

基于邻域多粒度粗糙集的知识发现模型

程 映<sup>1,2</sup> 刘 勇<sup>3</sup>

(四川大学计算机学院 成都 610000)<sup>1</sup> (四川建筑职业技术学院信息工程系 成都 610000)<sup>2</sup>  
(四川建筑职业技术学院电气工程系 四川 德阳 618000)<sup>3</sup>

**摘 要** 针对现有邻域多粒度粗糙集的定义及相应知识发现算法的不足,重新建立基于邻域多粒度粗糙集的知识发现模型。首先构建了多邻域半径下的乐观邻域多粒度粗糙集模型和悲观邻域多粒度粗糙集模型,讨论了相关性质;然后定义了邻域多粒度粗糙集的粒度重要性,并构造了粒度约简算法;最后通过实例解释了算法的运行机制,验证了算法的有效性。

**关键词** 多粒度,粗糙集,邻域

**中图法分类号** TP182 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.06.034

Knowledge Discovery Model Based on Neighborhood Multi-granularity Rough Sets

CHENG Yi<sup>1,2</sup> LIU Yong<sup>3</sup>

(College of Computer Science,Sichuan University,Chengdu 610000,China)<sup>1</sup>  
(Department of Information and Engineering,Sichuan College of Architectural Technology,Chengdu 610000,China)<sup>2</sup>  
(Department of Electrical Engineering,Sichuan College of Architectural Technology,Deyang,Sichuan 618000,China)<sup>3</sup>

**Abstract** It is the purpose of the present work to re-establish a knowledge discovery model based on neighborhood multi-granulation rough sets from the perspective of the deficiency with respect to the existing definition of neighborhood multi-granulation rough sets and the corresponding knowledge discovery algorithms. We firstly constructed the optimistic neighborhood multi-granulation rough set model and pessimistic neighborhood multi-granulation rough set model under multiple neighborhood radii,and discussed several pertinent properties. Then we gave a definition for the granularity importance of neighborhood multi-granulation rough sets,and constructed a granularity reduction algorithm. Finally we conducted a demonstration for the acting mechanism of the proposed algorithm by using an example,and verified its validity.

**Keywords** Multi-granulation,Rough sets,Neighborhood

1 引言

Pawlak 于 1982 年提出的粗糙集理论从新的视角对知识进行定义,把知识看作是不可分辨关系对论域的分类能力,并通过上、下近似集来刻画给定的概念。该理论的优势在于其所有参数均从所给数据集中获取,因而对问题的描述和处理比较客观,目前已被广泛应用于数据挖掘、决策分析、机器学习和知识发现等领域。过去几十年里,为解决实际应用中的问题,粗糙集已被推广到多种拓展模型。粗糙集理论只能处理符号型数据,无法直接处理数值型数据或符号型与数值型并存的混合型数据。为解决这个问题,Lin<sup>[1]</sup>于 1990 年利用邻域关系代替等价关系,提出了邻域粗糙集模型。

从粒度计算的观点来看,论域上的一个等价关系可视为一个粒度,论域上的一个划分可视为一个粒度空间。Pawlak 粗糙集是在单一粒度下通过上、下近似集来刻画一个概念。

当有两个属性子集  $P$  和  $Q$  时,由  $P$  和  $Q$  分别得到论域的两个粒度(即划分),通过  $U/(P \cup Q)$ (商集  $P \cup Q = \{P_i \cap Q_j : P_i \in U/P, Q_j \in U/Q, P_i \cap Q_j \neq \emptyset\}$ )可合并两个已知粒度,得到一个较细的粒度,目标概念在较细的粒度空间中描述。但细化的过程可能会破坏原始的粒度结构。另外,对于某些对象,属性集  $P$  和  $Q$  中的属性值可能存在冲突,此时无法应用  $P_i \cap Q_j$  来生成相应的粒度空间。对于同一信息系统,决策者之间可能是相互独立的,此时,两个商集的交可能导致决策过程中出现冗余。基于上述原因,根据用户需求,目标有时需要通过多个二元关系描述。钱宇华等<sup>[2-3]</sup>将 Pawlak 粗糙集模型推广到多粒度粗糙集模型,通过论域上的多个等价/二元关系定义了集合的近似。对应到信息系统,即采用一个属性集序列而非一个属性集来对论域进行分类,构造多粒度的论域空间,在此空间上进行目标概念的近似逼近,并分别基于“求同存异”和“求同排异”的近似策略提出乐观多粒度粗糙集和悲

到稿日期:2018-05-14 返修日期:2018-08-22 本文受国家自然科学基金(61071162)资助。

程 映(1977—),女,博士,副教授,主要研究方向为数据挖掘、粒计算、粗糙集理论及应用,E-mail:chengyimail@sohu.com(通信作者);刘 勇(1976—),男,硕士,主要研究方向为人工智能、数据挖掘。

观多粒度粗糙集两种模型<sup>[3]</sup>。此后,多种扩展的多粒度粗糙集模型和算法被提出<sup>[4-15]</sup>。本文将重点讨论邻域多粒度粗糙集及其知识发现模型。

文献[6]将邻域粗糙集模型扩展到多粒度空间,其目的是使模型能够处理同时具有符号型属性和数值型属性的多粒度信息。文献[6]以符号型和数值型属性的不同组合方式构建属性集序列,并以此作为粒化准则,提出了 1 型和 2 型邻域多粒度粗糙集。该模型的粒化沿袭了文献[3]中的方法,仍采用属性集序列构建论域的层次划分,只是在构建属性集序列时,考虑了属性集中符号型和数值型属性的不同组合方式<sup>[16]</sup>。但该模型对不同的属性集均采用相同的邻域半径。对于多源数据,数据的采集标准、描述方式和度量单位都可能存在差异,此时用相同的邻域半径来计算邻域存在不妥。另外,邻域半径的选取和属性集规模的大小对分类也有影响。文献[16]考虑了属性集和邻域半径对分类的共同作用,提出了一种基于双重粒化准则的邻域多粒度粗糙集模型,并给出了相应的属性约简算法。该算法的主要思想是针对  $n$  个邻域半径中的每个邻域半径  $\delta_i$ ,分别以属性重要度为启发因子进行前向搜索,计算出每个  $\delta_i$  上的约简集。对于文献[16]中的结论,笔者有以下 3 点疑义。1) 邻域多粒度粗糙集定义中,  $[x]_{B_i, \delta_j}$  表示对象  $x$  在属性集  $B_i$  和邻域半径  $\delta_j$  下的邻域粒,其中  $B = \{B_1, B_2, \dots, B_m\}$  是  $C$  的  $m$  个属性子集,  $\delta = \{\delta_1, \delta_2, \dots, \delta_n\}$  是  $n$  个邻域半径。 $\delta_j$  和  $B_i$  之间是否存在一一对应关系? 当  $i \neq j$  时,  $[x]_{B_i, \delta_j}$  难以从语义上进行解释。文献[16]没有对此进行进一步解释说明,这导致下近似集的定义也是语义不明的。2) 属性重要度是通过属性依赖度来定义的,但是依赖度的定义仍是基于单一的邻域半径,即属性依赖度的计算并没有体现出多个邻域半径对属性的影响。3) 基于属性重要度提出的属性约简算法是选定一个邻域半径,在这个邻域半径下对所有属性进行约简,得到约简集后再选择下一个邻域半径,对所有属性在该邻域半径下约简。因此,在属性约简的过程中,没有体现出不同邻域半径对不同属性集的影响。更重要的是,属性约简实际上仍是在单粒度空间中进行的,这使得在多粒度空间中刻画目标概念失去了原有的作用;而在多粒度空间中进行粒度约简显然比属性约简更有意义。粒度约简的本质就是在给定的多个粒空间上求取一个无冗余信息粒的粒空间集合。此后,文献[17]提出了双重粒化准则下的邻域多粒度粗糙集快速约简算法,该算法仍然是在文献[16]所定义的下、上近似集和依赖度基础上提出的,因此与文献[16]存在同样的问题。目前关于邻域多粒度粗糙集的研究还较少,本文拟重新建立基于多粒度邻域粗糙集的知识发现模型,以实现在多粒度空间中的数据挖掘。

## 2 相关知识

### 2.1 邻域粗糙集模型

Lin 等<sup>[1]</sup>利用邻域关系替代 Pawlak 粗糙集模型中的等价关系,建立了邻域粗糙集模型。

**定义 1<sup>[1]</sup>** 设  $\langle U, \Delta \rangle$  为非空度量空间,其中  $x \in U, \delta \geq 0$ 。称点集  $\delta(x) = \{y \mid \Delta(x, y) \leq \delta, y \in U\}$  为以  $x$  为中心、以  $\delta$  为半径的邻域。其中,  $\Delta$  为距离函数。

常见的距离函数有欧氏距离函数、闵科夫斯基距离函数、曼哈顿距离函数等。本文采用曼哈顿距离函数,其定义如下。

**定义 2** 两点  $x_i = \{x_{i1}, x_{i2}, \dots, x_{iN}\}$  和  $x_j = \{x_{j1}, x_{j2}, \dots, x_{jN}\}$  的距离函数  $\Delta$  定义为:  $\Delta(x_i, x_j) = \sum_{l=1}^N |x_{il} - x_{jl}|$ 。

给定一非空度量空间  $\langle U, \Delta \rangle$ , 如果有  $\delta_1 \leq \delta_2$ , 则有  $\forall x_i \in U, \delta_1(x_i) \subseteq \delta_2(x_i)$ 。邻域粒子族  $\{\delta(x_i) \mid i = 1, 2, \dots, n\}$  构成论域  $U$  的一个覆盖,一般不构成一个划分。

**定义 3<sup>[1]</sup>** 设  $\langle U, \Delta \rangle$  为非空度量空间,  $\forall x \in U$ , 设  $\delta \geq 0$ ,  $B_1 \subseteq C$  和  $B_2 \subseteq C$  分别是符号型属性和数值型属性,则  $x$  的邻域定义为:

$$n_{B_1}(x) = \{x_i \mid \Delta_{B_1}(x, x_i) = 0, x_i \in U\}$$

$$n_{B_2}(x) = \{x_i \mid \Delta_{B_2}(x, x_i) \leq \delta, x_i \in U\}$$

$$n_{B_1 \cup B_2}(x) = \{x_i \mid \Delta_{B_1}(x, x_i) = 0 \wedge \Delta_{B_2}(x, x_i) \leq \delta, x_i \in U\}$$

**定义 4<sup>[1]</sup>** 定义邻域决策信息系统为  $NDIS = (U, A = C \cup D, V, f, N)$ 。其中,  $C$  为条件属性集,  $D$  为决策属性集,  $C \cap D = \emptyset$ ;  $V = \bigcup_{a \in A} V_a$ ,  $V_a$  为属性  $a$  的值域;  $f: U \times A \rightarrow V$  为信息函数,对于  $\forall a \in A, x \in U, f(x, a) \in V_a$ ;  $N$  为由条件属性  $C$  生成的邻域关系。若由  $B \subseteq C$  生成  $U$  上的邻域关系为  $N_B$ , 则对于集合  $X \subseteq U$ ,  $X$  的下、上近似集和边界区域定义为:

$$\underline{N}_B(X) = \{x_i \in U \mid n_B(x_i) \subseteq X\}$$

$$\overline{N}_B(X) = \{x_i \in U \mid n_B(x_i) \cap X \neq \emptyset\}$$

$$BN_B(X) = \overline{N}_B(X) - \underline{N}_B(X)$$

下近似集  $\underline{N}_B(X)$  也称为  $X$  的正域,记为  $POS_B(X)$ 。当  $BN_B(X) = \emptyset$ , 即  $\overline{N}_B(X) = \underline{N}_B(X)$  时,称在邻域决策信息系统中  $X$  是可定义的,否则称  $X$  是粗糙的。

### 2.2 多粒度粗糙集模型

#### 2.2.1 乐观多粒度粗糙集

乐观多粒度粗糙集以“求同存异”的策略通过多个粒空间来近似刻画目标概念。某个对象可能同时属于多个信息粒,只要其中任意一个信息粒包含在目标概念中,就将该对象纳入下近似集;上近似集则定义为目标概念补集的下近似集的补集。

**定义 5<sup>[2-3]</sup>** 设  $S = (U, AT)$  是一个完备信息系统,其中  $R_1, R_2, \dots, R_m \subseteq AT$ 。对于  $\forall X \subseteq U$ ,  $X$  关于  $R_1, R_2, \dots, R_m$  的乐观多粒度下、上近似集和边界集分别为:

$$\underline{\sum_{i=1}^m R_i}^O(X) = \{x \mid [x]_{R_1} \subseteq X \vee [x]_{R_2} \subseteq X \vee \dots [x]_{R_m} \subseteq X, x \in U\}$$

$$\overline{\sum_{i=1}^m R_i}^O(X) = \sim(\sum_{i=1}^m R_i^O(\sim X))$$

$$Bn_{\sum_{i=1}^m R_i}^O(X) = \overline{\sum_{i=1}^m R_i}^O(X) - \underline{\sum_{i=1}^m R_i}^O(X)$$

#### 2.2.2 悲观多粒度粗糙集

悲观多粒度粗糙集以“求同排异”策略通过多个粒空间来近似刻画目标概念。某个对象所在的信息粒只有都包含在目标概念中时才能将其纳入下近似集;上近似集则定义为目标概念补集的下近似集的补集。

**定义 6<sup>[2-3]</sup>** 设  $S = (U, AT)$  是一个完备信息系统,其中  $R_1, R_2, \dots, R_m \subseteq AT$ 。对于  $\forall X \subseteq U$ ,  $X$  关于  $R_1, R_2, \dots, R_m$  的

悲观多粒度下、上近似集和边界集分别为：

$$\begin{aligned} \sum_{i=1}^m R_i^P(X) &= \{x \mid [x]_{R_1} \subseteq X \wedge [x]_{R_2} \subseteq X \wedge \cdots [x]_{R_m} \subseteq X, \\ &\quad x \in U\} \\ \overline{\sum_{i=1}^m R_i^P(X)} &= \sim(\sum_{i=1}^m R_i^P(\sim X)) \\ Bn_{\sum_{i=1}^m R_i^P}^P(X) &= \overline{\sum_{i=1}^m R_i^P(X)} - \sum_{i=1}^m R_i^P(X) \end{aligned}$$

2.3 邻域多粒度粗糙集

2.3.1 1 型邻域多粒度粗糙集

为了通过多个粒度空间处理符号型属性和数值型属性并存的混合数据,文献[6]提出了 1 型和 2 型邻域多粒度粗糙集。

定义 7<sup>[6]</sup> 设邻域信息系统  $NIS=(U,AT,N)$ ,其中, $U$  是非空论域, $A \subseteq AT$  是符号属性集, $B \subseteq AT$  是数值属性集, $A \cup B \subseteq AT$  是混合属性集。 $U/A$  表示论域  $U$  的划分, $U/B$  和  $U/(A \cup B)$  表示论域的两个覆盖。对于  $\forall X \subseteq U$ , $X$  的乐观多粒度下、上近似集和边界集定义为：

$$\begin{aligned} (A+B)^O X &= \{x \in U \mid n_A(x) \subseteq X \vee n_B(x) \subseteq X\} \\ \overline{(A+B)^O X} &= \sim(A+B)^O(\sim X) \\ Bn_{(A+B)}^O(X) &= \overline{(A+B)^O X} - (A+B)^O X \\ \text{称}((A+B)^O X, \overline{(A+B)^O X}) &\text{为乐观 1 型邻域多粒度粗糙集,其中 } n_A(x) \text{ 表示对象 } x \text{ 在属性集 } A \text{ 下的邻域。} \end{aligned}$$

定义 8<sup>[6]</sup> 设邻域信息系统  $NIS=(U,AT,N)$ ,其中, $U$  是非空论域, $A \subseteq AT$  是符号属性集, $B \subseteq AT$  是数值属性集, $A \cup B \subseteq AT$  是混合属性集。 $U/A$  表示论域  $U$  的划分, $U/B$  和  $U/(A \cup B)$  表示论域的两个覆盖。对于  $\forall X \subseteq U$ , $X$  的悲观多粒度下、上近似集和边界集定义为：

$$\begin{aligned} (A+B)^P X &= \{x \in U \mid n_A(x) \subseteq X \wedge n_B(x) \subseteq X\} \\ \overline{(A+B)^P X} &= \sim(A+B)^P(\sim X) \\ Bn_{(A+B)}^P(X) &= \overline{(A+B)^P X} - (A+B)^P X \\ \text{称}((A+B)^P X, \overline{(A+B)^P X}) &\text{为悲观 1 型邻域多粒度粗糙集。} \end{aligned}$$

2.3.2 2 型邻域多粒度粗糙集

定义 9<sup>[6]</sup> 设邻域信息系统  $NIS=(U,AT,N)$ , $N_1$  和  $N_2$  分别是由论域  $U$  上的属性集  $A_1, B_1$  和  $A_2, B_2$  所诱导的两个邻域关系,其中  $A_1$  和  $A_2$  是两个符号属性集, $B_1$  和  $B_2$  是两个数值属性集,则对于  $\forall X \subseteq U$ , $X$  的乐观多粒度下、上近似集和边界集定义为：

$$\begin{aligned} (N_1+N_2)^O X &= \{x \in U \mid n_{(A_1+B_1)}(x) \subseteq X \vee n_{(A_2+B_2)}(x) \subseteq X\} \\ \overline{(N_1+N_2)^O X} &= \sim(N_1+N_2)^O(\sim X) \\ Bn_{(N_1+N_2)}^O(X) &= \overline{(N_1+N_2)^O X} - (N_1+N_2)^O X \\ \text{称}((N_1+N_2)^O X, \overline{(N_1+N_2)^O X}) &\text{为基于两个邻域的乐观 2 型邻域多粒度粗糙集。其中, } n_{(A_i+B_i)}(x) = \{x \in U \mid n_{A_i}(x) \leq \delta \vee n_{B_i}(x) \leq \delta\}, i=1,2。 \end{aligned}$$

定义 10<sup>[6]</sup> 设邻域信息系统  $NIS=(U,AT,N)$ , $N_1$  和  $N_2$  分别是由论域  $U$  上的属性集  $A_1, B_1$  和  $A_2, B_2$  所诱导的两个邻域关系,其中  $A_1$  和  $A_2$  是两个符号属性集, $B_1$  和  $B_2$  是两个数值属性集,则对于  $\forall X \subseteq U$ , $X$  的悲观多粒度下、上近似

集和边界集定义为：

$$\begin{aligned} (N_1+N_2)^P X &= \{x \in U \mid n_{(A_1+B_1)}(x) \subseteq X \wedge n_{(A_2+B_2)}(x) \subseteq X\} \\ \overline{(N_1+N_2)^P X} &= \sim(N_1+N_2)^P(\sim X) \\ Bn_{(N_1+N_2)}^P(X) &= \overline{(N_1+N_2)^P X} - (N_1+N_2)^P X \\ \text{称}((N_1+N_2)^O X, \overline{(N_1+N_2)^O X}) &\text{为基于两个邻域的悲观 2 型邻域多粒度粗糙集。} \end{aligned}$$

2.4 基于双重粒化准则的邻域多粒度粗糙集

虽然定义 9 和定义 10 中的两个邻域关系  $N_1$  和  $N_2$  是由不同的属性集诱导出,但文献[6]采用了相同的邻域半径。对此,文献[16]提出了基于不同属性集序列和不同邻域半径的邻域多粒度粗糙集。

定义 11<sup>[16]</sup> 设  $NDIS=(U,A=C \cup D,V,f,N)$ , $B=\{B_1,B_2,\cdots,B_m\}$  是  $C$  的  $m$  个属性子集, $\delta=\{\delta_1,\delta_2,\cdots,\delta_n\}$  是  $n$  个邻域半径。定义  $X \subseteq U$  关于  $B$  和  $\delta$  的基于双重粒化准则的乐观邻域多粒度粗糙集下近似集和上近似集分别为：

$$\begin{aligned} \sum_i^j B_i \delta_j^O(X) &= \{x \in U \mid [x]_{B_1 \delta_1} \subseteq X \vee \cdots \vee [x]_{B_j \delta_j} \subseteq X \vee \cdots \\ &\quad \vee [x]_{B_n \delta_n} \subseteq X\} \\ \overline{\sum_i^j B_i \delta_j^O(X)} &= \sim \sum_i^j B_i \delta_j^O(\sim X) \end{aligned}$$

其中, $[x]_{B_i \delta_j}$  表示对象  $x$  在属性集  $B_i$  和邻域半径  $\delta_j$  下的邻域粒( $1 \leq i \leq m, 1 \leq j \leq n$ )。

定义 12<sup>[16]</sup> 设  $NDIS=(U,A=C \cup D,V,f,N)$ , $B=\{B_1,B_2,\cdots,B_m\}$  是  $C$  的  $m$  个属性子集, $\delta=\{\delta_1,\delta_2,\cdots,\delta_n\}$  是  $n$  个邻域半径。定义  $X \subseteq U$  关于  $B$  和  $\delta$  的基于双重粒化准则的悲观邻域多粒度粗糙集下近似集和上近似集分别为：

$$\begin{aligned} \sum_i^j B_i \delta_j^P(X) &= \{x \in U \mid [x]_{B_1 \delta_1} \subseteq X \wedge \cdots \wedge [x]_{B_j \delta_j} \subseteq X \wedge \cdots \\ &\quad \wedge [x]_{B_n \delta_n} \subseteq X\} \\ \overline{\sum_i^j B_i \delta_j^P(X)} &= \sim \sum_i^j B_i \delta_j^P(\sim X) \end{aligned}$$

定义 13<sup>[16]</sup> 设  $NDIS=(U,A=C \cup D,V,f,N)$ , $B \subseteq C$  是  $C$  的属性子集, $\delta$  是 1 个邻域半径, $Y=\{Y_1,Y_2,\cdots,Y_r\}$  是由决策属性集  $D$  在论域  $U$  上导出的划分,定义决策类  $Y$  在邻域半径  $\delta$  下对属性子集  $B$  的属性依赖度为：

$$\gamma_{B\delta}(Y) = \frac{|\sum_{k=1}^r B\delta(Y_k)|}{|U|}$$

定义 14<sup>[16]</sup> 设  $NDIS=(U,A=C \cup D,V,f,N)$ , $R \subseteq B \subseteq C$ , $\delta$  是 1 个邻域半径,若

- 1)  $\forall a \in R, \gamma_{(R-\{a\})\delta}(Y) < \gamma_{R\delta}(Y)$ ;
- 2)  $\gamma_{R\delta}(Y) = \gamma_{B\delta}(Y)$ 。

则称  $R$  是在  $\delta$  下  $B$  的一个约简。

定义 15<sup>[16]</sup> 设  $NDIS=(U,A=C \cup D,V,f,N)$ , $B \subseteq C$  是  $C$  的属性子集, $\delta$  是 1 个邻域半径, $Y=\{Y_1,Y_2,\cdots,Y_r\}$  是由决策属性集  $D$  在论域  $U$  上导出的划分,  $\forall a \in C-B$ ,定义属性  $a$  相对于属性集  $B$  的重要度为：

$$SIG(a,B,Y) = \gamma_{(B \cup \{a\})\delta} - \gamma_{B\delta}$$

下面给出文献[16]中的约简算法,其思想为:若  $\delta=\{\delta_1,\delta_2,\cdots,\delta_n\}$ ,针对  $n$  个邻域半径中的每个邻域半径  $\delta_i$ ,分别进

行以属性重要度为启发因子的前向搜索,计算出每个  $\delta_i$  上的属性约简集  $red[i]$ 。

**算法 1** 基于双重粒化准则的邻域多粒度粗糙集模型常规约简算法<sup>[16]</sup>

Input: NDIS=(U, A=C∪D, V, f, N),  $\delta=\{\delta_1, \delta_2, \dots, \delta_n\}$

Output: 约简  $red[i], 1 \leq i \leq n$

Step1 计算  $Y=\{Y_1, Y_2, \dots, Y_r\}$ , 对  $\delta=\{\delta_1, \delta_2, \dots, \delta_n\}$  排序, 得到升序序列  $\delta'=\{\delta_1', \delta_2', \dots, \delta_n'\}, 1 \rightarrow i$ ;

Step2  $\emptyset \rightarrow red[i]$ ;

Step3 对于任意  $a_j \in C - red[i]$ , 计算  $SIG(a_j, red[i], Y) = Y_{(red[i] \cup \{a_j\})\delta_i} - Y_{red[i]\delta_i}$ ;

Step4 选择满足  $SIG(a_k, red[i], Y) = \max(SIG(a_j, red[i], Y))$  的  $a_k$ ;

Step5 If  $SIG(a_k, red[i], Y) > 0$

$red[i] \cup \{a_k\} \rightarrow red[i]$

go to Step3

else if  $i=n$

break

else

$i++$

go to Step2

### 3 多邻域半径下的邻域多粒度粗糙集

与定义 9、定义 10 相比,定义 11 和定义 12 的优点在于考虑了不同的邻域半径。但是,正如本文引言所述,文献[16]的下、上近似集定义以及相应的属性约简算法都存在一些问题。第 2.4 节的算法表明:该算法是针对  $n$  个邻域半径中的每个邻域半径  $\delta_i$ , 分别进行以属性重要度为启发因子的前向搜索,计算出每个  $\delta_i$  上的属性约简集。因此在属性约简的过程中,该算法并没有体现出不同邻域半径对不同属性集的影响。更重要的是,属性约简实际上仍是在单粒度空间中进行的,这使得在多粒度空间中刻画目标概念失去了原有的意义。本文重新定义了多个不同邻域半径下的邻域多粒度粗糙集,并建立了相应的知识发现模型。

#### 3.1 邻域多粒度粗糙集的定义

##### 3.1.1 乐观邻域多粒度粗糙集

首先,重新定义邻域多粒度粗糙集。

**定义 16** 设  $NS=(U, AT, N)$  为一个邻域信息系统,

$X \subseteq U, \overset{\wedge}{R}_1, \overset{\wedge}{R}_2, \dots, \overset{\wedge}{R}_m$  是由  $m$  个属性子集  $R_1, R_2, \dots, R_m$  诱导的邻域关系,  $\delta=\{\delta_1, \delta_2, \dots, \delta_m\}$  是与  $R_1, R_2, \dots, R_m$  相对应的邻域半径。乐观邻域多粒度粗糙集下近似集和上近似集分别定义为:

$$\underline{X}_{\sum_{i=1}^m \overset{\wedge}{R}_i}^O = \{x \in U \mid \overset{\wedge}{R}_1(x) \subseteq X \vee \overset{\wedge}{R}_2(x) \subseteq X \vee \dots \vee \overset{\wedge}{R}_m(x) \subseteq X\}$$

$$\overline{X}_{\sum_{i=1}^m \overset{\wedge}{R}_i}^O = \sim(\sim X)_{\sum_{i=1}^m \overset{\wedge}{R}_i}^O$$

其中,  $\overset{\wedge}{R}_i(x)$  表示对象  $x$  在邻域半径  $\delta_i$  下由属性集  $R_i$  诱导的邻域粒 ( $1 \leq i \leq m$ )。

边界集定义如下:

$$Bn_{\sum_{i=1}^m \overset{\wedge}{R}_i}^O(X) = \overline{X}_{\sum_{i=1}^m \overset{\wedge}{R}_i}^O - \underline{X}_{\sum_{i=1}^m \overset{\wedge}{R}_i}^O$$

为了应用的方便,算法 2 给出计算邻域多粒度粗糙集下近似集的过程。设  $NS=(U, AT, N)$  为一个邻域信息系统,  $X \subseteq U, \overset{\wedge}{R}_1, \overset{\wedge}{R}_2, \dots, \overset{\wedge}{R}_m$  是由  $m$  个属性子集  $R_1, R_2, \dots, R_m$  诱导的邻域关系,  $\delta=\{\delta_1, \delta_2, \dots, \delta_m\}$  是与  $R_1, R_2, \dots, R_m$  相对应的邻域半径。

**算法 2** 计算邻域多粒度粗糙集下近似集的算法

输入:  $NS=(U, AT, N), X \subseteq U, m$  个属性子集  $R_1, R_2, \dots, R_m$

输出: 下近似集  $\underline{X}_{\sum_{i=1}^m \overset{\wedge}{R}_i}^O$

Step1 根据  $\delta_1, \delta_2, \dots, \delta_m$  分别计算  $\overset{\wedge}{R}_1, \overset{\wedge}{R}_2, \dots, \overset{\wedge}{R}_m$ ;

Step2 令  $i=1, j=1, L=\emptyset$ ;

Step3 For  $i=1$  to  $m$  Do

For  $j=1$  to  $|\overset{\wedge}{R}_i|$  Do

If  $Y_i^j \cap X = Y_i^j, (Y_i^j \in \overset{\wedge}{R}_i)$ , then

$L = L \cup \{Y_i^j\}$

Endif

Endfor

Endfor

Step4 输出下近似集  $L$ 。

**定理 1** 设  $NS=(U, AT, N)$  为一个邻域信息系统,  $X \subseteq$

$U, \overset{\wedge}{R}_1, \overset{\wedge}{R}_2, \dots, \overset{\wedge}{R}_m$  是由  $m$  个属性子集  $R_1, R_2, \dots, R_m$  诱导的邻域关系,  $\delta=\{\delta_1, \delta_2, \dots, \delta_m\}$  是与  $R_1, R_2, \dots, R_m$  相对应的邻域半径。下列性质成立:

$$1) \underline{X}_{\sum_{i=1}^m \overset{\wedge}{R}_i}^O = \bigcup_{i=1}^m \underline{X}_{\overset{\wedge}{R}_i}^O;$$

$$2) \overline{X}_{\sum_{i=1}^m \overset{\wedge}{R}_i}^O = \bigcap_{i=1}^m \overline{X}_{\overset{\wedge}{R}_i}^O;$$

$$3) \underline{X}_{\sum_{i=1}^m \overset{\wedge}{R}_i}^O \supseteq \underline{X}_{\overset{\wedge}{R}_i}^O;$$

$$4) \overline{X}_{\sum_{i=1}^m \overset{\wedge}{R}_i}^O \subseteq \overline{X}_{\overset{\wedge}{R}_i}^O;$$

$$5) \forall X_1 \subseteq X_2 \subseteq U \Rightarrow \underline{X}_{\sum_{i=1}^m \overset{\wedge}{R}_i}^O \subseteq \underline{X}_{\sum_{i=1}^m \overset{\wedge}{R}_i}^O;$$

$$6) \forall X_1 \subseteq X_2 \subseteq U \Rightarrow \overline{X}_{\sum_{i=1}^m \overset{\wedge}{R}_i}^O \subseteq \overline{X}_{\sum_{i=1}^m \overset{\wedge}{R}_i}^O。$$

证明:对于性质 1), 若  $x \in \underline{X}_{\sum_{i=1}^m \overset{\wedge}{R}_i}^O$ , 则  $x \in \{\overset{\wedge}{R}_1(x) \subseteq X \vee \overset{\wedge}{R}_2(x) \subseteq X \vee \dots \vee \overset{\wedge}{R}_m(x) \subseteq X\}$ , 即  $x \in \underline{X}_{\overset{\wedge}{R}_1}^O \cup \underline{X}_{\overset{\wedge}{R}_2}^O \cup \dots \cup \underline{X}_{\overset{\wedge}{R}_m}^O$ ;

对于性质 2),  $\overline{X}_{\sum_{i=1}^m \overset{\wedge}{R}_i}^O = \sim(\sim X)_{\sum_{i=1}^m \overset{\wedge}{R}_i}^O = \sim \bigcup_{i=1}^m (\sim X)_{\overset{\wedge}{R}_i}^O = \sim \bigcap_{i=1}^m (\sim \overline{X}_{\overset{\wedge}{R}_i}^O) = \bigcap_{i=1}^m \overline{X}_{\overset{\wedge}{R}_i}^O$ ;

性质 3)、性质 4) 可由性质 1)、性质 2) 直接得到; 性质 5)、性质 6) 的证明略。

##### 3.1.2 悲观邻域多粒度粗糙集

**定义 17** 设  $NS=(U, AT, N)$  为一个邻域信息系统,

$X \subseteq U, \overset{\wedge}{R}_1, \overset{\wedge}{R}_2, \dots, \overset{\wedge}{R}_m$  是由  $m$  个属性子集  $R_1, R_2, \dots, R_m$  诱导的邻域关系,  $\delta=\{\delta_1, \delta_2, \dots, \delta_m\}$  是与  $R_1, R_2, \dots, R_m$  相对应的邻域半径。悲观邻域多粒度粗糙集下近似集和上近似集分别定义为:

$$\underline{X}_{\sum_{i=1}^m \overset{\wedge}{R}_i}^P = \{x \in U \mid \overset{\wedge}{R}_1(x) \subseteq X \wedge \overset{\wedge}{R}_2(x) \subseteq X \wedge \dots \wedge \overset{\wedge}{R}_m(x) \subseteq X\}$$



$$\overline{X}_{\sum_{i=1}^m \wedge R_i}^P = \sim(\sim X)_{\sum_{i=1}^m \wedge R_i}^P$$

其中,  $R_i(x)$  表示对象  $x$  在邻域半径  $\delta_i$  下由属性集  $R_i$  诱导的邻域粒 ( $1 \leq i \leq m$ )。

边界集定义如下:

$$Bn_{\sum_{i=1}^m \wedge R_i}^P(X) = \overline{X}_{\sum_{i=1}^m \wedge R_i}^P - X_{\sum_{i=1}^m \wedge R_i}^P$$

与算法 2 类似, 可以计算悲观邻域多粒度粗糙集的下近似集。

**定理 2** 设  $NS=(U, AT, N)$  为一个邻域信息系统,  $X \subseteq U$ ,  $\hat{R}_1, \hat{R}_2, \dots, \hat{R}_m$  是由  $m$  个属性子集  $R_1, R_2, \dots, R_m$  诱导的邻域关系,  $\delta = \{\delta_1, \delta_2, \dots, \delta_m\}$  是与  $R_1, R_2, \dots, R_m$  相对应的邻域半径。下列性质成立:

- 1)  $X_{\sum_{i=1}^m \wedge R_i}^P = \bigcap_{i=1}^m X_{\wedge R_i}^P$ ;
- 2)  $\overline{X}_{\sum_{i=1}^m \wedge R_i}^P = \bigcup_{i=1}^m \overline{X}_{\wedge R_i}^P$ ;
- 3)  $X_{\sum_{i=1}^m \wedge R_i}^P \subseteq X_{\wedge R_i}^P$ ;
- 4)  $\overline{X}_{\sum_{i=1}^m \wedge R_i}^P \supseteq \overline{X}_{\wedge R_i}^P$ ;
- 5)  $\forall X_1 \subseteq X_2 \subseteq U \Rightarrow X_{\sum_{i=1}^m \wedge R_i}^P \subseteq X_{\sum_{i=1}^m \wedge R_i}^P$ ;
- 6)  $\forall X_1 \subseteq X_2 \subseteq U \Rightarrow \overline{X}_{\sum_{i=1}^m \wedge R_i}^P \supseteq \overline{X}_{\sum_{i=1}^m \wedge R_i}^P$ 。

证明: 对于性质 1), 若  $x \in X_{\sum_{i=1}^m \wedge R_i}^P$ , 则  $x \in \{R_1(x) \subseteq X \wedge R_2(x) \subseteq X \wedge \dots \wedge R_m(x) \subseteq X\}$ , 即  $x \in X_{\wedge R_1}^P \cap X_{\wedge R_2}^P \cap \dots \cap X_{\wedge R_m}^P$ ;

对于性质 2),  $\overline{X}_{\sum_{i=1}^m \wedge R_i}^P = \sim(\sim X)_{\sum_{i=1}^m \wedge R_i}^P = \sim \bigcap_{i=1}^m (\sim X)_{\wedge R_i}^P = \sim \bigcap_{i=1}^m (\sim \overline{X}_{\wedge R_i}^P) = \bigcup_{i=1}^m \overline{X}_{\wedge R_i}^P$ ;

性质 3)、性质 4) 可由性质 1)、性质 2) 直接得到; 性质 5)、性质 6) 的证明略。

**定理 3** 设  $NS=(U, AT, N)$  为一个邻域信息系统,  $X \subseteq U$ ,  $\hat{R}_1, \hat{R}_2, \dots, \hat{R}_m$  是由  $m$  个属性子集  $R_1, R_2, \dots, R_m$  诱导的邻域关系,  $\delta = \{\delta_1, \delta_2, \dots, \delta_m\}$  是与  $R_1, R_2, \dots, R_m$  相对应的邻域半径。记  $R = \{R_1, R_2, \dots, R_m\}$ , 若  $R' \subseteq R$ , 则:

- 1)  $X_{\sum_{R_i \in R} \wedge R_i}^O \supseteq X_{\sum_{R_i \in R'} \wedge R_i}^O$ ;
- 2)  $X_{\sum_{R_i \in R} \wedge R_i}^P \subseteq X_{\sum_{R_i \in R'} \wedge R_i}^P$ 。

证明: 由定理 1,  $X_{\sum_{R_i \in R} \wedge R_i}^O = \bigcup_{R_i \in R} X_{\wedge R_i}^O$ ,  $X_{\sum_{R_i \in R'} \wedge R_i}^O = \bigcup_{R_i \in R'} X_{\wedge R_i}^O$ , 由于  $R' \subseteq R$ , 因此  $\bigcup_{R_i \in R} X_{\wedge R_i}^O \supseteq \bigcup_{R_i \in R'} X_{\wedge R_i}^O$ , 即  $X_{\sum_{R_i \in R} \wedge R_i}^O \supseteq X_{\sum_{R_i \in R'} \wedge R_i}^O$ ; 类似可证 2) 成立。

3.2 基于邻域多粒度粗糙集的粒度约简模型

3.2.1 邻域多粒度粗糙集模型中的几个度量

刻画目标概念时的不确定性源于边界的存在, 边界域越大, 目标概念被表达的精确度越低。为了更准确地体现这一思想, 下面定义了精确度。

**定义 18** 设  $NS=(U, AT, N)$  为一个邻域信息系统,  $\emptyset \neq X \subseteq U$ ,  $\hat{R}_1, \hat{R}_2, \dots, \hat{R}_m$  是由  $m$  个属性子集  $R_1, R_2, \dots, R_m$  诱导的邻域关系,  $\delta = \{\delta_1, \delta_2, \dots, \delta_m\}$  是与  $R_1, R_2, \dots, R_m$  相对应的邻域半径。定义近似的精确度为:

$$\alpha_{\sum_{i=1}^m \wedge R_i}^{\Delta}(X) = \frac{|\overline{X}_{\sum_{i=1}^m \wedge R_i}^{\Delta}|}{|\overline{X}_{\sum_{i=1}^m \wedge R_i}^{\Delta}|}$$

其中,  $\Delta$  表示  $O$  或  $P$ , 即乐观和悲观两种情况,  $|X|$  表示集合  $X$  的基数。

**定义 19** 设  $NS=(U, AT \cup D, N)$  为一个邻域决策系统,  $D$  是决策属性集,  $U/D = \{Y_1, Y_2, \dots, Y_r\}$ ,  $\hat{R}_1, \hat{R}_2, \dots, \hat{R}_m$  是由  $m$  个属性子集  $R_1, R_2, \dots, R_m$  诱导的邻域关系,  $\delta = \{\delta_1, \delta_2, \dots, \delta_m\}$  是与  $R_1, R_2, \dots, R_m$  相对应的邻域半径。  $D$  对属性子集  $R_1, R_2, \dots, R_m$  的依赖度定义为:

$$\gamma_{\sum_{i=1}^m \wedge R_i}^{\Delta}(D) = \frac{|\bigcup_{j=1}^r (Y_j)_{\sum_{i=1}^m \wedge R_i}^{\Delta}|}{|U|}$$

其中,  $\Delta$  表示  $O$  或  $P$ , 即乐观和悲观两种情况。

**推论 1** 若  $R' \subseteq R$ , 则  $\gamma_{\sum_{i=1}^m \wedge R_i'}^O(D) \leq \gamma_{\sum_{i=1}^m \wedge R_i}^O(D)$ ,  $\gamma_{\sum_{i=1}^m \wedge R_i'}^P(D) \geq \gamma_{\sum_{i=1}^m \wedge R_i}^P(D)$ 。

证明: 根据定理 3 即证。

推论 1 说明, 随着粒度空间个数的增多, 依赖度也呈单调变化。

**定义 20** 设  $NS=(U, AT \cup D, N)$  为一个邻域决策系统,  $D$  是决策属性集,  $U/D = \{Y_1, Y_2, \dots, Y_r\}$ , 粒度集 (即属性子集构成的集合)  $R \subseteq AT$ , 若满足:

- 1)  $\gamma_R^{\Delta}(D) = \gamma_{AT}^{\Delta}(D)$ ;
- 2) 对  $\forall A \in R$ ,  $\gamma_{R-(A)}^{\Delta}(D) \neq \gamma_{AT}^{\Delta}(D)$ 。

则称  $R$  为信息系统  $NS$  的粒度约简集。

**定义 21** 设  $NS=(U, AT \cup D, N)$  为一个邻域决策系统,  $D$  是决策属性集, 粒度集  $R \subseteq AT$ ,  $\forall A \in AT - R$ , 粒度  $A$  在粒度集  $R$  中关于  $D$  的重要程度量定义为:

$$Sig_{outer}^{\Delta}(A, R, D) = |\gamma_{\sum_{R_i \in R \cup \{A\}} \wedge R_i}^{\Delta}(D) - \gamma_{\sum_{R_i \in R} \wedge R_i}^{\Delta}(D)|$$

其中,  $\Delta$  表示  $O$  或  $P$ 。显然,  $Sig_{outer}^{\Delta}(A, R, D) \geq 0$ 。粒度重要程度量是粒度本身、粒度相对的粒度集以及决策变量三者构成的一个函数。

3.2.2 基于邻域多粒度粗糙集模型的粒度约简算法

基于粒度重要程度量, 可构造粒度约简算法。该算法以空集为起点, 每次计算全部剩余粒度的重要度, 从中选择粒度重要度值最大的粒度加入约简集合中, 直到所有剩余粒度的重要度为 0, 即加入任何新的粒度, 系统的依赖度不再变化为止。前向搜索算法能够确保重要的粒度首先被加入到约简中, 从而不损失重要的特征。算法 3 采用乐观邻域多粒度粗糙集模型。

设  $NS=(U, AT \cup D, N)$  为一个邻域决策系统,  $D$  是决策属性集,  $\hat{R}_1, \hat{R}_2, \dots, \hat{R}_m$  是由  $m$  个属性子集  $R_1, R_2, \dots, R_m$  诱导的邻域关系,  $\delta = \{\delta_1, \delta_2, \dots, \delta_m\}$  是与  $R_1, R_2, \dots, R_m$  相对应的邻域半径。

**算法 3** 基于邻域多粒度粗糙集模型的粒度约简算法  
输入:  $NS=(U, AT \cup D, N)$ ,  $m$  个属性子集  $R_1, R_2, \dots, R_m$ , 记  $R = \{R_1, R_2, \dots, R_m\}$

输出:粒度约简 Red

Step1 计算  $U/D=\{Y_1,Y_2,\cdots,Y_r\}$ , 根据  $\delta_1,\delta_2,\cdots,\delta_m$  分别计算  $R_1,\hat{R}_2,\cdots,\hat{R}_m$ , 由算法 2 分别计算决策类  $Y_1,Y_2,\cdots,Y_r$  的下近似集合;

Step2 令  $Red=\emptyset$ ;

Step3 对于  $\forall R_i \in R-Red$ , 计算  $Sig^O(R_i,Red,D)=\gamma^O_{\sum_{R_i \in Red \cup \{R_i\}} \hat{R}_i}(D)-\gamma^O_{\sum_{R_i \in Red} \hat{R}_i}(D)$ //定义  $\gamma_{\emptyset}(D)=0$ ;

Step4 选择  $R_k$ , 使其满足:  $Sig^O(R_k,Red,D)=\max_{i=1}^m\{Sig(R_i,Red,D)\}$ ; 若  $|Sig^O(R_k,Red,D)| \neq 1$ , 则选择使  $\gamma_{\hat{R}_i}^{\wedge}(D)$  取最大值的  $R_i$  作为  $R_k$ ;

Step5 若  $Sig^O(R_k,Red,D)>0$ , 令  $Red=Red \cup \{R_k\}$ , 转 Step3; 否则, 转 Step6;

Step6 输出 Red.

4 实例分析

为了证明算法的可行性, 将邻域多粒度粗糙集的粒度约简算法应用于方案评估。某公司有 10 种方案, 分别在 5 个站点上进行评估, 站点 1—站点 5 的评估数据见表 1—表 5。将整个评估表看作一个完备信息系统, 论域  $U=\{x_1,x_2,\cdots,x_{10}\}$  表示 10 种方案; 条件属性集  $C=\{a_1,a_2,a_3,a_4,a_5\}$  表示评价指标, 其中  $a_1,a_2,a_4$  为符号型属性,  $a_3$  和  $a_5$  为数值型属性。各站点自行选择评价指标进行评价, 表 1—表 5 分别对应了 5 个属性子集  $R_1,R_2,\cdots,R_5$ , 决策属性  $D$  表示方案是否通过评估。在评估过程中, 将每个站点的评估看成一个粒度空间, 在多粒度空间中进行粒度约简, 找到最小评估站点集。

表 1 站点 1 评估数据

Table 1 Evaluation data for site 1

| 方案    | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|
| $a_1$ | 1     | 1     | 1     | 2     | 3     | 2     | 3     | 3     | 3     | 3        |
| $a_2$ | 2     | 2     | 2     | 1     | 2     | 1     | 2     | 2     | 2     | 2        |
| $a_3$ | 0.40  | 0.42  | 0.54  | 0.12  | 0.37  | 0.21  | 0.46  | 0.51  | 0.88  | 0.99     |

表 2 站点 2 评估数据

Table 2 Evaluation data for site 2

| 方案    | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|
| $a_1$ | 1     | 1     | 2     | 1     | 2     | 2     | 2     | 2     | 2     | 3        |
| $a_3$ | 0.45  | 0.56  | 0.71  | 0.49  | 0.69  | 0.22  | 0.82  | 0.19  | 0.31  | 0.93     |
| $a_4$ | 1     | 1     | 3     | 1     | 3     | 2     | 3     | 2     | 2     | 2        |

表 3 站点 3 评估数据

Table 3 Evaluation data for site 3

| 方案    | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|
| $a_2$ | 1     | 2     | 2     | 1     | 2     | 1     | 3     | 2     | 3     | 3        |
| $a_3$ | 0.11  | 0.55  | 0.66  | 0.23  | 0.59  | 0.19  | 0.79  | 0.17  | 0.84  | 0.88     |
| $a_5$ | 0.53  | 0.81  | 0.79  | 0.48  | 0.92  | 0.45  | 0.54  | 0.20  | 0.60  | 0.61     |

表 4 站点 4 评估数据

Table 4 Evaluation data for site 4

| 方案    | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|
| $a_1$ | 2     | 2     | 2     | 1     | 2     | 1     | 1     | 1     | 3     | 3        |
| $a_2$ | 2     | 2     | 2     | 1     | 2     | 3     | 3     | 3     | 2     | 2        |
| $a_3$ | 0.43  | 0.55  | 0.39  | 0.79  | 0.51  | 0.44  | 0.47  | 0.54  | 0.85  | 0.90     |
| $a_4$ | 1     | 1     | 1     | 2     | 1     | 1     | 1     | 1     | 2     | 2        |

表 5 站点 5 评估数据

Table 5 Evaluation data for site 5

| 方案    | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|
| $a_1$ | 1     | 2     | 1     | 1     | 3     | 1     | 2     | 3     | 3     | 3        |
| $a_2$ | 2     | 2     | 2     | 2     | 2     | 2     | 2     | 2     | 3     | 3        |
| $a_5$ | 0.51  | 0.75  | 0.42  | 0.47  | 0.53  | 0.56  | 0.87  | 0.49  | 0.13  | 0.21     |

粒度集  $AT=\{R_1,R_2,R_3,R_4,R_5\}$ ,  $\hat{R}_1,\hat{R}_2,\cdots,\hat{R}_5$  是属性子集  $R_1,R_2,\cdots,R_5$  诱导的邻域关系,  $\delta_1=0.125,\delta_2=0.11,\delta_3=0.125,\delta_4=0.10,\delta_5=0.12$  是与  $R_1,R_2,\cdots,R_m$  相对应的邻域半径。  $U/D=\{Y_1,Y_2\}$ , 其中,  $Y_1=\{x_1,x_2,x_4,x_6,x_8\}$ ,  $Y_2=\{x_3,x_5,x_7,x_9,x_{10}\}$ 。采用乐观融合策略, 利用算法 3 进行粒度约简。

令  $Red=\emptyset$ , 定义  $\gamma_{\emptyset}(D)=0$ , 计算粒度集  $R_1,R_2,\cdots,R_5$  关于  $D$  的重要性可得:

$$Sig^O(R_1,Red,D)=\gamma_{\hat{R}_1}^O(D)-0=\frac{2}{5}-0=\frac{2}{5}$$
$$Sig^O(R_2,Red,D)=\gamma_{\hat{R}_2}^O(D)-0=\frac{4}{5}-0=\frac{4}{5}$$
$$Sig^O(R_3,Red,D)=\gamma_{\hat{R}_3}^O(D)-0=\frac{7}{10}-0=\frac{7}{10}$$
$$Sig^O(R_4,Red,D)=\gamma_{\hat{R}_4}^O(D)-0=\frac{3}{10}-0=\frac{3}{10}$$
$$Sig^O(R_5,Red,D)=\gamma_{\hat{R}_5}^O(D)-0=\frac{3}{10}-0=\frac{3}{10}$$

$Sig^O(R_2,Red,D)$ 最大, 选择  $R_2$  加入约简集  $Red$ , 即  $Red=\{R_2\}$ 。重复以上过程, 相对于约简集  $Red$ , 分别计算属性重要性可得:

$$Sig^O(R_1,Red,D)=\gamma^O_{\sum_{R_i \in Red \cup \{R_1\}} \hat{R}_i}(D)-\gamma^O_{\sum_{R_i \in Red} \hat{R}_i}(D)=\gamma^O_{\sum_{R_i \in R_1 \cup R_2} \hat{R}_i}(D)-\gamma_{\hat{R}_2}^O(D)=1-\frac{4}{5}=\frac{1}{5}$$
$$Sig^O(R_3,Red,D)=\gamma^O_{\sum_{R_i \in Red \cup \{R_3\}} \hat{R}_i}(D)-\gamma^O_{\sum_{R_i \in Red} \hat{R}_i}(D)=\gamma^O_{\sum_{R_i \in R_2 \cup R_3} \hat{R}_i}(D)-\gamma_{\hat{R}_2}^O(D)=1-\frac{4}{5}=\frac{1}{5}$$
$$Sig^O(R_4,Red,D)=\gamma^O_{\sum_{R_i \in Red \cup \{R_4\}} \hat{R}_i}(D)-\gamma^O_{\sum_{R_i \in Red} \hat{R}_i}(D)=\gamma^O_{\sum_{R_i \in R_2 \cup R_4} \hat{R}_i}(D)-\gamma_{\hat{R}_2}^O(D)=\frac{9}{10}-\frac{4}{5}=\frac{1}{10}$$
$$Sig^O(R_5,Red,D)=\gamma^O_{\sum_{R_i \in Red \cup \{R_5\}} \hat{R}_i}(D)-\gamma^O_{\sum_{R_i \in Red} \hat{R}_i}(D)=\gamma^O_{\sum_{R_i \in R_2 \cup R_5} \hat{R}_i}(D)-\gamma_{\hat{R}_2}^O(D)=1-\frac{4}{5}=\frac{1}{5}$$

由于  $Sig^O(R_1,Red,D)=Sig^O(R_3,Red,D)=Sig^O(R_5,Red,D)$ , 根据算法 3, 选择依赖度最大的  $R_3$  加入约简集  $Red$ , 即  $Red=\{R_2,R_3\}$ , 再次重复以上过程, 可得  $Sig^O(R_1,Red,D)=Sig^O(R_4,Red,D)=Sig^O(R_5,Red,D)=0$ , 算法终止, 可得  $Red=\{R_2,R_3\}$  为乐观融合策略下的邻域多粒度粗糙集的粒度约简集。

**结束语** 针对现有邻域多粒度粗糙集模型的定义及其应用的不足, 本文重新定义了乐观和悲观邻域多粒度粗糙集, 讨论了相关性质并提出了基于邻域多粒度粗糙集模型的粒度约简算法; 通过实例解释了算法的运行机制和有效性。需要注

意的是,在算法 3 中,邻域半径的选取是事先设定的,实际应用时,可根据问题的需要,由先验知识确定或由领域专家指定。在邻域多粒度粗糙集中进行决策分析将是以后研究工作的重点。

参 考 文 献

[1] LIN T Y,LIU Q,HUANG K J,et al. Rough sets neighborhood systems and approximation[C]// Proc of the 5th International Symposium on Methodologies of Intelligent Systems. Knoxville, 1990:130-141.

[2] QIAN Y H,LIANG J Y. Rough set method based on multi-granulations[C]// Proceedings of 5th IEEE Conference on Cognitive Informatics. New York:IEEE,2006:297-304.

[3] QIAN Y H,LIANG J Y,YAO Y Y,et al. MGRS:A multigranulation rough set[J]. Information Sciences,2010,180(6):949-970.

[4] QIAN Y H,LIANG J Y,DANG C Y. Incomplete multigranulation rough set[J]. IEEE Transation on Systems,Man and Cybernetics,2010,40(2):420-431.

[5] QIAN Y H,ZHANG H,SANG Y L,et al. Multigranulation decision-theoretic rough sets[J]. International Journal of Approximate Reasoning,2014,55(1):225-237.

[6] LIN G P,QIAN Y H,LI J J. NMGRS:Neighborhood-based multigranulation rough set[J]. International Journal of Approximate Reasoning,2012,53(7):1080-1093.

[7] YANG X B,SONG X N,DOU H L,et al. Multi-granulation rough set;from crisp to fuzzy case[J]. Annals of Fuzzy Mathematics and Informatics,2011,1(1):55-70.

[8] XU W H,WANG Q R,ZHANG X T. Multi-granulation rough sets based on tolerance relations[J]. Soft Compute,2013,17(7):1241-1252.

[9] LIU C H,MIAO D Q. On multi-granulation covering rough sets[J]. International Journal of Approximate Reasoning,2014,55(6):1404-1418.

[10] SHE Y H,HE X L,SHI H X,et al. A multiple-valued logic approach for multigranulation rough set model[J]. International Journal of Approximate Reasoning,2017,82:270-284.

[11] YAO Y Y,SHE Y H. Rough set models in multigranulation spaces[J]. Information Sciences,2016,327:40-56.

[12] SUN B Z,MA W M,QIAN Y H. Multigranulation fuzzy rough set over two universes and its application to decision making[J]. Knowledge-Based Systems,2017,123:61-74.

[13] SHE Y H,HE X L,SHI H X,et al. A multiple-valued logic approach for multigranulation rough set model[J]. International Journal of Approximate Reasoning,2017,82:270-284.

[14] QIAN Y H,LIANG X Y,LIN G P,et al. Local multigranulation decision-theoretic rough sets[J]. International Journal of Approximate Reasoning,2017,82:119-137.

[15] FENG T,MI J S. Variable precision multigranulation decision-theoretic fuzzy rough sets[J]. Knowledge-Based Systems,2016,91:93-101.

[16] XU Y,YANG H J,JI X. Neighborhood multi-granulation rough set model based on double granulate criterion[J]. Control and Decision,2015,30(8):1469-1478. (in Chinese)  
徐怡,杨宏健,纪霞. 基于双重粒化准则的邻域多粒度粗糙集模型[J]. 控制与决策,2015,30(8):1469-1478.

[17] MA F M,CHEN J W,ZHANG T F. Quick attribute reduction algorithm for neighbor-hood multi-granulation rough set based on double granulate criterion[J]. Control and Decision,2017,32(6):1121-1127. (in Chinese)  
马福民,陈静雯,张腾飞. 基于双重粒化准则的邻域多粒度粗糙集快速约简算法[J]. 控制与决策,2017,32(6):1121-1127.