不能是 x 最小描述中另一个子集的子集，即 $Md(x)=\{K_1,K_2,\cdots,K_n\}$，$K_j\neq\bigcup K_i, K_j\not\subset K_i, i,j=1,2,\cdots,n$. 这一点根据最小描述的定义，很容易可以推断出来. 而且，x 的最小描述的子集不可能是论域 U 上其它元素的最小描述的子集的并.

这两种方法都能得到一个唯一的并可约粒度空间的基，它们相互等价. 特别地，一个划分是个特殊的覆盖. 划分中任意基本粒的交为空，然而在覆盖下却未必是. 如果将划分视为基本粒族，那么在划分下通过基本粒的并运算，我们可以得到一个并可约粒结构，即 σ-代数. 这也是粗糙集理论的主要论题.

4.2 交可约粒度空间

交可约粒度空间的形成是与并可约粒度空间构造相对应的对偶过程.

定义 7. 假设覆盖 C，令 $S_0=\{S_{01},S_{02},\cdots,S_{0n}\}$. 我们称$(U,S_0,S)$是交可约的，如果它满足

(1) $\bigcup_{i=1,2,\cdots,n} S_{0i}=U$；

(2) $S_{0j}\neq\bigcap S_{0i}, i,j=1,2,\cdots,n, i\neq j$.

三元组(U,S_0,S)被称为交可约粒度空间.

式(1)表示基 S_0 中的所有基本粒的并是整个论域. 式(2)说明基 S_0 中任意一个基本粒都不能由其它基本粒的交得到. 一个基本粒可以是另一个基本粒的子集(如 $U=\{a,b,c,d\}$，$\{\{a,b\},\{a,b,c\},\{b,d\}\}$可以是一个交可约粒结构的基). 从基本粒构造交可约粒结构的过程是一个从“大粒”到“小粒”的过程，与并可约粒结构的形成过程相比较，它是一个对偶过程. 但对一般的粒结构都有一个共同的性质，那就是所有基本粒的并覆盖整个论域 U. 一般来说，多数理论模型可归结为并可约粒度空间，而交可约粒度空间则出现较少.

对一个覆盖，可以用删除法来得到交可约的基. 具体来说，如果覆盖中的一个子集可由覆盖中的其它子集的交得到，则删除它，直到不存在任何一个子集可由剩下子集的交得到. 我们用形式语言描述如下.

方法 3(删除法 2).

假设在论域 U 上有个覆盖记为 $C=\{C_1,C_2,\cdots,$

$C_n\}$. 为了找到交可约粒度空间的基，我们删除 C_j 如果 $C_j=\bigcap C_i, i,j=1,2,\cdots,n, i\neq j$，直到任何 C_j 都不能由其它剩下的 C_i 的交得到. 这个过程很容易被理解. 所有剩下的集合组成交可约的基，即基本粒. 很显然，剩下的子集覆盖整个论域 U，并且没有一个子集是其它子集的交，满足交可约的两个条件.

采用算法描述可以表示如下：

```
1. 令 C={C1,C2,…,Cn};
2. m=0;
   while |n-m|>0
     m=n;
     for j=1 to n
       if Cj=∩Ci, i=1,2,…,n, i≠j
         C=C-{Cj};
         n=card(C);
         break;
       end
     end
   end
3. RETURN C.
```

5 并可约粒度空间与交可约粒度空间的上下近似

为了定义并可约粒度空间与交可约粒度空间的上下近似，我们先研究扩展的粗糙集的上下近似在粒度空间中的表述，这里的扩展粗糙集模型采用子系统来描述.

5.1 扩展的粗糙集上下近似

5.1.1 定 义

近似空间 $apr=(U,U/E,\sigma(U/E))$ 唯一决定一个拓扑空间 $(U,\sigma(U/E))$，其中 $\sigma(U/E)$ 是所有开集和闭集集族[28]. 等式(8)定义的下近似算子是合理的，只要该子系统对并封闭. 类似的，上近似算子定义要合理只需要子系统对交运算封闭. 因此，我们可以采用两个子系统[29-30]，使得下近似算子所在的子系统对并封闭，上近似算子所在的子系统对交封闭. 为了保证近似算子的对偶性，两个子系统的子集必须通过补运算相互联系[29]. 为进一步扩展基于子系统的定义，我们删除那些限制.

扩展粗糙集上下近似的定义与 Cattaneo[20] 定义的抽象近似空间表述相关. 在这里，我们主要讨论集合论背景，所以删除抽象近似空间的一些抽象约束.

令 $S_l, S_u\subseteq 2^U$ 是 2^U 上的两个子系统. 三元组 $apr=(U,S_l,S_u)$ 称为近似空间. 我们对 S_l 和 S_u 限定两个条件：

① $\varnothing\in S_l, \varnothing\in S_u$；

② $U\in S_l, U\in S_u$.

S_l 中的子集可以理解为一族可定义的或可辨识的集合. S_u 中的子集可理解为另一族可定义的或可辨识的集合. 我们的目的是用 S_l 中的可定义集从下去近似 2^U-S_l 中的不可定义集，用 S_u 中的可定义集从上去近似 2^U-S_u 中的不可定义集.

定义 8. 在抽象近似空间 $apr=(U,S_l,S_u)$ 中，下近似和上近似分别定义为

$$\underline{apr}(X)=\{Y\in S_l\,|\,Y\subseteq X, \forall Y'\in S_l(Y\subset Y'\Rightarrow Y'\not\subset X)\},$$
$$\overline{apr}(X)=\{Y\in S_u\,|\,X\subseteq Y, \forall Y'\in S_u(Y'\subset Y\Rightarrow X\not\subset Y')\} \tag{14}$$

为了方便，我们对扩展的上下近似采用相同的符号. 下近似 $\underline{apr}(X)$ 是 $\{Y\in S_l\,|\,Y\subseteq X\}$ 中的所有最大集合组成的集族，上近似 $\overline{apr}(X)$ 是 $\{Y\in S_u\,|\,X\subseteq Y\}$ 中的所有最小的集合组成的集族. 这个定义是等式(8)的一个推广. 推广的上下近似算子具有以下性质：

(1) $\underline{apr}(\varnothing)=\{\varnothing\}$, $\overline{apr}(\varnothing)=\{\varnothing\}$；

(2) $\underline{apr}(U)=\{U\}$, $\overline{apr}(U)=\{U\}$；

(3) $X\subseteq Z\Rightarrow(\exists A\in\underline{apr}(X), \exists B\in\underline{apr}(Z), A\subseteq B)X\subseteq Z\Rightarrow(\exists A\in\overline{apr}(X), \exists B\in\overline{apr}(Z), A\subseteq B)$；

(4) $Y\in\underline{apr}(X)\Rightarrow Y\subseteq X, Y\in\overline{apr}(X)\Rightarrow X\subseteq Y$；

(5) $Y\in\underline{apr}(X)\Rightarrow\underline{apr}(Y)=\{Y\}$,
$Y\in\overline{apr}(X)\Rightarrow\overline{apr}(Y)=\{Y\}$；

(6) $Y\in\overline{apr}(X)\Rightarrow\underline{apr}(Y)=\{Y\}$,
$Y\in\underline{apr}(X)\Rightarrow\overline{apr}(Y)=\{Y\}$.

这些性质都很容易从推广的近似算子的定义中得到.

5.1.2 特 例

我们讨论几种不同条件下推广的粗糙集上下近似形式.

第 1 种情况：S_l 对并封闭，S_u 对交封闭.

如果 S_l 对并封闭，下近似根据定义 8 由一个集合组成，即

$$\underline{apr}(X)=\{\bigcup\{Y\in S_l\,|\,Y\subseteq X\}\} \tag{15}$$

类似的，如果 S_u 对交封闭，上近似根据定义 8 也由一个集合组成，即

$$\overline{apr}(X)=\{\bigcap\{Y\in S_u\,|\,X\subseteq Y\}\} \tag{16}$$

第 2 种情况：S_l 和 S_u 是对偶的子系统，即 $S_u=\{Y^c\,|\,Y\in S_l\}$ 且 $S_l=\{Y^c\,|\,Y\in S_u\}$. 近似算子满足如下

性质：

$$Y \in \underline{apr}(X) \Rightarrow Y^c \in \overline{apr}(X^c) \tag{17}$$

第 3 种情况：$S_l=S_u$. 当 $S_l=S_u=S$ 时，我们得到近似空间 $apr=(U,S)$. 我们可以定义近似算子如下：

$$\underline{apr}(X)=\{Y\in S \mid Y\subseteq X, \forall Y'\in S(Y\subset Y'\Rightarrow Y'\not\subset X)\},$$
$$\overline{apr}(X)=\{Y\in S \mid X\subseteq Y, \forall Y'\in S(Y'\subset Y\Rightarrow X\not\subset Y')\} \tag{18}$$

第 4 种情况：$S_l=S_u=S$ 且对补封闭. 近似算子的定义与等式(19)相同. 它也满足等式(18)的性质.

第 5 种情况：$S_u=S$ 对交封闭且 $S_l=S^c$ 对并封闭. 我们定义：

$$\underline{apr}(X)=\{\bigcup\{Y\in S^c \mid Y\subseteq X\}\},$$
$$\overline{apr}(X)=\{\bigcap\{Y\in S \mid X\subseteq Y\}\} \tag{19}$$

它们对应于闭式系统中的粗糙近似[1]. 由于闭式系统只对交运算封闭，上下近似算子满足较少的性质，有性质(3)～(6).

第 6 种情况：$S_l=S$ 对交和并封闭，$S_u=S^c$ 对交并也封闭. 我们定义：

$$\underline{apr}(X)=\{\bigcup\{Y\in S \mid Y\subseteq X\}\},$$
$$\overline{apr}(X)=\{\bigcap\{Y\in S^c \mid X\subseteq Y\}\} \tag{20}$$

它们对应于拓扑空间中的粗糙近似[27]. 实际上它们是拓扑内闭算子，满足性质(1)～(6).

第 7 种情况：$S_l=S_u=S$ 且对交、并、补都封闭. 这是标准的 Pawlak 粗糙集模型.

5.2　并可约粒度空间的上下近似

显然的，覆盖 C、基 S_0 和并可约粒度空间 S 之间的关系满足 $S_0\subseteq C\subseteq S$. 假定一个覆盖 C，我们可以得到唯一的一个并可约粒度空间(U,S_0,S). 并可约粒度空间中的所有子集可被认为是可定义粒，即可定义集.

在近似空间 $apr=(U,S_0,S)$下，我们分别从上下用一对可定义集来描述一个不可定义集. 由于对并封闭，正好是第一种情况. 下近似是唯一的而上近似由一个集族表示. 因此，我们有如下定义.

定义 9.　在近似空间 $apr=(U,S_0,S)$中，一对近似算子$\underline{apr},\overline{apr}$: $2^U\rightarrow 2^U$定义为

$$\underline{apr}(X)=\bigcup\{Y \mid Y\in S\wedge Y\subseteq X\},$$
$$\overline{apr}(X)=$$
$$\{Y \mid Y\in S\wedge X\subseteq Y, \forall Y'\in S(Y'\subset Y\Rightarrow X\not\subset Y')\} \tag{21}$$

下近似是 S 中包含于 X 的最大集合. 上近似则由 S 中所有包含了 X 的最小集合组成. 在这种情况下，包含在 X 中的最大集是唯一的，而包含了 X 的最小集合不再是唯一的. 如果 X 是可定义集，即 $X\in S$，满足$\underline{apr}(X)=X$ 且 $\overline{apr}(X)=X$[3].

5.3　交可约粒度空间的上下近似

从交可约粒度空间的定义可知，$S\subseteq S_0\subseteq C$. 与并可约粒度空间相比，$S$ 的位置颠倒了. 有了基本粒，我们可以组成交可约粒度空间(U,S_0,S)的基. 我们也可以通过对任意基本粒的交运算，加上空集$\varnothing$和全集 U 构造粒度空间(U,S_0,S). 所有交可约粒度空间的子集是粒，也是可定义集.

在近似空间 $apr=(U,S_0,S)$中，我们用可定义集分别从上下对不可定义集进行描述. 由于对交封闭，符合第一种情况，上近似是唯一的而下近似是一个集族. 因此，我们有如下定义.

定义 10.　在近似空间 $apr=(U,S_0,S)$中，一对近似算子$\underline{apr},\overline{apr}$: $2^U\rightarrow 2^U$定义为

$$\underline{apr}(X)=\{Y\in S \mid Y\subseteq X, \forall Y'\in S(Y\subset Y'\Rightarrow Y'\not\subset X)\},$$
$$\overline{apr}(X)=\bigcap\{Y \mid Y\in S\wedge X\subseteq Y\} \tag{22}$$

下近似由 S 中包含于 X 的所有最大集合组成. 上近似由 S 中包含了 X 的最小集合组成. 在这种情况下，包含了 X 的最小集合是唯一的，包含在 X 中的最大集合不再是唯一的. 如果 X 是一个可定义集，即 $X\in S$，满足$\underline{apr}(X)=X$ 且 $\overline{apr}(X)=X$.

5.4　William Zhu 的覆盖粗糙集模型下的上下近似的合理性

William Zhu 提出了 4 种基于覆盖的粗糙集模型[13-16]. 所有这些模型都有相同的下近似定义而不同的上近似定义. 下近似定义为$\underline{apr}(X)=\bigcup\{Y \mid Y\in S\wedge Y\subseteq X\}$. 我们假设这些模型都是在并可约粒度空间下讨论的. 否则，一般的覆盖不可能有对并封闭的性质. 对任意的一个集合 X，下近似不再是包含于 X 中的最大集合，而是包含于 X 的一些最大集合形成的集族.

若 X 是可定义集，那么应该$\underline{apr}(X)=X=\overline{apr}(X)$，但我们发现在有些模型中，有下近似满足$\underline{apr}(X)=X$. 然而，上近似的定义则不都满足$\overline{apr}(X)=X$. 故近似空间$(U,S_0,S)$的成员不再都是可定义集. 因此，某些模型可能存在不合理性. 我们将逐个对模型进行讨论.

第 1 种覆盖上近似定义如下：

$$\overline{apr}(X)=\underline{apr}(X)\bigcup\{Md(x) \mid x\in X-\underline{apr}(X)\} \tag{23}$$

如果X是可定义集，$\underline{apr}(X)=X$，$X-\underline{apr}(X)=\varnothing$，则 $Md(x)=\varnothing$. 因此，$\overline{apr}(X)=X$. 第 1 种覆盖广义粗糙集是合理的.

第 2 种覆盖上近似定义为

$$\overline{apr}(X)=\bigcup\{Y\mid Y\in S,Y\cap X\neq\varnothing\}\tag{24}$$

如果 X 是可定义集，可能存在多个可定义集满足 $Y\cap X\neq\varnothing$. 因此，上近似 $\overline{apr}(X)$ 可能不等于 X. 第 2 种覆盖广义粗糙集模型是不合理的.

第 3 种覆盖上近似定义为

$$\overline{apr}(X)=\bigcup\{Md(x)\mid x\in X\}\tag{25}$$

如果 X 是可定义集，对 $x\in X$，$\bigcup Md(x)$ 可能不包含在 X 内. 因此，$\overline{apr}(X)$ 可能不等于 X. 即第 3 种覆盖广义粗糙集是不合理的.

第 4 种覆盖上近似定义为

$$\overline{apr}(X)=\underline{apr}(X)\cup\{Y\mid Y\cap(X-\underline{apr}(X))\neq\varnothing\}\tag{26}$$

如果 X 是可定义集，$\underline{apr}(X)=X$，即 $(X-\underline{apr}(X))=\varnothing$，$\forall Y\cap\varnothing=\varnothing$. 因此，$\overline{apr}(X)=\underline{apr}(X)=X$. 第 4 种覆盖广义粗糙集是合理的.

6 形式概念分析的粒结构与上下近似

形式概念分析主要研究数据的形式化表达和分析[8-9]. 它主要研究基于一组属性的一组对象集的可定义性，或者一组对象集的属性集的可定义性.

令 U 和 V 为任意两个有限集. U 的元素为对象，V 的元素称为属性. 对象和属性之间的关系用 U 和 V 间的二元关系描述，其是笛卡尔乘积 $U\times V$ 的子集. 对元素对 $x\in U$，$y\in V$，如果 $(x,y)\in R$，记为 xRy，我们说 x 具有属性 y 或属性 y 为对象 x 所具有. 三元组 (U,V,R) 称为一个形式背景.

基于一个二元关系，我们为一个对象关联一些属性. 对象 $x\in U$ 具有属性集：

$$xR=\{y\in V\mid xRy\}\subseteq V\tag{27}$$

类似的，属性 y 由对象集所有

$$Ry=\{x\in U\mid xRy\}\subseteq U\tag{28}$$

通过扩展这些符号，我们可以建立对象子集和属性子集之间的关系. 这就产生两个算子，一个是从 2^U 到 2^V 上，另一个从 2^U 到 2^U.

定义 11. 假设 (U,V,R) 是一个形式背景. 对对象子集，我们将它和一些属性联系起来：

$$\begin{aligned}X^*&=\{y\in V\mid\forall x\in U(x\in X\Rightarrow xRy)\}\\&=\{y\in V\mid X\subseteq Ry\}=\bigcap_{x\in X}xR\end{aligned}\tag{29}$$

对属性子集，我们将一些对象与之关联：

$$\begin{aligned}Y^*&=\{x\in U\mid\forall y\in V(y\in Y\Rightarrow xRy)\}\\&=\{x\in U\mid Y\subseteq xR\}=\bigcap_{y\in Y}Ry\end{aligned}\tag{30}$$

假设 $X\subseteq U$，$Y\subseteq V$，如果 $X=Y^*$ 且 $Y=X^*$，则 (X,Y) 对称为形式背景 (U,V,R) 的一个形式概念. $X=ex(X,Y)$ 称为概念的外延，$Y=in(X,Y)$ 称为概念的内涵. 所有形式概念形成一个完全格，称为概念格，记为 $L(U,V,R)$ 或简记为 L.

定义 12. 对一个形式概念格 L，所有外延形成的集族为

$$EX(L)=\{ex(X,Y)\mid(X,Y)\in L\}\tag{31}$$

6.1 形式概念分析中的粒结构

系统 $EX(L)$ 包括空集 $\varnothing$，全集 U，并且对交运算封闭. 在这种情况下，它是一个交可约粒结构. 此时，覆盖与交可约粒结构相同，用删除法，我们很容易得到交可约粒度空间的基，记为 $EX_0(L)$. 所有 $EX(L)$ 中的子集都可以通过 $EX_0(L)$ 中的子集的交运算得到. $EX_0(L)$ 中的所有子集被认为是基本粒，而 $EX(L)$ 中的子集是粒. 因此，我们有粒度空间：

$$(U,EX_0(L),EX(L)),$$

其中，U 是论域，$EX_0(L)\subseteq 2^U$ 是交可约粒度空间，即所有的外延的基，$EX(L)\subseteq 2^U$ 是所有外延的集族，称为粒.

三元组 $(U,EX_0(L),EX(L))$ 是形式概念分析中的交可约粒度空间.

6.2 形式概念分析中的上下近似

一个完整的形式概念由对象的一个可定义集和属性的一个可定义集组成. 概念格是所有这样的可定义概念族. 给定任意一个对象集，有可能不是形式概念的外延. 因此，这个集合可认为是不可定义对象集. 这样一个对象子集可以由可定义对象集近似(即形式概念的外延). 近似算子可由粗糙集理论的基于子系统的形式来定义，这个定义是结合前述的第一种和第 3 种情况的.

$EX(L)$ 包含空集 $\varnothing$，全集 U，且对交运算封闭. 它定义了一个近似空间 $apr=(U,EX_0(L),EX(L))$. 我们如果对上下近似保持直观解释，那么下近似是 $EX(L)$ 中包含在 X 里的一个最大可定义集合，上近似是 $EX(L)$ 中包含了 X 的一个最小可定义集合. 由于系统 $EX(L)$ 对并运算不封闭，包含 X 的最小集合是唯一的，而包含在 X 里的最大集合不再是唯一的.

定义 13. 在近似空间 $apr=(U,EX_0(L),EX(L))$ 中，对对象子集 $X\subseteq U$，它的上下近似定义为

$$\begin{aligned}&\overline{apr}(X)=\{\bigcap\{Y\in EX(L)\mid X\subseteq Y\}\},\\&\underline{apr}(X)=\{Y\in EX(L)\mid Y\subseteq X,\\&\qquad\forall Y'\in EX(L)(Y\subset Y'\Rightarrow Y\not\subset X)\}\end{aligned}\tag{32}$$

因此，在形式概念分析中，一个集合可能被多个可定义对象集从下近似.

6.3 实 例

表1是一个形式背景. 对象的论域为 $U=\{1,2,3,4,5,6,7,8\}$，属性的论域为 $P=\{a,b,c,d,e,f,g,h,i\}$. 根据形式概念的定义，所有外延族 $EX(L)$ 为

$$EX(L)=\{\varnothing,\{3\},\{4\},\{6\},\{7\},\{2,3\},\{3,4\},\{3,6\},\{5,6\},\{6,8\},\{1,2,3\},\{2,3,4\},\{6,7,8\},\{5,6,8\},\{1,2,3,4\},\{5,6,7,8\},\{1,2,3,5,6\},\{3,4,6,7,8\},U\}.$$

采用删除法，我们可以得到交可约粒度空间的基. 即

$$EX_0(L)=\{\{7\},\{2,3,4\},\{5,6,8\},\{1,2,3,4\},\{5,6,7,8\},\{1,2,3,5,6\},\{3,4,6,7,8\}\}.$$

表 1 形式背景[28]

	a	b	c	d	E	f	g	h	i
1	*	*					*		
2	*	*					*	*	
3	*	*	*				*	*	
4	*		*				*	*	*
5	*	*		*		*			
6	*	*	*	*		*			
7	*		*	*	*				
8	*		*	*		*			

我们有如下交可约粒度空间：

$$(U,EX_0(L),EX(L)),$$

其中，$U=\{1,2,3,4,5,6,7,8\}$，$EX_0(L)=\{\{7\},\{2,3,4\},\{5,6,8\},\{1,2,3,4\},\{5,6,7,8\},\{1,2,3,5,6\},\{3,4,6,7,8\}\}$，$EX(L)=\{\varnothing,\{3\},\{4\},\{6\},\{7\},\{2,3\},\{3,4\},\{3,6\},\{5,6\},\{6,8\},\{1,2,3\},\{2,3,4\},\{6,7,8\},\{5,6,8\},\{1,2,3,4\},\{5,6,7,8\},\{1,2,3,5,6\},\{3,4,6,7,8\},U\}$.

对对象子集 $X=\{2,3,4,6\}$. 它不是一个形式概念的外延. 它可被看成是不可定义粒. 我们用 $EX(L)$ 中的一对可定义对象集从上下近似它：

$$\begin{aligned}\overline{apr}(X)&=\{\bigcap\{Y\in EX(L)\mid X\subseteq Y\}\}\\&=\{1,2,3,4,5,6,7,8\},\\\underline{apr}(X)&=\{Y\in EX(L)\mid Y\subseteq X,\forall Y'\in EX(L)\\&\quad(Y\subset Y'\Rightarrow Y\not\subset X)\}\\&=\{\{2,3,4\},\{3,6\}\}.\end{aligned}$$

7 知识空间中的粒结构和上下近似

知识空间[10-12]是数学心理学的一种新的范例. 它是一种对知识进行评估的系统方法. 例如，老师为一个刚从国外回来的新的学生选择合适的年级，那么哪个年级比较适合这个学生呢？这个老师可以先问一个问题，然后根据学生的回答再继续问下一个问题. 经过一些问题后，这个学生所掌握知识的大致情况将被了解. 知识空间的主要思想就是对这样一个评估过程的构建用数学理论来解释.

知识空间中研究的对象是一个有限论域集和一些论域子集族，记为集对 $(Q,\mathbb{K})$，其中 $\mathbb{K}\subseteq 2^Q$ 是 Q 的幂集的子集族. Q 的元素称为问题，集族 $\mathbb{K}$ 中的集合称为知识状态，记为 K. 个体的知识状态由个体所能解决的问题集表示. 每个知识状态可被看成是一个粒. 所有知识状态的集族，加上空集 $\varnothing$ 和全集 U 称为一个知识结构，从粒计算观点出发，可以看成是具有粒度的知识结构. 这个知识结构是一个覆盖，由所有知识状态，包括空集 $\varnothing$ 和全集 Q 组成. 知识空间是并可约粒度空间的实例解释.

知识空间里有两种知识结构. 一个为基于推测关系的知识结构，对交和并封闭，另一个是基于推测系统的知识结构，只对并封闭，通常情况下我们把这种知识结构简称为知识空间. 我们分别具体讨论在这两种情况下的并可约粒度空间.

7.1 基于推测关系的知识空间的粒结构及其上下近似

7.1.1 基于推测关系的知识空间的粒结构

我们首先介绍知识空间的一些概念. 了解知识结构的一个直观的方法是从推测关系开始的. 在知识空间中，问题集 Q 上的推测关系是 Q 上任何具有自反性和传递性的关系 R. aRb 表示如果学生能正确回答问题 b，则我们可以推测他也能正确回答问题 a. 推测关系限定了对应的知识结构. 例如，aRb 意味着如果一个知识状态包含 b，它一定包含 a.

定义 14. 对有限问题集 Q 上的推测关系 R，对应的知识结构 $\mathbb{K}$ 定义为

$$\mathbb{K}=\{K\mid\forall q,q'\in Q((qRq',q'\in K)\Rightarrow q\in K)\}\tag{33}$$

基于推测关系的知识结构包括空集 $\varnothing$，全集 U，且对交、并封闭.

对每个 Q 中的问题 q，在推测关系下，我们可以找到唯一一个由前件组成的问题集合 $R_p(q)=\{q'\mid q'Rq\}$. 所有问题的前件组成的集合构成的集族记为 B，它是问题 Q 上的一个覆盖. 每个问题的前件组成的集合称为基本粒. 通过对一些问题的前件组成的集合采用并操作，可以得到基于推测关系的知识结构 $\mathbb{K}$. 它定义了一个并可约粒度空间 $(Q,B,\mathbb{K})$. 所有知识状态称为 $(Q,B,\mathbb{K})$ 中的粒.

基于推测关系的知识结构是一个覆盖.在这种情况下,覆盖与并可约粒结构相同,我们可以用最小描述法得到并可约粒度空间的基.问题 q 的最小描述只有一个集合,与问题 q 的前件组成的集合相同.

对任意问题的前件组成的集合的并是基于推测关系的知识结构中的一个知识状态.这个知识结构是一个并可约粒结构.基于推测关系,我们有并可约粒度空间:

$$(Q,B,\mathbb{K}),$$

其中,Q 是问题集, $B\subseteq 2^Q$ 是前件组成的集合的集族,即 $B=\{R_P(q)\mid q\in Q\}$, $\mathbb{K}\subseteq 2^Q$ 是知识状态的集族.

注意:$B\subseteq\mathbb{K}$ 且每个知识状态可由 B 的一些子集的并来表示.

7.1.2 推测关系下的上下近似

基于推测关系的知识结构定义了近似空间 $apr=(Q,B,\mathbb{K})$.由于这个结构对交、并封闭,上下近似都是唯一的.

定义 15. 近似空间 $apr=(Q,B,\mathbb{K})$ 中,对象子集 $X\subseteq Q$,定义:

$$\underline{apr}(X)=\{\bigcup\{K\in\mathbb{K}\mid K\subseteq X\}\},$$
$$\overline{apr}(X)=\{\bigcap\{K\in\mathbb{K}\mid X\subseteq K\}\} \tag{34}$$

这个定义是建立在第一种情况基础上的.如果 X 是一个知识状态,即一个可定义集,则 $\underline{apr}(X)=X=\overline{apr}(X)$.下近似 $\underline{apr}(X)$ 是所有包含于 X 的可定义粒的并.上近似 $\overline{apr}(X)$ 是包含 X 的所有可定义粒的交.基于推测关系的知识结构不再对补封闭,即不满足对偶性.

知识空间的上下近似物理解释如下:假设 X 是一个学生正确回答的问题集,如果它不是一个知识状态,我们可用两个知识状态近似它.下近似表示学生肯定掌握的问题子集,而上近似描述的是这个学生可能掌握的问题子集.

7.1.3 实 例

假设 $Q=\{a,b,c,d,e\}$.给定一个推测关系:

$aRa,aRd,bRb,bRc,bRd,bRe,cRd,$
$cRd,cRe,dRd,eRe.$

我们得到知识结构:

$$\mathbb{K}=\{\varnothing,\{a\},\{b\},\{a,b\},\{b,c\},\{a,b,c\},\{b,c,e\},\{a,b,c,e\},\{a,b,c,d\},Q\}.$$

很容易看出这个知识结构对交、并运算是封闭的.它是基于推测关系的知识结构.因此,我们对每个问题都可以找到各自的前件组成的集合:

$$R_P(a)=\{a\},\ R_P(b)=\{b\},\ R_P(c)=\{b,c\},$$
$$R_P(d)=\{a,b,c,d\},\ R_P(e)=\{b,c,e\}.$$

因此,我们基于推测关系得到一个粒度空间:

$$(Q,B,\mathbb{K}),$$

其中, $Q=\{a,b,c,d,e\}$, $B=\{\{a\},\{b\},\{b,c\},\{a,b,c,d\},\{b,c,e\}\}$, $\mathbb{K}=\{\varnothing,\{a\},\{b\},\{a,b\},\{b,c\},\{a,b,c\},\{b,c,e\},\{a,b,c,e\},\{a,b,c,d\},Q\}$.

$\mathbb{K}$ 中的粒可由 B 中的一些基本粒通过并运算得到.如 $\{a,b\}=\{a\}\cup\{b\}$,$\{a,b,c,e\}=\{a\}\cup\{b,c,e\}$.

假设正确回答的问题集 $X=\{a,b,d\}$.由于它不是一个知识状态,我们用两个知识状态近似它:

$$\underline{apr}(X)=\{\bigcup\{K\in\mathbb{K}\mid K\subseteq X\}\}=\{a,b\},$$
$$\overline{apr}(X)=\{\bigcap\{K\in\mathbb{K}\mid X\subseteq K\}\}=\{a,b,c,d\} \tag{35}$$

7.2 基于推测系统的知识空间的粒结构与上下近似

7.2.1 基于推测系统的知识空间的粒结构

对推测关系,一个问题只有一个前件组成的集合.有的时候这是不合适的.实际中,我们假设知识结构只对并运算封闭.在上文中说道,一般情况下,我们都假设知识空间具有这样的知识结构.所以在这一小节,我们将给予推测系统的知识空间,简称为知识空间.因此,知识空间是弱化的基于推测关系的知识结构.

在有限集 Q 上的推测系统是一个映射 σ,这个映射将任意元素 $q\in Q$ 与 Q 上的非空子集 $\sigma(q)$ 关联起来,且映射满足以下 3 个条件:

(1) $C\in\sigma(q)\Rightarrow q\in C$;

(2) $(C\in\sigma(q),q'\in C)\Rightarrow\exists C'\in\sigma(q')(C'\subseteq C)$;

(3) $C\in\sigma(q)\Rightarrow\forall C'\in\sigma(q)(C'\not\subset C)$,

其中 C 是 $\sigma(q)$ 中的一个子集,称为问题 q 的子句.推测系统类似于邻域系统.问题 q 的一个子句实际上是 q 的一个前件组成的集合.每个问题可能有几个前件组成的集合,即问题 q 的子句不一定是唯一的.条件(1)推广了对关系的自反性条件,而条件(2)扩展了传递性的概念.条件(3)要求问题 q 的子句都是最大的集合.

定义 16. 对推测系统 (Q,σ),对应的知识结构为

$$\mathbb{K}=\{K\mid\forall q\in Q(q\in K\Rightarrow\exists C\in\sigma(q)(C\subseteq K))\} \tag{36}$$

其对并运算封闭.任何对并运算封闭的知识结构都称为知识空间.Q 上的推测系统和 Q 上的知识空间是一一对应的.

通常存在 $\mathbb{K}$ 的一个最小的子集族 B.对这个子集族 B,子集族中任何一个知识状态都不能由 B 中其它的知识状态的并得到.而知识空间中所有知识

状态都可以由这个最小的子集族中的一些子集的并得到. 我们称这样的子集族为$\mathbb{K}$的基. 很容易知道这个基是 Q 的一个覆盖. 基中的子集称为基本粒. 对应的知识空间定义了一个并可约粒度空间$(Q,B,\mathbb{K})$. 知识空间中的知识状态称为粒.

我们在知识空间中有并可约粒度空间：

$$(Q,B,\mathbb{K}),$$

其中,Q 是问题集,$B\subseteq 2^Q$ 是知识空间的基,称为基本粒,$\mathbb{K}\subseteq 2^Q$ 是知识空间中的知识状态集族,称为粒.

已经证明知识空间的基是所有由前件组成的集合的集合(集族)[12]. 每个子句是基 B 的一个子集. 因此,基的每个子集是某个问题的一个子句. 令 B_q 表示以 B 为基的知识空间$\mathbb{K}$中,B 内包含问题 q 的所有子集形成的集族. 因此,$B_q\subseteq\mathbb{K}_q$,其中$\mathbb{K}_q$是包含 q 的所有状态的集族. 我们对任意问题 q,也有 $B_q\subseteq B$,q 的所有子句的集族$\sigma(q)$可由以下形式设定：

$$\sigma(q)=B_q \tag{37}$$

等式(37)具体说明从知识空间的基是如何构造子句的.

7.2.2 推测系统下的上下近似

知识空间定义了近似空间$(Q,B,\mathbb{K})$. 和形式概念分析中的系统对交运算封闭不同的是,知识空间对并运算封闭. 因此,知识空间的下近似是唯一的而上近似是一个集族[23].

定义 17. 假定(Q,σ)是推测系统,且$\mathbb{K}\subseteq 2^Q$是相对应的知识结构,对并运算封闭. 在近似空间 $apr=(Q,B,\mathbb{K})$中,对象子集 $X\subseteq Q$,它的上下近似定义为

$$\underline{apr}(X)=\{\bigcup\{K\in\mathbb{K}\mid K\subseteq X\}\},$$
$$\overline{apr}(X)=\{K\in\mathbb{K}\mid X\subseteq K,\forall K'\in\mathbb{K}(K'\subset K\Rightarrow X\not\subset K')\} \tag{38}$$

这个定义是基于第 1 种和第 3 种情况的结合. 如果 X 是一个知识状态,我们有$\underline{apr}(X)=X=\overline{apr}(X)$. 下近似是$\mathbb{K}$中包含于 X 的最大集合,上近似是$\mathbb{K}$中包含 X 的最小集合组成的集族. 包含在 X 里的最大集是唯一的,而包含 X 的最小集合却不是唯一的.

7.2.3 实 例

假定 $Q=\{a,b,c,d,e\}$. 给定推测系统：

$$\sigma(a)=\{\{a\}\},$$
$$\sigma(b)=\{\{b,d\},\{a,b,c\},\{b,c,e\}\},$$
$$\sigma(c)=\{\{a,b,c\},\{b,c,e\}\},$$
$$\sigma(d)=\{\{b,d\}\},$$
$$\sigma(e)=\{\{b,c,e\}\}.$$

我们得到知识结构

$$\mathbb{K}=\{\varnothing,\{a\},\{b,d\},\{a,b,c\},\{b,c,e\},\{a,b,d\},\{a,b,c,d\},\{a,b,c,e\},\{b,c,d,e\},Q\},$$

很容易验证出知识空间对并运算封闭.

我们在知识空间中有一个粒度空间：

$$(Q,B,\mathbb{K}),$$

其中,$Q=\{a,b,c,d,e\}$,$B=\{\{a\},\{b,d\},\{a,b,c\},\{b,c,e\}\}$,$\mathbb{K}=\{\varnothing,\{a\},\{b,d\},\{a,b,c\},\{b,c,e\},\{a,b,d\},\{a,b,c,d\},\{a,b,c,e\},\{b,c,d,e\},Q\}$.

$\mathbb{K}$可通过对基 B 中的任意子句的并运算构造.

考虑正确回答的问题集 $X=\{b,c,d\}$. 由于它不是一个知识状态,我们用两个知识状态近似它：

$$\underline{apr}(X)=\{\bigcup\{K\in\mathbb{K}\mid K\subseteq X\}\}=\{b,d\},$$
$$\overline{apr}(X)=\{K\in\mathbb{K}\mid X\subseteq K,\forall K'\in\mathbb{K}(K'\subset K\Rightarrow X\not\subset K')\}=\{\{a,b,c,d\},\{b,c,d,e\}\} \tag{39}$$

8 结 论

粒和粒结构是粒计算研究中两个基本的概念.

本文提出了粒度空间的集合论框架,即用一个三层模型(U,S_0,S)对论域 U 上的特性进行描述. U 是论域,S_0是基本集族,即基本粒(也叫基). 可定义粒族 S 是基本集的超集族或者子集族,称为粒结构,S 的子集即为可定义集.

然后,我们将经典粗糙集在三层粒度空间模型中表示出来,并在覆盖粗糙集基础上,根据不同的构造规则,研究了两种粒度空间:并可约粒度空间和交可约粒度空间. 讨论了这两种粒度空间下,覆盖、基本粒和对应粒结构之间的关系,以及从覆盖求基本粒的方法.

此外,我们将粗糙集上下近似基于子系统的定义作了最一般的推广. 我们将上下近似分别在两个不同的子系统下进行定义,讨论了广义粗糙近似算子的基本性质. 新定义的粗糙集上下近似将不再由一个集合构成,而是依据近似空间的性质,可能由多个集合构成. 对新定义的近似算子加上某些限定条件,探讨了 7 种特例,并指出闭式系统、拓扑空间和粗糙集下的粗糙近似算子是我们新定义的特例. 另外,我们分别在并可约粒度空间和交可约粒度空间下给出上下近似的具体定义,并且指出 William Zhu 提出的 4 种基于覆盖的粗糙集模型中有两个是不合理的.

作为交可约粒度空间的范例,我们探讨了形式概念分析中的三层粒度空间模型及其上下近似,并

通过一个实例具体说明.作为并可约粒度空间的范例,我们研究了知识空间中的三层粒度空间模型及其上下近似,其中包括基于推测关系的知识结构下的三层粒度空间模型及其上下近似和基于推测系统下的三层粒度空间模型及其上下近似,并通过实例分别进行详细说明.

这个三层粒度空间模型使我们可以在同一环境下研究粗糙集、形式概念分析和知识空间.这 3 个理论具有相似的粒度空间,主要的区别在于对基本粒以及粒结构的构造与解释不同.这为我们开辟了研究这 3 个理论的新的途径.我们将上下近似的概念从粗糙集引入形式概念分析和知识空间中,一方面推广了粗糙集上下近似,提出一种新的广义粗糙近似算子;另一方面为形式概念分析中概念的外延和内涵的研究以及知识空间中的知识状态评估提出了一种新的方法.

参 考 文 献

[1] Yao Y Y. On generalizing rough set theory//Proceedings of the 9th International Conference (RSFDGrC). Chongqing, China, Springer, 2003: 44-51

[2] Yao Y Y. Information granulation and rough set approximation. International Journal of Intelligent Systems, 2001, 16(1): 87-104

[3] Yao Y Y, Zhou B. A logic language of granular computing//Proceedings of the 6th IEEE International Conference on Cognitive Informatics. Beijing, China, 2006: 178-185

[4] Miao Duo-Qian, Wang Guo-Yin, Liu Qing et al. Granular Computing: Past, Present and Future. Beijing: Science Press, 2007(in Chinese)
(苗夺谦,王国胤,刘清等. 粒计算的过去,现在与展望. 北京:科学出版社,2007)

[5] Feng Qin-Rong, Miao Duo-Qian, Cheng Yi. Presentation of relative partition granularity of attributes reduction for decision table. Journal of Chinese Computer Systems, 2008, 29(12): 2305-2308(in Chinese)
(冯琴荣,苗夺谦,程昳. 决策表属性约简的相对划分粒度表示. 小型微型计算机系统,2008,29(12):2305-2308)

[6] Feng Qin-Rong, Miao Duo-Qian, Cheng Yi, Xu Fei-Fei. Knowledge representation using partition granularity. Pattern Recognition and Artificial Intelligence, 2009, 22(1): 64-69(in Chinese)
(冯琴荣,苗夺谦,程昳,徐菲菲. 知识的划分粒度表示法. 模式识别与人工智能,2009,22(1):64-69)

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

[8] Ganter B, Wille R. Formal Concept Analysis: Mathematical Foundations. New York: Springer-Verlag, 1999

[9] Wille R. Restructuring lattice theory: An approach based on hierarchies of concepts//Ivan Rival. Ordered Sets. Dordecht-Boston: Reidel, 1982: 445-470

[10] Doignon J P, Falmagne J C. Spaces for the assessment of knowledge. International Journal of Man-Machine Studies, 1985, 23(2): 175-196

[11] Doignon J P, Falmagne J C. Knowledge Spaces. German: Springer, 1999

[12] Duntsch I, Gediga G. A note on the correspondences among entail relations, rough set dependencies, and logical consequence. Mathematical Psychology, 2001, 45(2): 393-401

[13] Zhu W, Wang F Y. Properties of the first type of covering-based rough sets//Proceedings of the ICDM 06. Hong Kong, China, 2006: 407-411

[14] Zhu W. Properties of the second type of covering-based rough sets//Proceedings of GrC and BI 06, WI 06, Hong Kong, China, 2006: 494-497

[15] Zhu W. Properties of the fourth type of covering-based rough sets//Proceedings of the HIS'06. Auckland, New Zealand, 2006: 43—43.

[16] Zhu W, Wang F Y. Properties of the third type of covering-based rough sets//Proceedings of the 6th International Conference on Machine Learning and Cybernetics. Hong Kong, China, 2007: 3746-3751

[17] Hobbs J R. Granularity//Proceedings of the 9th International Joint Conference on Artificial Intelligence. Los Angeles, California, USA, 1985: 432-435

[18] Orlowska E. Logical aspects of learning concepts. International Journal of Approximate Reasoning, 1988, 2: 349-364

[19] Yao Y Y. A note on definability and approximations//Peter J F et al. Transactions on Rough Sets VII. Lecture Notes in Computer Science 4400. Springer, 2007: 274-282

[20] Cattaneo G. Generalized rough sets: Preclusivity fuzzy-intuitionistic (BZ) lattices. Studia Logica, 1997, 58(1): 47-77

[21] Ziarko W. Variable precision rough set model. Journal of Computer and System Sciences, 1993, 46(1): 39-59

[22] Kondo M. On the structure of generalized rough sets. Information Sciences, 2005, 176(5): 589-600

[23] Kryszkiewicza M. Rough set approach to incomplete information systems. Information Sciences, 1998, 112(1-4): 39-49

[24] Skowron A, Stepaniuk J. Generalized approximation spaces//Lin T Y, Wildberger A M eds. Soft Computing. San Diego: The Society for Computer Simulation, 1995: 18-21

[25] Skowron A, Stepaniuk J. Tolerance approximation spaces. Fundamenta Informaticae, 1996, 27(2-3): 254-253

[26] Slowinski R, Vanderpooten D. Similarity relation as a basis for rough approximations//Wang P P eds. Advances in Machine Intelligence and Soft-Computing. Berlin: Springer, 1997: 17-33

[27] Wybraniec-Skardowska U. On a generalization of approximation space. Bulletin of the Polish Academy of Sciences, 1989, 37: 51-61

[28] Skowron A. On topology in information system. Bulletin of the Polish Academy of Sciences, Mathematics, 1988, 36: 477-479

[29] Yao Y. On generalizing pawlak approximation operators//Proceedings of the 1st International Conference on Rough Sets and Current Trends in Computing. Warsaw, Poland, 1998: 298-307

[30] Yao Y, Chen Y. Subsystem based generalizations of rough set approximations//Proceedings of the ISMIS 2005. New York, 2005: 210-218

MIAO Duo-Qian, born in 1964, professor, Ph. D. supervisor. His main research interests include rough set theory, pattern recognition, artificial intelligence etc.

XU Fei-Fei, born in 1983, Ph. D., lecturer. Her current research interests include fuzzy rough set theory, data mining, artificial intelligence etc.

YAO Yi-Yu, born in 1962, professor, Ph. D. supervisor. His main research interests include granular computing, rough sets and Web intelligence.

WEI Lai, born in 1980, Ph. D., lecturer. His current research interests include manifold learning, data mining, artificial intelligence etc.

Background

Granular computing, a rising research direction of artificial intelligence, is a new computing paradigm of problem solving and information processing. One of its central notions is hierarchical multilevel granularity. However, most of the existed work on granular computing has been done based on a single level. It is against the idea of granular computing. Though the formulation and interpretation of granularity are much application dependent, one may study some mathematical models independent of specific applications. However, to our best knowledge, there still lacks of a unified model which can describe the granular computing theory.

Therefore, in this paper, we propose a set-theoretic framework for granular computing. We assume that a granule is represented by a subset of a universal set and the granular structure is constructed based on the standard set-inclusion relation on a family of subsets of the universal set. With the proposed framework, we are able to describe and represent of multilevel granularity in granular computing. And we can also investigate granular structures used in rough sets, formal concept analysis and knowledge spaces.

The paper is partial work of the National Natural Science Foundation of China "Research on multi-level data model based granular computing". The project tries to introduce the multi-level data model into granular computing to provide a concrete expression method for multi-granularity information. Our group has published more than 130 papers on granular computing, including 80 papers indexed by EI or SCI.