

# 可变论域信息粒和粗糙集的数字特征

常雪尔, 孔庆钊, 王婉婷

(集美大学理学院, 福建 厦门 361021)

[摘要] 为了更准确直观地刻画信息粒, 从数字描述的角度, 研究可变论域信息表中信息粒和粗糙集的个数, 以及信息粒空间和粗糙集空间的基、维数等数字特征。

[关键词] 信息粒; 可变论域; 粗糙集; 数字特征

[中图分类号] TP 182

## Numerical Characteristics of Information Granules and Rough Sets Based on Variable Universe

CHANG Xueer, KONG Qingzhao, WANG Wanting

(School of Science, Jimei University, Xiamen 361021, China)

**Abstract:** In order to describe information granules more accurately and intuitively, from the perspective of digital description, this paper studies the number of information granules and rough sets in the variable universe information table, and the number characteristics of information granules space and rough set space, such as basis and dimension.

**Keywords:** information granule; variable universe; rough set; numerical characteristics

### 0 引言

在处理大量复杂信息时, 由于人类认知能力有限, 人们通过物理学中把大型物质划分为颗粒、分子、原子的思想引入到信息领域, 将大量复杂信息按其各自特征和性能划分为若干较为简单的块, 每个被分出来的块就被看作一个粒。这种处理信息的过程称为信息粒化, 它把复杂问题抽象、划分从而转化为若干较为简单的问题, 有助于人们更好地分析和解决问题<sup>[1-2]</sup>。1979年, Zadeh首次提出了信息粒的概念, 认为很多领域都存在信息粒, 只是在不同领域中的表现形式不同<sup>[3-6]</sup>。基于信息粒, 人们建立了粗糙集模型<sup>[7-10]</sup>、形式概念分析模型<sup>[11-14]</sup>、商空间模型<sup>[15]</sup>、三支决策模型<sup>[16-18]</sup>等许多粒计算模型来应对不同的复杂数据信息问题。

粗糙集作为一类重要的粒计算模型是由 Pawlak<sup>[7]</sup>于1982年首先提出的。粗糙集理论在分析和处理不精确、不一致、不完整信息与知识方面起着重要作用。该理论为数据分析、机器学习、数据挖掘、模式识别等方面的应用提供了一种有效的数学方法<sup>[19-21]</sup>。

在 Pawlak 粗糙集模型中, 通常把信息表中所有对象看作一个论域。但是, 在许多实际问题中, 并不需要选择信息表中全部的对象, 而是选择部分对象就能完成对不确定性知识的描述。显然,

[收稿日期] 2022-02-25

[基金项目] 福建省自然科学基金项目“信息系统中基于不同粒结构知识发现理论的比较研究”(2020J01707)

[作者简介] 通信作者: 孔庆钊(1978—), 副教授, 博士, 从事粒计算、智能信息处理和人工智能研究。E-mail: kongqingzhao@163.com

http://xuebaobangong.jmu.edu.cn/zkb

部分对象作为论域的模型会比传统模型在分析数据时更能节省时间,分析结果也会更加准确,而且问题的需求不同也会导致选择的论域不同。根据这一思想,基于可变论域的粗糙集模型应运而生<sup>[22]</sup>,并且基于该模型的上下近似算子的性质、属性约简、可变论域的极大部分等被深入研究。在概率论中,人们不仅需要研究随机变量的概率分布、密度函数、分布函数,还要考察其数学期望、方差、相关系数等数字特征。同理,对于一个从信息表中诱导出的粗糙集模型而言,除了要研究上下近似集、属性约简、规则提取等内容,也需要探索信息表中的信息粒和粗糙集模型的数字性质。因此,本文在文献[22]基础上,从数字描述的角度进一步讨论了基于可变论域信息粒和粗糙集模型的若干数字特征。

## 1 预备知识

一个信息表<sup>[23]</sup>通常写作  $I = (OB, AT, \{V_a : a \in AT\}, \{I_a : a \in AT\})$ , 其中  $OB (\neq \emptyset)$ : 是一个包含所有对象的有限集, 称为论域;  $AT (\neq \emptyset)$  为一个包含所有属性的有限属性集;  $V = \bigcup_{a \in AT} V_a$ ,  $V_a$  是属性  $a$  的所有属性值的集合;  $I_a : OB \rightarrow V_a$  是一个信息函数, 对象  $x$  关于属性  $a$  的属性值通常用  $I_a(x)$  表示,  $x \in OB, a \in AT$ 。

对任意属性集  $A \subseteq AT$ , 可以定义  $OB$  上的一个等价关系  $E_A^{[24]} : x E_A y \Leftrightarrow I_a(x) = I_a(y), \forall a \in A$ , 其中,  $x, y \in OB$ 。对  $\forall x \in OB$ , 关于  $x$  的等价类表示  $[x]_A = \{y \in OB \mid x E_A y\}$ 。

对  $\forall A \subseteq AT$ , 用  $OB/E_A = \{[x]_A \mid x \in OB\}$  表示由属性集  $A$  确定的  $OB$  上的一个划分。

人们对信息粒的研究早在粗糙集理论提出之前就已经开始了。从知识表示的角度, 信息表中的信息粒可分为可定义信息粒和不可定义信息粒。可定义信息粒也被称作可描述集<sup>[25-26]</sup>或者可定义集<sup>[27]</sup>。当前信息粒的研究仍然是一个热点问题<sup>[28]</sup>。

**定义 1<sup>[27]</sup>** 在一个信息表  $I = (OB, AT, \{V_a : a \in AT\}, \{I_a : a \in AT\})$  中, 对  $\forall P \subseteq OB/E_A$ , 称  $\cup P$  为一个可定义信息粒。所有可定义信息粒组成的集合称作可定义信息粒空间, 用符号  $DG_A$  表示, 即  $DG_A = \{\cup P \mid P \subseteq OB/E_A\}$ 。

**定义 2<sup>[27]</sup>** 在一个信息表  $I = (OB, AT, \{V_a : a \in AT\}, \{I_a : a \in AT\})$  中, 对  $\forall W \in 2^{OB/DG_A}$ , 称  $W$  是一个不可定义信息粒。所有不可定义信息粒组成的集合称作不可定义信息粒空间, 用符号  $UDG_A$  表示, 即  $UDG_A = 2^{OB/DG_A}$ 。

Pawlak 首先利用论域上的划分构造了上、下近似两个算子, 提出了经典的 Pawlak 粗糙集模型<sup>[7]</sup>。该模型已成为分析数据和提取规则的有效方法, 任何集合都可以用该集合的两个近似集来描述。

**定义 3<sup>[7]</sup>** 在一个信息表  $I = (OB, AT, \{V_a : a \in AT\}, \{I_a : a \in AT\})$  中,  $A \subseteq AT$  是一个属性集, 对  $\forall X \subseteq OB$ ,  $\underline{\text{apr}}_A(X) = \{x \in OB \mid [x]_A \subseteq X\}$ 、 $\overline{\text{apr}}_A(X) = \{x \in OB \mid [x]_A \cap X \neq \emptyset\}$  分别称为  $X$  关于  $A$  的下近似和上近似。

对  $\forall X \subseteq OB$ ,  $\underline{\text{apr}}_A(X) \subseteq X \subseteq \overline{\text{apr}}_A(X)$ 。如果  $\underline{\text{apr}}_A(X) = \overline{\text{apr}}_A(X)$ , 称  $X$  是精确的。如果  $\underline{\text{apr}}_A(X) \neq \overline{\text{apr}}_A(X)$ , 称  $X$  是粗糙的。而且,  $\underline{\text{apr}}_A(X)$  是所有包含于  $X$  的精确集中的最大者, 而  $\overline{\text{apr}}_A(X)$  是所有包含  $X$  的精确集中的最小者。

**定义 4<sup>[7]</sup>** 在一个信息表  $I = (OB, AT, \{V_a : a \in AT\}, \{I_a : a \in AT\})$  中,  $A \subseteq AT$  是一个属性集, 对  $\forall X \subseteq OB$ ,  $\alpha_A(X) = |\underline{\text{apr}}_A(X)| / |\overline{\text{apr}}_A(X)|$  称为  $X$  关于  $A$  的精度, 其中  $|\cdot|$  表示一个集合的基数。

$\alpha_A(X)$  描述了利用可定义信息粒表示  $X$  的精确性。显然,  $X$  是一个可定义信息粒, 当且仅当  $\alpha_A(X) = 1$  时, 即  $|\underline{\text{apr}}_{A,U}(X)| = |\overline{\text{apr}}_{A,U}(X)|$ 。

## 2 可变论域的划分及信息粒的数字特征

通常,人们把信息表中所有对象看作一个论域,进而作划分探讨其性质,可是在许多时候,根据实际问题和需要确定部分对象作为论域更为合理,即论域是可以变化的。本节主要介绍基于可变论域的划分及信息粒的数字特征。

在一个信息表  $I = (OB, AT, \{V_a: a \in AT\}, \{I_a: a \in AT\})$  中,对任意属性集  $A \subseteq AT (A \neq \emptyset)$  和任意论域  $U \subseteq OB (U \neq \emptyset)$ , 可以定义  $U$  上的一个等价关系  $E_{A,U}: x E_{A,U} y \Leftrightarrow I_a(x) = I_a(y), \forall a \in A$ , 其中  $x, y \in U$ 。对  $\forall x \in U$ , 关于  $x$  的等价类表示为  $[x]_{A,U} = \{y \in U \mid x E_{A,U} y\}$ 。对  $\forall A \subseteq AT$  和  $\forall U \subseteq OB$ , 可以得到由属性集  $A$  确定的论域  $U$  上的一个划分  $U/E_{A,U} = \{[x]_{A,U} \mid x \in U\}$ 。显然,  $U/E_{A,U} = (OB/E_A) \cap U = \{[x]_A \cap U \mid x \in U\}$ 。

**例 1** 在一个信息表  $I = (OB, AT, \{V_a: a \in AT\}, \{I_a: a \in AT\})$  中, 假设  $OB = \{x_1, x_2, \dots, x_9\}$ , 且  $OB/E_A = \{\{x_1, x_2, x_7, x_9\}, \{x_3, x_4, x_5, x_8\}, \{x_6\}\}$ 。令  $U = \{x_1, x_4, x_5\}$ , 则  $U/E_{A,U} = (OB/E_A) \cap U = \{\{x_1\}, \{x_4, x_5\}\}$ 。

下面给出基于可变论域的可定义信息粒的概念。

**定义 5** 在一个信息表  $I = (OB, AT, \{V_a: a \in AT\}, \{I_a: a \in AT\})$  中,  $A \subseteq AT$  是一个属性集,  $U \subseteq OB$  是一个论域, 对  $\forall P \subseteq U/E_{A,U}$ , 称  $UP$  为关于  $U$  的一个可定义信息粒。所有关于  $U$  的可定义信息粒组成的集合称作关于  $U$  的可定义信息粒空间, 用符号  $DG_{A,U}$  表示, 即  $DG_{A,U} = \{UP \mid P \subseteq U/E_{A,U}\}$ 。

同样, 基于可变论域的不可定义信息粒可做如下定义。

**定义 6** 在一个信息表  $I = (OB, AT, \{V_a: a \in AT\}, \{I_a: a \in AT\})$  中,  $A \subseteq AT$  是一个属性集,  $U \subseteq OB$  是一个论域, 对  $\forall W \in 2^{U/DG_{A,U}}$ , 称  $W$  是关于  $U$  的一个不可定义信息粒。所有关于  $U$  的不可定义信息粒组成的集合称作关于  $U$  的不可定义信息粒空间, 用符号  $UDG_{A,U}$  表示, 即  $UDG_{A,U} = 2^{U/DG_{A,U}}$ 。

在一个信息表  $I = (OB, AT, \{V_a: a \in AT\}, \{I_a: a \in AT\})$  中, 对任意论域  $U \subseteq OB$ ,  $U$  中的每个子集都是一个信息粒。在上面的分析中, 所有关于  $U$  的信息粒可以分为可定义信息粒和不可定义信息粒两部分。 $U$  中所有的子集的数目是  $2^{|U|}$ 。也就是说, 所有关于  $U$  的信息粒的个数是  $2^{|U|}$ 。下面给出一个信息表中对任意论域  $U \subseteq OB$  关于  $U$  的可定义信息粒和不可定义信息粒的数量。

**定理 1** 在一个信息表  $I = (OB, AT, \{V_a: a \in AT\}, \{I_a: a \in AT\})$  中,  $A \subseteq AT$  是一个属性集,  $U \subseteq OB$  是一个论域, 则: 1) 关于  $U$  的所有可定义信息粒的个数是  $2^{|U/E_{A,U}|}$ , 即  $|DG_{A,U}| = 2^{|U/E_{A,U}|}$ ; 2) 关于  $U$  的所有不可定义信息粒的个数是  $2^{|U|} - 2^{|U/E_{A,U}|}$ , 即  $|UDG_{A,U}| = 2^{|U|} - 2^{|U/E_{A,U}|}$ 。

由定理 1 可发现, 关于  $U$  的可定义信息粒的个数为  $2^{|U/E_{A,U}|}$ , 与  $U/E_{A,U}$  中等价类个数有关, 并与  $U$  中对象个数是无关的; 不可定义信息粒的个数为  $2^{|U|} - 2^{|U/E_{A,U}|}$ , 与  $U$  中对象个数和  $U/E_{A,U}$  中等价类个数都有关。

**例 2** 在例 1 中, 关于  $U$  的可定义信息粒有 4 个, 它们是  $\emptyset, \{x_1\}, \{x_4, x_5\}, U$ 。另一方面, 关于  $U$  的不可定义信息粒有 4 个, 分别是  $\{x_4\}, \{x_5\}, \{x_1, x_4\}, \{x_1, x_5\}$ 。

那么, 当信息表中论域  $U$  的对象个数增加或减少时, 关于  $U$  的可定义信息粒、不可定义信息粒的个数是如何变化的呢?

**定理 2** 在一个信息表  $I = (OB, AT, \{V_a: a \in AT\}, \{I_a: a \in AT\})$  中,  $A \subseteq AT$  是一个属性集,  $U_1, U_2 \subseteq OB$  是两个论域, 且  $U_1 \subseteq U_2$ , 则: 1)  $|DG_{A,U_1}| \leq |DG_{A,U_2}|$ ; 2)  $|UDG_{A,U_1}| \leq |UDG_{A,U_2}|$ 。

**证明** 1) 当  $U_1 \subseteq U_2$  时, 有  $|U_1/E_{A,U_1}| \leq |U_2/E_{A,U_2}|$ 。因为  $|DG_{A,U}| = 2^{|U/E_{A,U}|}$ , 那么  $2^{|U_1/E_{A,U_1}|} \leq 2^{|U_2/E_{A,U_2}|}$ , 即  $|DG_{A,U_1}| \leq |DG_{A,U_2}|$ 。

2) 对  $U_1 \subseteq U_2$ , 分两种情况进行讨论。

第一种情况, 当  $U_1 = U_2$  时, 显然有  $|UDG_{A,U_1}| = |UDG_{A,U_2}|$ 。

第二种情况, 当  $U_1 \subset U_2$  时, 令  $r = |U_2| - |U_1|$  和  $t = |U_2/E_{A,U_2}| - |U_1/E_{A,U_1}|$ , 则  $r \geq 1$ ,  $t \geq 0$ , 且  $r \geq t$ 。  $|\text{UDG}_{A,U_2}| - |\text{UDG}_{A,U_1}| = 2^{|U_2|} - 2^{|U_2/E_{A,U_2}|} - (2^{|U_1|} - 2^{|U_1/E_{A,U_1}|}) = 2^{|U_1|+r} - 2^{|U_1/E_{A,U_1}|+t} - 2^{|U_1|} + 2^{|U_1/E_{A,U_1}|} = (2^r - 1)[2^{|U_1|} - (2^t - 1)/(2^r - 1)2^{|U_1/E_{A,U_1}|}]$ 。

因为  $r \geq 1, t \geq 0$ , 且  $r \geq t$ , 那么  $2^r - 1 \geq 1, 0 \leq (2^t - 1)/(2^r - 1) \leq 1$ , 则  $|\text{UDG}_{A,U_2}| - |\text{UDG}_{A,U_1}| = (2^r - 1)[2^{|U_1|} - (2^t - 1)/(2^r - 1)2^{|U_1/E_{A,U_1}|}] \geq 2^{|U_1|} - (2^t - 1)/(2^r - 1)2^{|U_1/E_{A,U_1}|} \geq 2^{|U_1|} - 2^{|U_1/E_{A,U_1}|} \geq 0$ , 即  $|\text{UDG}_{A,U_1}| \leq |\text{UDG}_{A,U_2}|$ 。

由定理2可以看出, 当信息表中论域  $U$  的对象个数增加(减少)时, 关于  $U$  的可定义信息粒、不可定义信息粒的个数也增加(减少)。下面通过例3对定理2做进一步说明。

**例3** 在一个信息表  $I = (\text{OB}, \text{AT}, \{V_a: a \in \text{AT}\}, \{I_a: a \in \text{AT}\})$  中, 假设  $\text{OB} = \{x_1, x_2, \dots, x_9\}$ , 且  $\text{OB}/E_A = \{\{x_1, x_2, x_7, x_9\}, \{x_3, x_4, x_5, x_8\}, \{x_6\}\}$ 。令  $U_1 = \{x_4, x_5, x_6\}$ ,  $U_2 = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ ,  $U_1 \subseteq U_2$ , 则:  $U_1/E_{A,U_1} = \{\{x_4, x_5\}, \{x_6\}\}$ ,  $U_2/E_{A,U_2} = \{\{x_1, x_2\}, \{x_3, x_4, x_5\}, \{x_6\}\}$ 。

由定理1可知,  $|\text{DG}_{A,U_1}| = 2^2 = 4$ ,  $|\text{DG}_{A,U_2}| = 2^3 = 8$ ;  $|\text{UDG}_{A,U_1}| = 2^3 - 2^2 = 4$ ,  $|\text{UDG}_{A,U_2}| = 2^6 - 2^3 = 56$ 。从而有  $|\text{DG}_{A,U_1}| < |\text{DG}_{A,U_2}|$ ,  $|\text{UDG}_{A,U_1}| < |\text{UDG}_{A,U_2}|$ 。

接下来讨论关于  $U$  的信息粒在交、并、补运算下的结构性质。

**定理3** 在一个信息表  $I = (\text{OB}, \text{AT}, \{V_a: a \in \text{AT}\}, \{I_a: a \in \text{AT}\})$  中,  $A \subseteq \text{AT}$  是一个属性集,  $U \subseteq \text{OB}$  是一个论域, 则关于  $U$  的可定义信息粒在交、并、补运算下是封闭的。

**定理4** 在一个信息表  $I = (\text{OB}, \text{AT}, \{V_a: a \in \text{AT}\}, \{I_a: a \in \text{AT}\})$  中,  $A \subseteq \text{AT}$  是一个属性集,  $U \subseteq \text{OB}$  是一个论域, 则关于  $U$  的不可定义信息粒在补运算下是封闭的。

**证明** 对  $W \in \text{UDG}_{A,U}$ , 假设  $W^c = U - W$  是一个关于  $U$  的可定义信息粒。根据定理3, 关于  $U$  的可定义信息粒在补运算下是封闭的, 那么  $(W^c)^c = W$  是一个关于  $U$  的可定义信息粒。这与假设矛盾, 即关于  $U$  的不可定义信息粒的补集仍然是关于  $U$  的不可定义信息粒。

那么关于  $U$  的不可定义信息粒在交、并运算下是否封闭呢? 下面的例4对此做出说明。

**例4** 在一个信息表  $I = (\text{OB}, \text{AT}, \{V_a: a \in \text{AT}\}, \{I_a: a \in \text{AT}\})$  中, 假设  $\text{OB} = \{x_1, x_2, \dots, x_9\}$ , 且  $\text{OB}/E_A = \{\{x_1, x_2, x_7, x_9\}, \{x_3, x_4, x_5, x_8\}, \{x_6\}\}$ 。令  $U = \{x_1, x_4, x_5, x_6\}$ , 则  $U/E_{A,U} = \{\{x_1\}, \{x_4, x_5\}, \{x_6\}\}$ 。

关于  $U$  的两个不可定义信息粒  $X_1 = \{x_1, x_4\}$  和  $X_2 = \{x_4, x_6\}$ , 可知  $X_1 \cap X_2 = \{x_4\}$  和  $X_1 \cup X_2 = \{x_1, x_4, x_6\}$  都是关于  $U$  的不可定义信息粒, 而对关于  $U$  的两个不可定义信息粒  $X_3 = \{x_1, x_4\}$  和  $X_4 = \{x_1, x_5\}$ , 不难发现,  $X_3 \cap X_4 = \{x_1\}$  和  $X_3 \cup X_4 = \{x_1, x_4, x_5\}$  都是关于  $U$  的可定义信息粒。

例4说明, 关于  $U$  的不可定义信息粒在交、并运算下是不封闭的。

为了更好地理解一个  $n$  维实向量空间  $V_n = \{\{v_i, v_2, \dots, v_n\} | v_i \in \mathbf{R}, i = 1, 2, \dots, n\}$ , 人们在向量的线性运算的基础上引入了向量空间的基<sup>[29]</sup>这一概念。例如,  $\{V_1 = \{1, 0, 0, \dots, 0\}, V_2 = \{0, 1, 0, \dots, 0\}, \dots, V_n = \{0, 0, 0, \dots, 1\}\}$  是  $V_n$  的一个基空间。 $V_n$  中的任何向量都可以由基空间中的向量通过线性运算表示出来, 即对任意  $V = \{v_1, v_2, \dots, v_n\} \in V_n$ , 有  $V = v_1 \cdot V_1 + v_2 \cdot V_2 + \dots + v_n \cdot V_n$ , 且  $n$  维实向量空间的基不是唯一的。

那么在一个信息表  $I = (\text{OB}, \text{AT}, \{V_a: a \in \text{AT}\}, \{I_a: a \in \text{AT}\})$  中, 对任意论域  $U \subseteq \text{OB}$ 。结合定理3, 对关于  $U$  的可定义信息粒空间在交、并、补运算下基的形式和基的数量进行了研究。首先给出关于  $U$  的可定义信息粒空间在交、并、补运算下基的定义, 再讨论关于  $U$  的可定义信息粒空间的基、维数等。

**定义7** 在一个信息表  $I = (\text{OB}, \text{AT}, \{V_a: a \in \text{AT}\}, \{I_a: a \in \text{AT}\})$  中,  $A \subseteq \text{AT}$  是一个属性集,  $U \subseteq \text{OB}$  是一个论域, 且  $\text{DG}_{A,U}$  是由  $A$  确定的关于  $U$  的可定义信息粒空间, 若存在  $\text{BDG}_{A,U} \subseteq \text{DG}_{A,U}$  满足下列条件: 1) 对  $\forall X \in \text{DG}_{A,U}$ , 存在  $\text{BDG}'_{A,U} \subseteq \text{BDG}_{A,U}$ , 使得  $X$  可由  $\text{BDG}'_{A,U}$  中的信息粒通过并、交、

补运算表示出来; 2) 对  $\forall Y \in \text{BDG}_{A,U}$ , 不存在  $\text{BDG}_{A,U}'' \subseteq \text{BDG}_{A,U}/\{Y\}$ , 使得  $Y$  可由  $\text{BDG}_{A,U}''$  中的信息粒通过并、交、补运算表示出来, 则称  $\text{BDG}_{A,U}$  为  $\text{DG}_{A,U}$  在交、并、补运算下的一个基空间。

**定理 5** 在一个信息表  $I = (\text{OB}, \text{AT}, \{V_a: a \in \text{AT}\}, \{I_a: a \in \text{AT}\})$  中,  $A \subseteq \text{AT}$  是一个属性集,  $U \subseteq \text{OB}$  是一个论域。令  $U/E_{A,U} = \{E_1, E_2, \dots, E_{|U/E_{A,U}|}\}$ , 则: 1)  $\text{BDG}_{A,U} = \{E_1, E_2, \dots, E_{i-1}, E_{i+1}, \dots, E_{|U/E_{A,U}|}\}$ ,  $i \in \{1, 2, \dots, |U/E_{A,U}|\}$ ; 2)  $|\text{BDG}_{A,U}| = |U/E_{A,U}| - 1$ , 即  $\text{DG}_{A,U}$  是  $|U/E_{A,U}| - 1$  维的; 3)  $\text{DG}_{A,U}$  的基空间有  $|U/E_{A,U}|$  个。

由定理 5 发现, 关于  $U$  的可定义信息粒空间的基空间个数为  $|U/E_{A,U}|$ , 并且基空间的维数为  $|U/E_{A,U}| - 1$ , 都与  $U/E_{A,U}$  中等价类个数有关, 与  $U$  中对象个数无关。

**例 5** 根据例 2, 有  $\text{DG}_{A,U} = \{\emptyset, \{x_1\}, \{x_4, x_5\}, U\}$ 。由定理 5 可知,  $\text{BDG}_{A,U,1} = \{\{x_1\}\}$  和  $\text{BDG}_{A,U,2} = \{\{x_4, x_5\}\}$  都是  $\text{DG}_{A,U}$  在交、并、补运算下的基空间, 因此,  $\text{DG}_{A,U}$  的基空间是不唯一的, 且  $\text{DG}_{A,U}$  是 1 维的。

同样, 结合定理 4, 对可变论域的不可定义信息粒空间在补运算下基的形式和基的数量进行了研究。下面首先给出可变论域的不可定义信息粒空间在补运算下基的定义, 然后再讨论可变论域的不可定义信息粒空间的基、维数等。

**定义 8** 在一个信息表  $I = (\text{OB}, \text{AT}, \{V_a: a \in \text{AT}\}, \{I_a: a \in \text{AT}\})$  中,  $A \subseteq \text{AT}$  是一个属性集,  $U \subseteq \text{OB}$  是一个论域, 且  $\text{UDG}_{A,U}$  是由  $A$  确定的关于  $U$  的不可定义信息粒空间, 若存在  $\text{BUDG}_{A,U} \subseteq \text{UDG}_{A,U}$  满足下列条件: 1) 对任意  $X \in \text{UDG}_{A,U}$ , 存在  $X' \in \text{BUDG}_{A,U}$ , 使得  $X = (X')^c$  或  $X = ((X')^c)^c$ , 其中  $(X')^c = U - X'$ ; 2) 对任意  $Y \in \text{BUDG}_{A,U}$ , 不存在  $Y' \in \text{BUDG}_{A,U}/\{Y\}$ , 使得  $Y = (Y')^c$ , 其中  $(Y')^c = U - Y'$ , 则称  $\text{BUDG}_{A,U}$  为  $\text{UDG}_{A,U}$  在补运算下的一个基空间。

**定理 6** 在一个信息表  $I = (\text{OB}, \text{AT}, \{V_a: a \in \text{AT}\}, \{I_a: a \in \text{AT}\})$  中,  $A \subseteq \text{AT}$  是一个属性集,  $U \subseteq \text{OB}$  是一个论域。令  $U/E_{A,U} = \{E_1, E_2, \dots, E_{|U/E_{A,U}|}\}$ , 则: 1)  $\text{UDG}_{A,U} = \{X_1, X_1^c, X_2, X_2^c, \dots, X_t, X_t^c\}$ ,  $t = (2^{|U|} - 2^{|U/E_{A,U}|})/2$ , 其中  $X^c = U - X$ ; 2)  $\text{BUDG}_{A,U} = \{X'_1, X'_2, \dots, X'_t\}$ , 其中  $X'_i = X_i$  或  $X_i^c$ ,  $i = 1, 2, \dots, t$ ; 3)  $|\text{BUDG}_{A,U}| = (2^{|U|} - 2^{|U/E_{A,U}|})/2$ , 即  $\text{UDG}_{A,U}$  是  $(2^{|U|} - 2^{|U/E_{A,U}|})/2$  维的; 4)  $\text{UDG}_{A,U}$  的基空间有  $2^{(2^{|U|} - 2^{|U/E_{A,U}|})/2}$  个。

由定理 6 发现, 关于  $U$  的不可定义信息粒空间的基空间个数为  $2^{(2^{|U|} - 2^{|U/E_{A,U}|})/2}$ , 并且基空间的维数为  $(2^{|U|} - 2^{|U/E_{A,U}|})/2$ , 与  $U$  中对象个数和  $U/E_{A,U}$  中等价类个数都有关。

**例 6** 根据例 2, 有  $\text{UDG}_{A,U} = \{\{x_4\}, \{x_1, x_5\}, \{x_5\}, \{x_1, x_4\}\}$ 。由定理 6 可知,  $\text{BUDG}_{A,U,1} = \{\{x_4\}, \{x_5\}\}$ 、 $\text{BUDG}_{A,U,2} = \{\{x_4\}, \{x_1, x_4\}\}$ 、 $\text{BUDG}_{A,U,3} = \{\{x_1, x_5\}, \{x_5\}\}$ 、 $\text{BUDG}_{A,U,4} = \{\{x_1, x_5\}, \{x_1, x_4\}\}$  都是  $\text{UDG}_{A,U}$  在补运算下的基空间, 因此,  $\text{UDG}_{A,U}$  的基空间是不唯一的, 并且  $\text{UDG}_{A,U}$  是 2 维的。

同样, 当信息表中论域  $U$  的对象个数增加或减少时, 关于  $U$  的可定义信息粒空间和不可定义信息粒空间的基空间中信息粒个数是如何变化的呢?

**定理 7** 在一个信息表  $I = (\text{OB}, \text{AT}, \{V_a: a \in \text{AT}\}, \{I_a: a \in \text{AT}\})$  中,  $A \subseteq \text{AT}$  是一个属性集,  $U_1, U_2 \subseteq \text{OB}$  是两个论域, 且  $U_1 \subseteq U_2$ , 则: 1)  $|\text{BDG}_{A,U_1}| \leq |\text{BDG}_{A,U_2}|$ ; 2)  $|\text{BUDG}_{A,U_1}| \leq |\text{BUDG}_{A,U_2}|$ 。

**证明** 1) 当  $U_1 \subseteq U_2$  时, 有  $|U_1/E_{A,U_1}| \leq |U_2/E_{A,U_2}|$ 。因为  $|\text{BDG}_{A,U}| = |U/E_{A,U}| - 1$ , 那么  $|\text{BDG}_{A,U_1}| \leq |\text{BDG}_{A,U_2}|$ 。

2) 当  $U_1 \subseteq U_2$  时, 由定理 1 和定理 6 可知, 对  $\forall U \subseteq \text{OB}$ ,  $|\text{BUDG}_{A,U}| = (2^{|U|} - 2^{|U/E_{A,U}|})/2 = |\text{UDG}_{A,U}|/2$ 。由定理 2 可知,  $|\text{UDG}_{A,U_1}| \leq |\text{UDG}_{A,U_2}|$ 。显然,  $|\text{BUDG}_{A,U_1}| \leq |\text{BUDG}_{A,U_2}|$ 。

由定理 7 可以看出, 当信息表中论域  $U$  的对象个数增加 (减少) 时, 关于  $U$  的可定义信息粒空间和不可定义信息粒空间的基空间中信息粒个数也增加 (减少)。



例7 在例3中, 由定理5可知,  $|\text{BDG}_{A,U_1}| = 1$ ,  $|\text{BDG}_{A,U_2}| = 2$ 。由定理6可得,  $|\text{BUDG}_{A,U_1}| = (2^4 - 2^2)/2 = 6$ ,  $|\text{BUDG}_{A,U_1}| = (2^6 - 2^3)/2 = 28$ , 从而有  $|\text{BDG}_{A,U_1}| < |\text{BDG}_{A,U_2}|$ ,  $|\text{BUDG}_{A,U_1}| < |\text{UDG}_{A,U_2}|$ 。

### 3 可变论域粗糙集模型的数字特征

定义9<sup>[22]</sup> 在一个信息表  $I = (\text{OB}, \text{AT}, \{V_a: a \in \text{AT}\}, \{I_a: a \in \text{AT}\})$  中,  $A \subseteq \text{AT}$  是一个属性集,  $U \subseteq \text{OB}$  是一个论域, 则对  $\forall X \subseteq U$ ,  $\text{apr}_{A,U}(X) = \cup \{[x]_{A,U} \mid ([x]_{A,U} \subseteq X) \wedge (x \in U)\}$ ,  $\overline{\text{apr}}_{A,U}(X) = \cup \{[x]_{A,U} \mid ([x]_{A,U} \cap X \neq \emptyset) \wedge (x \in U)\}$  分别称为  $X$  关于  $U$  和  $A$  的下近似和上近似。为了突出本文研究的是可变论域模型这一特点, 下文在不影响理解的情况下, 把“ $X$  关于  $U$  和  $A$  的下近似和上近似”简称为“ $X$  关于  $U$  的下近似和上近似”。

显然, 当  $U = \text{OB}$  时, 定义9中的可变论域粗糙集模型就是经典的 Pawlak 粗糙集模型。

由定义9可以看出, 关于  $U$  的上、下近似是由  $U/E_{A,U}$  中若干个等价类的并定义的。令  $\text{APR}_{A,U}$ 、 $\overline{\text{APR}}_{A,U}$  分别为信息表中关于  $U$  的所有下近似和上近似的全体, 则有  $\text{APR}_{A,U} = \overline{\text{APR}}_{A,U} = \text{DG}_{A,U}$ , 由定理1, 即  $|\text{APR}_{A,U}| = |\overline{\text{APR}}_{A,U}| = 2^{|U/E_{A,U}|}$ , 且关于  $U$  的上、下近似和关于  $U$  的可定义信息粒具有相同的数字特征。

对任意  $X \subseteq U$ ,  $X$  可由  $\text{apr}_{A,U}(X)$ 、 $\overline{\text{apr}}_{A,U}(X)$  描述。同样, 称  $(\text{apr}_{A,U}(X), \overline{\text{apr}}_{A,U}(X))$  为  $X$  关于  $U$  的粗糙集, 且  $\{(\text{apr}_{A,U}(X), \overline{\text{apr}}_{A,U}(X)) \mid X \subseteq U\}$  为关于  $U$  的粗糙集全体, 记为  $\text{RS}_{A,U}$ 。

接下来考虑在一个信息表  $I = (\text{OB}, \text{AT}, \{V_a: a \in \text{AT}\}, \{I_a: a \in \text{AT}\})$  中, 对任意论域  $U \subseteq \text{OB}$  关于  $U$  的粗糙集的个数。

定理8 在一个信息表  $I = (\text{OB}, \text{AT}, \{V_a: a \in \text{AT}\}, \{I_a: a \in \text{AT}\})$  中,  $A \subseteq \text{AT}$  是一个属性集,  $U \subseteq \text{OB}$  是一个论域。令  $U/E_{A,U} = \{E_{11}, E_{12}, \dots, E_{1m}, E_{21}, E_{22}, \dots, E_{2n}\}$ , 其中  $E_{11}, E_{12}, \dots, E_{1m}$  均为单点集,  $E_{21}, E_{22}, \dots, E_{2n}$  均为非单点集, 则  $|\text{RS}_{A,U}| = \sum_{i=0}^m \sum_{j=0}^n C_m^i \cdot C_n^j \cdot 2^{n-j}$ 。

证明 对  $\forall (\text{apr}_{A,U}(X), \overline{\text{apr}}_{A,U}(X)) \in \text{RS}_{A,U}$ , 并且假设  $\overline{\text{apr}}_{A,U}(X)$  满足条件

$$\overline{\text{apr}}_{A,U}(X) = (\cup_{E \in P} E) \cup (\cup_{E' \in Q} E'), \quad (1)$$

其中:  $P \subseteq \{E_{11}, E_{12}, \dots, E_{1m}\}$ ;  $|P| = i, i \in \{0, 1, \dots, m\}$ ;  $Q \subseteq \{E_{21}, E_{22}, \dots, E_{2n}\}$ ,  $|Q| = j, j \in \{0, 1, \dots, n\}$ 。显然, 满足条件(1)的粗糙集个数等于  $C_m^i \cdot C_n^j \cdot 2^{n-j}$ 。因为  $i \in \{0, 1, \dots, m\}$ ,  $j \in \{0, 1, \dots, n\}$ , 那么所有粗糙集的个数是  $\sum_{i=0}^m \sum_{j=0}^n C_m^i \cdot C_n^j \cdot 2^{n-j}$ 。

例8 在例1中,  $U/E_{A,U}$  中单点集有1个, 非单点集有1个。由定理8可知,  $|\text{RS}_{A,U}| = C_1^0 \cdot C_1^0 \cdot 2^{1-0} + C_1^0 \cdot C_1^1 \cdot 2^{1-1} + C_1^1 \cdot C_1^0 \cdot 2^{1-0} + C_1^1 \cdot C_1^1 \cdot 2^{1-1} = 6$ , 且所有关于  $U$  的粗糙集分别为  $(\emptyset, \emptyset), (\emptyset, \{x_4, x_5\}), (\{x_1\}, \{x_1\}), (\{x_4, x_5\}, \{x_4, x_5\}), (\{x_1\}, U), (U, U)$ 。

那么, 当信息表中论域  $U$  的对象个数增加或减少时, 关于  $U$  的粗糙集的个数是如何变化的呢?

定理9 在一个信息表  $I = (\text{OB}, \text{AT}, \{V_a: a \in \text{AT}\}, \{I_a: a \in \text{AT}\})$  中,  $A \subseteq \text{AT}$  是一个属性集,  $U_1, U_2 \subseteq \text{OB}$  是两个论域, 且  $U_1 \subseteq U_2$ , 则  $|\text{RS}_{A,U_1}| \leq |\text{RS}_{A,U_2}|$ 。

由定理9可以看出, 当信息表中论域  $U$  的对象个数增加(减少)时, 关于  $U$  的粗糙集的个数也增加(减少)。下面通过例9对定理9做进一步的说明。

例9 在例3中,  $U_1/E_{A,U_1}$  中单点集有1个, 非单点集有1个;  $U_2/E_{A,U_2}$  中单点集有1个, 非单点集有2个。由定理8可知,  $|\text{RS}_{A,U_2}| = C_1^0 \cdot C_1^0 \cdot 2^1 + C_1^0 \cdot C_1^1 \cdot 2^0 + C_1^1 \cdot C_1^0 \cdot 2^1 + C_1^1 \cdot C_1^1 \cdot 2^0 = 6$ ,

$|\text{RS}_{A,U_2}| = C_1^0 \cdot C_2^0 \cdot 2^2 + C_1^0 \cdot C_2^1 \cdot 2^1 + C_1^0 \cdot C_2^2 \cdot 2^0 + C_1^1 \cdot C_2^0 \cdot 2^2 + C_1^1 \cdot C_2^1 \cdot 2^1 + C_1^1 \cdot C_2^2 \cdot 2^0 = 18$ , 从而有  $|\text{RS}_{A,U_1}| < |\text{RS}_{A,U_2}|$ 。

范世栋等<sup>[30]</sup>给出了粗糙集在交、并、补运算下的表示定义。在此基础上, 对关于  $U$  的粗糙集模型引入相应的概念。

**定义 10** 在一个信息表  $I = (\text{OB}, \text{AT}, \{V_a: a \in \text{AT}\}, \{I_a: a \in \text{AT}\})$  中,  $A \subseteq \text{AT}$  是一个属性集,  $U \subseteq \text{OB}$  是一个论域, 对  $\forall X, Y \subseteq U$ ,  $(\underline{\text{apr}}_{A,U}(X), \overline{\text{apr}}_{A,U}(X))$  和  $(\underline{\text{apr}}_{A,U}(Y), \overline{\text{apr}}_{A,U}(Y))$  交集、并集的定义为:  $(\underline{\text{apr}}_{A,U}(X), \overline{\text{apr}}_{A,U}(X)) \cap (\underline{\text{apr}}_{A,U}(Y), \overline{\text{apr}}_{A,U}(Y)) = (\underline{\text{apr}}_{A,U}(X) \cap \underline{\text{apr}}_{A,U}(Y), \overline{\text{apr}}_{A,U}(X) \cap \overline{\text{apr}}_{A,U}(Y))$ ,  $(\underline{\text{apr}}_{A,U}(X), \overline{\text{apr}}_{A,U}(X)) \cup (\underline{\text{apr}}_{A,U}(Y), \overline{\text{apr}}_{A,U}(Y)) = ((\underline{\text{apr}}_{A,U}(X) \cup \underline{\text{apr}}_{A,U}(Y)), (\overline{\text{apr}}_{A,U}(X) \cup \overline{\text{apr}}_{A,U}(Y)))$ 。

根据定义 10, 有定理 10。

**定理 10** 在一个信息表  $I = (\text{OB}, \text{AT}, \{V_a: a \in \text{AT}\}, \{I_a: a \in \text{AT}\})$  中,  $A \subseteq \text{AT}$  是一个属性集,  $U \subseteq \text{OB}$  是一个论域, 则关于  $U$  的粗糙集在交、并运算下是封闭的。

**证明** 对  $\forall X, Y \subseteq U$ , 令  $W = (X \cup \{x \mid [x]_{A,U} \cap (X \cap Y) = \emptyset, \text{ 且 } [x]_{A,U} \cap X \neq \emptyset, [x]_{A,U} \cap Y \neq \emptyset\}) \cap Y$ ,  $Z = (X - \{x \mid [x]_{A,U} \subseteq X \cap Y, \text{ 且 } [x]_{A,U} \not\subseteq X, [x]_{A,U} \not\subseteq Y\}) \cup Y$ 。则有  $(\underline{\text{apr}}_{A,U}(W), \overline{\text{apr}}_{A,U}(W)) = (\underline{\text{apr}}_{A,U}(X), \overline{\text{apr}}_{A,U}(X)) \cap (\underline{\text{apr}}_{A,U}(Y), \overline{\text{apr}}_{A,U}(Y))$ ,  $(\underline{\text{apr}}_{A,U}(Z), \overline{\text{apr}}_{A,U}(Z)) = (\underline{\text{apr}}_{A,U}(X), \overline{\text{apr}}_{A,U}(X)) \cup (\underline{\text{apr}}_{A,U}(Y), \overline{\text{apr}}_{A,U}(Y))$ 。

**定义 11** 在一个信息表  $I = (\text{OB}, \text{AT}, \{V_a: a \in \text{AT}\}, \{I_a: a \in \text{AT}\})$  中,  $A \subseteq \text{AT}$  是一个属性集,  $U \subseteq \text{OB}$  是一个论域, 对  $\forall X \subseteq U$ ,  $(\underline{\text{apr}}_{A,U}(X), \overline{\text{apr}}_{A,U}(X))$  的补集定义为  $(\underline{\text{apr}}_{A,U}(X), \overline{\text{apr}}_{A,U}(X))^c = (\underline{\text{apr}}_{A,U}(X^c), \overline{\text{apr}}_{A,U}(X^c))$ 。

由定理 10 和定义 11 可知, 基于可变论域的粗糙集在交、并、补运算下是封闭的。下面自然可以定义可变论域的粗糙集空间在交、并、补运算下的基。

**定义 12** 在一个信息表  $I = (\text{OB}, \text{AT}, \{V_a: a \in \text{AT}\}, \{I_a: a \in \text{AT}\})$  中,  $A \subseteq \text{AT}$  是一个属性集,  $U \subseteq \text{OB}$  是一个论域, 若存在  $\text{BRS}_{A,U} \subseteq \text{RS}_{A,U}$  满足下列条件: 1) 对任意  $(\underline{\text{apr}}_{A,U}(X), \overline{\text{apr}}_{A,U}(X)) \in \text{RS}_{A,U}$ , 存在  $\text{BRS}'_{A,U} \subseteq \text{BRS}_{A,U}$ , 使得  $(\underline{\text{apr}}_{A,U}(X), \overline{\text{apr}}_{A,U}(X))$  可由  $\text{BRS}'_{A,U}$  中的粗糙集通过交、并、补运算表示出来; 2) 对任意  $(\underline{\text{apr}}_{A,U}(Y), \overline{\text{apr}}_{A,U}(Y)) \in \text{BRS}_{A,U}$ , 不存在  $\text{BRS}''_{A,U} \subseteq \text{BRS}_{A,U} \setminus \{(\underline{\text{apr}}_{A,U}(Y), \overline{\text{apr}}_{A,U}(Y))\}$ , 使得  $(\underline{\text{apr}}_{A,U}(Y), \overline{\text{apr}}_{A,U}(Y))$  可由  $\text{BRS}''_{A,U}$  中的粗糙集通过交、并、补运算表示出来, 则称  $\text{BRS}_{A,U}$  为  $\text{RS}_{A,U}$  在交、并、补运算下的一个基空间。

**定理 11** 在一个信息表  $I = (\text{OB}, \text{AT}, \{V_a: a \in \text{AT}\}, \{I_a: a \in \text{AT}\})$  中,  $A \subseteq \text{AT}$  是一个属性集,  $U \subseteq \text{OB}$  是一个论域,  $U/E_{A,U} = \{E_{11}, E_{12}, \dots, E_{1m}, E_{21}, E_{22}, \dots, E_{2n}\}$ , 其中  $E_{11}, E_{12}, \dots, E_{1m}$  均为单点集,  $E_{21}, E_{22}, \dots, E_{2n}$  均为非单点集, 且  $\text{BRS}'_{A,U} = \{(\emptyset, E_{21}), (\emptyset, E_{22}), \dots, (\emptyset, E_{2n})\}$ ,  $\text{BRS}''_{A,U} = \{(E_{11}, E_{11}), (E_{12}, E_{12}), \dots, (E_{1m}, E_{1m}), (E_{21}, E_{21}), (E_{22}, E_{22}), \dots, (E_{2n}, E_{2n})\} \setminus \{(\{E\}, \{E\})\}$ , 其中  $E \in U/E_{A,U}$ , 则: 1)  $\text{BRS}_{A,U} = \text{BRS}'_{A,U} \cup \text{BRS}''_{A,U}$ ; 2)  $|\text{BRS}_{A,U}| = 2n + m - 1$ , 即  $\text{RS}_{A,U}$  是  $2n + m - 1$  维的; 3)  $\text{RS}_{A,U}$  的基空间有  $m + n$  个。

由定理 11 发现, 关于  $U$  的粗糙集空间的基空间个数为  $m + n$ , 并且基空间的维数为  $2n + m - 1$ , 与  $U$  中对象个数有关, 其中  $m, n$  分别表示  $U/E_{A,U}$  中单点集和非单点集的个数。

**例 10** 根据例 8, 可知  $\text{RS}_{A,U} = \{(\emptyset, \emptyset), (\emptyset, \{x_4, x_5\}), (\{x_1\}, \{x_1\}), (\{x_4, x_5\}, \{x_4, x_5\}), (\{x_1\}, U), (U, U)\}$ 。由定理 11,  $\text{BRS}_{A,U,1} = \{(\emptyset, \{x_4, x_5\}), (\{x_4, x_5\}, \{x_4, x_5\})\}$ ,  $\text{BRS}_{A,U,2} = \{(\emptyset, \{x_4, x_5\}), (\{x_1\}, \{x_1\})\}$  都是  $\text{RS}_{A,U}$  在交、并、补运算下的基空间, 因此  $\text{RS}_{A,U}$  的基空间是不唯一的,

且  $RS_{A,U}$  是 2 维的。

同样, 当信息表中论域  $U$  的对象个数增加或减少时, 关于  $U$  的粗糙集空间的基空间中粗糙集的个数是如何变化的呢?

**定理 12** 在一个信息表  $I = (OB, AT, \{V_a: a \in AT\}, \{I_a: a \in AT\})$  中,  $A \subseteq AT$  是一个属性集,  $U_1, U_2 \subseteq OB$  是两个论域, 且  $U_1 \subseteq U_2$ , 则  $|BRS_{A,U_1}| \leq |BRS_{A,U_2}|$ 。

**证明** 当  $U_1 \subseteq U_2$  时, 有  $|U_1/E_{A,U_1}| \leq |U_2/E_{A,U_2}|$ , 且  $U_2/E_{A,U_2}$  比  $U_1/E_{A,U_1}$  中的非单点集个数只会增多或不变。由定理 11, 可得  $|BRS_{A,U_1}| \leq |BRS_{A,U_2}|$ 。

由定理 12 可以看出, 当信息表中论域  $U$  的对象个数增加 (减少) 时, 关于  $U$  的粗糙集空间的基空间中粗糙集的个数也增加 (减少)。

**例 11** 在例 3 中, 由定理 11 可知,  $|BRS_{A,U_1}| = 2 + 1 - 1 = 2$ ,  $|BRS_{A,U_2}| = 4 + 1 - 1 = 4$ , 从而有  $|BRS_{A,U_1}| < |BRS_{A,U_2}|$ 。

接下来, 在 Pawlak 模型中信息粒精度的基础上, 对关于  $U$  的粗糙集模型引入精度的概念。

**定义 13** 在一个信息表  $I = (OB, AT, \{V_a: a \in AT\}, \{I_a: a \in AT\})$  中,  $A \subseteq AT$  是一个属性集,  $U \subseteq OB$  是一个论域, 则对  $\forall X \subseteq U$ ,  $\alpha_{A,U}(X) = |\underline{\text{apr}}_{A,U}(X)| / |\overline{\text{apr}}_{A,U}(X)|$ , 称为  $X$  关于  $U$  和  $A$  的粗糙集的精度, 其中  $|\cdot|$  表示一个集合的基数。为了突出本文研究的是可变论域模型这一特点, 下文在不影响理解的情况下, 把“ $X$  关于  $U$  和  $A$  的粗糙集的精度”简称为“ $X$  关于  $U$  的粗糙集的精度”。

显然, 对  $\forall X \subseteq U$ , 有  $0 \leq \alpha_{A,U}(X) \leq 1$ 。为了讨论关于  $U$  的粗糙集的精度, 首先引入引理 1。

**引理 1**<sup>[22]</sup> 在一个信息表  $I = (OB, AT, \{V_a: a \in AT\}, \{I_a: a \in AT\})$  中,  $A \subseteq AT$  是一个属性集,  $U_1, U_2 \subseteq OB$  是两个论域, 且  $U_1 \subseteq U_2$ , 则对  $\forall X \subseteq U_1$ , 1)  $\underline{\text{apr}}_{A,U_2}(X) \subseteq \underline{\text{apr}}_{A,U_1}(X)$ ; 2)  $\overline{\text{apr}}_{A,U_1}(X) \subseteq \overline{\text{apr}}_{A,U_2}(X)$ 。

根据引理 1, 有定理 13。

**定理 13** 在一个信息表  $I = (OB, AT, \{V_a: a \in AT\}, \{I_a: a \in AT\})$  中,  $A \subseteq AT$  是一个属性集,  $U_1, U_2 \subseteq OB$  是两个论域, 且  $U_1 \subseteq U_2$ , 则对  $\forall X \subseteq U_1$ ,  $\alpha_{A,U_2}(X) \leq \alpha_{A,U_1}(X)$ 。

**证明** 由引理 1,  $\alpha_{A,U_2}(X) = |\underline{\text{apr}}_{A,U_2}(X)| / |\overline{\text{apr}}_{A,U_2}(X)| \leq |\underline{\text{apr}}_{A,U_1}(X)| / |\overline{\text{apr}}_{A,U_1}(X)| = \alpha_{A,U_1}(X)$ , 即  $\alpha_{A,U_2}(X) \leq \alpha_{A,U_1}(X)$ 。

由定理 13 可知, 当信息表中论域  $U$  的对象个数增加 (减少) 时, 关于  $U$  的粗糙集的精度减小 (增大)。

**例 12** 在例 3 中, 令  $X = \{x_4, x_5, x_6\} \subseteq U_1$ , 有:  $\underline{\text{apr}}_{A,U_1}(X) = \{x_4, x_5, x_6\}$ ,  $\overline{\text{apr}}_{A,U_1}(X) = \{x_4, x_5, x_6\}$ ;  $\underline{\text{apr}}_{A,U_2}(X) = \{x_6\}$ ,  $\overline{\text{apr}}_{A,U_2}(X) = \{x_3, x_4, x_5, x_6\}$ ;  $\alpha_{A,U_1}(X) = |\underline{\text{apr}}_{A,U_1}(X)| / |\overline{\text{apr}}_{A,U_1}(X)| = 1$ ,  $\alpha_{A,U_2}(X) = |\underline{\text{apr}}_{A,U_2}(X)| / |\overline{\text{apr}}_{A,U_2}(X)| = 1/4$ , 从而有  $\alpha_{A,U_2}(X) < \alpha_{A,U_1}(X)$ 。

## 4 结论

本文在可变论域概念的基础上, 深入研究了可变论域信息粒和粗糙集模型的数字特征, 得到许多有意义的结果。当可变论域的对象个数增加时, 可定义信息粒、不可定义信息粒的个数, 以及可定义信息粒空间、不可定义信息粒空间的基空间中信息粒的个数和维数也增加; 并且在粗糙集模型中, 粗糙集的个数、粗糙集空间的基空间中粗糙集的个数和维数随可变论域中对象个数的增加而增加, 信息粒的精度随可变论域中对象个数的增加而减小。也就是说, 论域中对象个数的变化会影响信息表中的数字特征, 从微观层面上也让人们对信息表中信息粒的数字特征有了更详细和深刻的认识。今后, 也可以根据可变论域的思想, 进一步讨论其他粗糙集模型上、下近似的性质、属性约简、规则提取等内容。



## [ 参考文献 ]

- [1] 苗夺谦,范世栋. 知识的粒度计算及其应用[J]. 系统工程理论与实践,2002,22(1):48-56.
- [2] 王国胤,张清华,胡军. 粒计算研究综述[J]. 智能系统学报,2007,2(6):8-26.
- [3] ZADEH L A. Fuzzy sets and information granularity[J]. Advances in Fuzzy Set Theory and Applications,1979,11:3-18.
- [4] ZADEH L A. Fuzzy sets[J]. Information and Control,1965,8:338-353.
- [5] ZADEH L A. Outline of a new approach to the analysis of complex systems and decision processes[J]. IEEE Transactions on Systems, Man, and Cybernetics,1973(1):28-44.
- [6] ZADEH L A. Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems[J]. Soft Computing,1998(2):23-25.
- [7] PAWLAK Z. Rough sets[J]. International Journal of Computer & Information Sciences,1982,11:341-356.
- [8] QIAN Y H, LIANG J Y, YAO Y Y, et al. MGRS: a multi-granulation set[J]. Information Sciences,2010,180:949-971.
- [9] QIAN Y H, LIANG X Y, WANG Q, et al. Local rough set: a solution to rough data analysis in big data[J]. International Journal of Approximate Reasoning,2018,97:38-63.
- [10] KONG Q Z, ZHANG X W, XU W H, et al. Attribute reducts of multi-granulation information system[J]. Artificial Intelligence Review,2020,53:1353-1371.
- [11] WILLE R. Restructuring lattice theory: an approach based on hierarchies of concepts[M]. Reidel: Dordrecht-Boston,1982.
- [12] 徐伟华,李金海,魏玲,等. 形式概念分析理论与应用[M]. 北京:科学出版社,2016.
- [13] LI J H, HUANG C C, QI J J, et al. Three-way cognitive concept learning via multi-granulation[J]. Information Sciences,2017,378:244-263.
- [14] LI J H, MEI C L, XU W H, et al. Concept learning via granular computing: a cognitive viewpoint[J]. Information Sciences,2015,298:447-467.
- [15] 张钊,张铃. 问题求解理论及应用[M]. 北京:清华大学出版社,1990.
- [16] YAO Y Y. Three-way decisions with probabilistic rough sets[J]. Information Sciences,2010,180:341-353.
- [17] YAO Y Y. Three-way decisions and cognitive computing[J]. Cognitive Computation,2016,8:543-554.
- [18] YAO Y Y. Three-way decision and granular computing[J]. International Journal of Approximate Reasoning,2018,103:107-123.
- [19] 张文修,吴伟志,梁吉业. 粗糙集理论与方法[M]. 北京:科学出版社,2001.
- [20] 张文修,吴伟志. 粗糙集理论介绍和研究综述[J]. 模糊系统与数学,2000,14(4):1-12.
- [21] 王国胤,姚一豫,于洪. 粗糙集理论与应用研究综述[J]. 计算机学报,2009,32(7):1229-1246.
- [22] KONG Q Z, CHANG X E. Rough set model based on variable universe[J]. CAAI Transactions on Intelligence Technology,2021(3):503-511.
- [23] PAWLAK Z. Information systems theoretical foundations[J]. Information Systems,1981,6(3):205-218.
- [24] PAWLAK Z. Rough sets: theoretical aspects of reasoning about data[M]. Berlin: Springer Science & Business Media,1991.
- [25] MAREK W, PAWLAK Z. Information storage and retrieval systems: mathematical foundations[J]. Theoretical Computer Science,1976,1(4):331-354.
- [26] PAWLAK Z. Mathematical foundations of information retrieval[C]//Proceedings of the International Symposium and Summer School on Mathematical Foundations of Computer Science. Strbské Pleso, High Tatras, Czechoslovakia: DBLP,1973:135-136.
- [27] YAO Y Y. A note on definability and approximations: transactions on rough sets VII[M]. Berlin, Heidelberg: Springer,2007:274-282.
- [28] YAO Y Y. The two sides of the theory of rough sets[J]. Knowledge-Based Systems,2015,80:67-77.
- [29] TRYBULEC W A. Basis of vector space[J]. Formalized Mathematics,1990,1:883-885.
- [30] 范世栋,支天云. 粗糙集的代数性质[J]. 山西大学学报(自然科学版),2001,24(2):116-119.

(责任编辑 马建华 英文审校 黄振坤)