

可变粒度粗糙集

张 明^{1,2} 唐振民¹ 徐维艳³ 杨习贝^{1,2}

(南京理工大学计算机科学与技术学院 南京 210094)¹

(江苏科技大学计算机科学与工程学院 镇江 212003)² (江苏科技大学数理学院 镇江 212003)³

摘 要 分析乐观多粒度粗糙集和悲观多粒度粗糙集的不足之后,提出一种可变粒度粗糙集模型,定义了可变粒度粗糙集的下、上近似集,研究了可变粒度粗糙集与这两种多粒度粗糙集的性质,证明了可变粒度粗糙集是多粒度粗糙集的泛化,最后给出几种可变粒度粗糙集的度量因子,研究了变粒度粗糙集与多粒度粗糙集度量之间的关系。

关键词 可变粒度粗糙集,多粒度粗糙集,度量

中图法分类号 TP18 **文献标识码** A

Variable Granulation Rough Set

ZHANG Ming^{1,2} TANG Zhen-min¹ XU Wei-yan³ YANG Xi-bei^{1,2}

(School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China)¹

(School of Computer Science and Engineering, Jiangsu University of Science and Technology, Zhenjiang 212003, China)²

(School of Mathematics and Physics, Jiangsu University of Science and Technology, Zhenjiang 212003, China)³

Abstract By analyzing the weakness of optimistic Muiltl-granulation rough set and pessimistic Muiltl-granulation rough set, the Variable granulation rough set was proposed, the lower and the upper approximation were defined. Furthermore, the properties of these three kinds of rough sets were discussed, and that the Variable granulation Rough Set is the generalization of Muiltl-granulation rough set was proved. At last, some measures of those rough sets were defined, the relation of their measures was discussed.

Keywords Variable granulation rough set, Muiltl-granulation rough set, Measure

1 引言

粗糙集理论^[1]是 Pawlak 教授于 1982 年提出的一种能够有效分析和处理不精确、不一致、不完整信息的数学方法。经过近 30 年的发展,粗糙集理论已在模式识别、机器学习、决策分析、知识获取和数据挖掘等领域得到了广泛应用。经典的粗糙集是通过论域上的等价关系导出的知识,以下、上近似集的形式来粗糙表示未知的知识(概念)。为了处理不同类型的信息系统,许多学者对 Pawlak 粗糙集模型进行了扩充,分别有容差关系、相似关系、限制容差关系、优势关系和模糊关系粗糙集等^[2-4],以上的粗糙集模型都是基于单个二元关系导出的知识粒,来定义粗糙集的下、上近似的。Qian 从粒计算^[5]的角度,分析单个知识粒度下粗糙集的不足,提出了多粒度粗糙集的概念^[6-8],并给出了乐观多粒度粗糙集模型和悲观多粒度粗糙集模型^[6-8],同时证明了 Pawlak 的经典粗糙集模型是多粒度粗糙集的特例。

本文经过分析,认为乐观多粒度粗糙集下近似的定义过于宽松(即在 m 个知识粒度 $[x]_{A_1}, [x]_{A_2}, \dots, [x]_{A_m}$ 下,只要

存在某个知识粒 $[x]_{A_i}, 1 \leq i \leq m$, 使得 $[x]_{A_i} \subseteq X$ 成立,则 x 就属于 X 的下近似),而悲观多粒度粗糙集下近似的定义又过于严格(要求在 m 个知识粒度 $[x]_{A_1}, [x]_{A_2}, \dots, [x]_{A_m}$ 下,对任意的知识粒 $[x]_{A_i} \subseteq X$, 则 x 才属于 X 的下近似)。本文认为在 m 个知识粒度 $[x]_{A_1}, [x]_{A_2}, \dots, [x]_{A_m}$ 下,只要有相对多数的知识粒 $[x]_{A_i} \subseteq X$ 成立时,我们就认为 x 属于 X 的下近似,这也符合我们生活中的少数服从多数的思想。为此提出了一种可变粒度粗糙集模型,它可以克服乐观多粒度粗糙集下近似的定义过于宽松和悲观多粒度粗糙集下近似的定义又过于严格的缺陷,同时还可以证明变粒度粗糙集是乐观和悲观多粒度粗糙集的泛化,并进一步研究了变粒度粗糙集的一些性质,讨论了变粒度粗糙集与多粒度粗糙集之间的关系。

2 基本概念

2.1 pawlak's 粗糙集

信息系统(Information System, IS)可以定义为一个四元组: $IS = \langle U, AT, V, f \rangle$, 其中 U 是一个被称为论域的非空有限对象集合; AT 是非空有限属性集合,对 $\forall a \in AT, V_a$ 表示属性

到稿日期:2010-11-29 返修日期:2011-02-26 本文受国家自然科学基金重点项目(90820306),中国博士后科学基金项目(20100481149)资助。

张 明(1978—),男,博士生,讲师,主要研究方向为粗糙集理论及应用、粒计算等, E-mail: zm_fred@163.com; 唐振民(1961—),男,博士,教授,博士生导师,主要研究方向为模式识别与智能信息系统等; 徐维艳(1978—),女,硕士,讲师,主要研究方向为模糊分析、粗糙集及粒计算; 杨习贝(1980—),男,博士后,主要研究方向为粗糙集理论及应用、粒计算等。

a 的值域; V 表示全体属性的值域集合, 即 $V = V_{AT} = \bigcup_{a \in AT} V_a$; f 为信息函数, 对 $\forall x \in U, a \in AT$, 定义 $f(x, a)$ 表示 x 在属性 a 上的取值, 则有 $f(x, a) \in V_a$ 。如果 $AT = C \cup D$, 其中 C 是条件属性集合, D 是决策属性集合, 则 $\langle U, C \cup D, V, f \rangle$ 又被称为决策信息系统(DIS)。

定义 1 设 $IS = \langle U, AT, V, f \rangle$, $\forall A \subseteq AT$, 定义 A 上的不可分辨关系 $IND(A)$ 。 $IND(A) = \{(x, y) \in U^2 : \forall a \in A, f(x, a) = f(y, a)\}$ 。

不可分辨关系满足自反性、对称性和传递性, 是 U 上的一个等价关系, 由此等价关系可以导出 U 上的一个划分 $U/IND(A)$, 且 $U/IND(A)$ 中的等价类记为 $[x]_A$, $[x]_A = \{y \in U : f(x, y) = IND(A)\}$ 。从粒计算的角度看, $U/IND(A) = \{[x]_A : x \in U\}$ 是一个粒空间, 等价类记为 $[x]_A$, 被称为知识(颗)粒。

定义 2 设 $IS = \langle U, AT, V, f \rangle$, $X \subseteq U, A \subseteq AT$, 定义 X 关于属性 A 的下、上近似和边界域集分别记为:

$$\begin{aligned}\underline{A}(X) &= \{x \in U : [x]_A \subseteq X\} \\ \overline{A}(X) &= \{x \in U : [x]_A \cap X \neq \emptyset\} \\ BND_A(X) &= \overline{A}(X) - \underline{A}(X)\end{aligned}$$

则概念 X 可以通过其下、上近似来粗糙表示, 称序偶 $\langle \underline{A}(X), \overline{A}(X) \rangle$ 为概念 X 关于属性 A 的粗糙集。

2.2 多粒度粗糙集

在经典的粗糙集理论中, 论域 U 上的任一概念 X 是通过单个不可分辨关系导出的等价类(单个知识粒度)来粗糙表示的, 通过概念的下、上近似来逼近 X 。Qian 从粒计算的角度出发, 将单个知识粒度的粗糙集扩展到由一组二元关系导出的多个等价类(多个知识粒度)来近似表示概念 X , 并将其命名为多粒度粗糙集。在文献[6, 7]中分别定义了乐观多粒度粗糙集模型和悲观多粒度粗糙集模型。

2.2.1 乐观多粒度粗糙集^[6]

定义 3 设 $IS = \langle U, AT, V, f \rangle$, 令 $A_1, A_2, \dots, A_m$ 是 AT 的 m 个属性子集, 定义 $X \subseteq U$ 关于属性 $A_1, A_2, \dots, A_m$ 的乐观多粒度下、上近似集分别记为:

$$\begin{aligned}\sum_{i=1}^m \underline{A_i}^O(X) &= \{x \in U : [x]_{A_1} \subseteq X \vee [x]_{A_2} \subseteq X \vee \dots \vee [x]_{A_m} \subseteq X\} \\ \sum_{i=1}^m \overline{A_i}^O(X) &= \sim \sum_{i=1}^m \underline{A_i}^O(\sim X)\end{aligned}$$

式中, $[x]_{A_i} (1 \leq i \leq m)$ 是由属性子集 A_i 导出的等价类, $\sim X$ 是 X 的补集。 $\langle \sum_{i=1}^m \underline{A_i}^O(X), \sum_{i=1}^m \overline{A_i}^O(X) \rangle$ 称为 X 关于属性 $A_1, A_2, \dots, A_m$ 的乐观多粒度粗糙集。

定理 1 设 $IS = \langle U, AT, V, f \rangle$, 令 $A_1, A_2, \dots, A_m$ 是 AT 的 m 个属性子集, $\forall X \subseteq U$, 可得:

$$\sum_{i=1}^m \underline{A_i}^O(X) = \{x \in U : [x]_{A_1} \cap X \neq \emptyset \wedge [x]_{A_2} \cap X \neq \emptyset \wedge \dots \wedge [x]_{A_m} \cap X \neq \emptyset\}$$

证明: 由定义 3 可知:

$$\begin{aligned}x \in \sum_{i=1}^m \underline{A_i}^O(X) &\Leftrightarrow x \notin \sum_{i=1}^m \underline{A_i}^O(\sim X) \\ &\Leftrightarrow [x]_{A_1} \not\subseteq (\sim X) \neq \emptyset \wedge [x]_{A_2} \not\subseteq (\sim X) \wedge \dots \wedge [x]_{A_m} \not\subseteq (\sim X) \\ &\Leftrightarrow [x]_{A_1} \cap X \neq \emptyset \wedge [x]_{A_2} \cap X \neq \emptyset \wedge \dots \wedge [x]_{A_m} \cap X \neq \emptyset\end{aligned}$$

通过 X 的乐观多粒度下、上近似集, 我们还可以得到其边界域集合。

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

定理 2 设 $IS = \langle U, AT, V, f \rangle$, 令 $A_1, A_2, \dots, A_m$ 是 AT 的 m 个属性子集, $\forall X \subseteq U$ 关于属性 $A_1, A_2, \dots, A_m$ 的乐观多粒度粗糙集有如下性质^[6]:

- 1) $\sum_{i=1}^m \underline{A_i}^O(X) \subseteq X \subseteq \sum_{i=1}^m \overline{A_i}^O(X)$
- 2) $\sum_{i=1}^m \underline{A_i}^O(\emptyset) = \sum_{i=1}^m \overline{A_i}^O(\emptyset) = \emptyset$
 $\sum_{i=1}^m \underline{A_i}^O(U) = \sum_{i=1}^m \overline{A_i}^O(U) = U$
- 3) $\sum_{i=1}^m \underline{A_i}^O(X) = \bigcup_{i=1}^m \underline{A_i}(X)$
 $\sum_{i=1}^m \overline{A_i}^O(X) = \bigcap_{i=1}^m \overline{A_i}(X)$
- 4) $\sum_{i=1}^m \underline{A_i}^O(\sum_{i=1}^m \underline{A_i}^O(X)) = \sum_{i=1}^m \underline{A_i}^O(X)$
 $\sum_{i=1}^m \overline{A_i}^O(\sum_{i=1}^m \overline{A_i}^O(X)) = \sum_{i=1}^m \overline{A_i}^O(X)$
- 5) $\sum_{i=1}^m \underline{A_i}^O(\sim X) = \sim \sum_{i=1}^m \overline{A_i}^O(X)$
 $\sum_{i=1}^m \overline{A_i}^O(\sim X) = \sim \sum_{i=1}^m \underline{A_i}^O(X)$

2.2.2 悲观多粒度粗糙集^[8]

定义 4 设 $IS = \langle U, AT, V, f \rangle$, 令 $A_1, A_2, \dots, A_m$ 是 AT 的 m 个属性子集, $\forall X \subseteq U$ 关于属性 $A_1, A_2, \dots, A_m$ 的悲观多粒度下、上近似集分别记为:

$$\begin{aligned}\sum_{i=1}^m \underline{A_i}^P(X) &= \{x \in U : [x]_{A_1} \subseteq X \subseteq [x]_{A_2} \subseteq X \dots \subseteq [x]_{A_m}\} \\ \sum_{i=1}^m \overline{A_i}^P(X) &= \sim \sum_{i=1}^m \underline{A_i}^P(\sim X)\end{aligned}$$

序偶 $\langle \sum_{i=1}^m \underline{A_i}^P(X), \sum_{i=1}^m \overline{A_i}^P(X) \rangle$ 称为 X 关于属性 $A_1, A_2, \dots, A_m$ 的悲观多粒度粗糙集。

定理 3 设 $IS = \langle U, AT, V, f \rangle$, 令 $A_1, A_2, \dots, A_m$ 是 AT 的 m 个属性子集, $\forall X \subseteq U$ 可得:

$$\sum_{i=1}^m \underline{A_i}^P(X) = \{x \in U : [x]_{A_1} \cap X \neq \emptyset \vee [x]_{A_2} \cap X \neq \emptyset \vee \dots \vee [x]_{A_m} \cap X \neq \emptyset\}$$

证明: 由定义 4 可知:

$$\begin{aligned}x \in \sum_{i=1}^m \underline{A_i}^P(X) &\Leftrightarrow x \notin \sum_{i=1}^m \underline{A_i}^P(\sim X) \\ &\Leftrightarrow [x]_{A_1} \not\subseteq (\sim X) \neq \emptyset \vee [x]_{A_2} \not\subseteq (\sim X) \vee \dots \vee [x]_{A_m} \not\subseteq (\sim X) \\ &\Leftrightarrow [x]_{A_1} \cap X \neq \emptyset \vee [x]_{A_2} \cap X \neq \emptyset \vee \dots \vee [x]_{A_m} \cap X \neq \emptyset\end{aligned}$$

通过 X 的悲观多粒度下、上近似集, 我们还可以得到其边界域集合。

$$BND_{\sum_{i=1}^m A_i}^P(X) = \sum_{i=1}^m \overline{A_i}^P(X) - \sum_{i=1}^m \underline{A_i}^P(X)$$

定理 4 设 $IS = \langle U, AT, V, f \rangle$, 令 $A_1, A_2, \dots, A_m$ 是 AT 的 m 个属性子集, $\forall X \subseteq U$ 关于属性 $A_1, A_2, \dots, A_m$ 的悲观多粒度粗糙集有如下性质^[8]:

- 1) $\sum_{i=1}^m \underline{A_i}^P(X) \subseteq X \subseteq \sum_{i=1}^m \overline{A_i}^P(X)$
- 2) $\sum_{i=1}^m \underline{A_i}^P(\emptyset) = \sum_{i=1}^m \overline{A_i}^P(\emptyset) = \emptyset$
 $\sum_{i=1}^m \underline{A_i}^P(U) = \sum_{i=1}^m \overline{A_i}^P(U) = U$
- 3) $\sum_{i=1}^m \underline{A_i}^P(X) = \bigcap_{i=1}^m \underline{A_i}(X)$
 $\sum_{i=1}^m \overline{A_i}^P(X) = \bigcup_{i=1}^m \overline{A_i}(X)$
- 4) $\sum_{i=1}^m \underline{A_i}^P(\sum_{i=1}^m \underline{A_i}^P(X)) = \sum_{i=1}^m \underline{A_i}^P(X)$
 $\sum_{i=1}^m \overline{A_i}^P(\sum_{i=1}^m \overline{A_i}^P(X)) = \sum_{i=1}^m \overline{A_i}^P(X)$
- 5) $\sum_{i=1}^m \underline{A_i}^P(\sim X) = \sim \sum_{i=1}^m \overline{A_i}^P(X)$
 $\sum_{i=1}^m \overline{A_i}^P(\sim X) = \sim \sum_{i=1}^m \underline{A_i}^P(X)$

3 可变粒度粗糙集

在 $IS = \langle U, AT, V, f \rangle$ 中, 令 $A_1, A_2, \dots, A_m$ 是 AT 的 m 个属性子集。Qian 的乐观多粒度粗糙集的下近似只要判断是

否存在某个知识粒 $[x]_{A_i}$,使得 $[x]_{A_i} \subseteq X$ 成立,如存在,则 x 属于 X 的下近似,这个条件过于宽松;而悲观多粒度粗糙集下近似则要求在所有的知识粒 $[x]_{A_i}$ ($1 \leq i \leq m$),使得 $[x]_{A_i} \subseteq X$ 都成立,此时 x 才属于 X 的下近似,这个要条件又过于严格。本文认为在 m 知识粒度 $[x]_{A_1}, [x]_{A_2}, \dots, [x]_{A_m}$ 下,只要存在相对多数(比如50%)知识粒 $[x]_{A_i}$,使得 $[x]_{A_i} \subseteq X$ 成立,我们就认为 x 属于 X 的下近似,这也符合民主社会生活中的少数服从多数的思想。至于如何量化有多少的知识粒满足 $[x]_{A_i} \subseteq X$,可以通过参数 β 来设定,因此,将其称为 β 可变粒度粗糙集,简称变粒度粗糙集。

3.1 变粒度粗糙集的定义

定义5 设 $IS = \langle U, AT, V, f \rangle$,令 $A_1, A_2, \dots, A_m$ 是 AT 的 m 个属性子集, $A = \{A_1, A_2, \dots, A_m\}$, $0 < \beta \leq 1$,定义 $X \subseteq U$ 关于 A 的 β 可变粒度粗糙集的下、上近似记为

$$\underline{\sum_{i=1}^m A_i}^\beta(X) = \{x \in U : \forall A_i \in T, [x]_{A_i} \subseteq X\}$$

式中, $T \subseteq A \wedge \frac{|T|}{m} \geq \beta$ 。

$$\overline{\sum_{i=1}^m A_i}^\beta(X) = \sim \underline{\sum_{i=1}^m A_i}^\beta(\sim X)$$

注1:定义5中的 $|X|$ 表示集合 X 的基数,集合 T 中的元素就是由所有满足 $[x]_{A_i} \subseteq X$ 中的 A_i 构成的,且满足 $|T|/m \geq \beta$ 。当 $(1-1/m) < \beta \leq 1$ 时,可变粒度粗糙集就退化为乐观多粒度粗糙集;当 $0 < \beta < 1/m$ 时,可变粒度粗糙集就退化为悲观多粒度粗糙集。因此乐观和悲观多粒度粗糙集是可变粒度粗糙集的两种特殊形式,变粒度粗糙集是乐观和悲观多粒度粗糙集的泛化。

通过 X 的 β 可变粒度粗糙集的下、上近似集,还可以得到其边界域集合。

$$BND_{\sum_{i=1}^m A_i}^\beta(X) = \overline{\sum_{i=1}^m A_i}^\beta(X) - \underline{\sum_{i=1}^m A_i}^\beta(X)$$

定理5 设 $IS = \langle U, AT, V, f \rangle$,令 $A_1, A_2, \dots, A_m$ 是 AT 的 m 个属性子集, $A = \{A_1, A_2, \dots, A_m\}$, $0 < \beta \leq 1$, $\forall X \subseteq U$,可得

$$\overline{\sum_{i=1}^m A_i}^\beta(X) = \{x \in U : \forall A_i \in T, [x]_{A_i} \cap X \neq \emptyset\}$$

式中, $T \subseteq A \wedge |T|/m \geq \beta$ 。

证明:由定义5可知

$$\forall x \in \underline{\sum_{i=1}^m A_i}^\beta(X) \Leftrightarrow x \notin \overline{\sum_{i=1}^m A_i}^\beta(\sim X)$$

$$\Leftrightarrow \forall A_i \in T, [x]_{A_i} \not\subseteq (\sim X), \text{ 且 } T \subseteq A, |T|/m \geq \beta$$

$$\Leftrightarrow \forall A_i \in T, [x]_{A_i} \cap X \neq \emptyset, \text{ 且 } T \subseteq A, |T|/m \geq \beta$$

定理6 设 $IS = \langle U, AT, V, f \rangle$,令 $A_1, A_2, \dots, A_m$ 是 AT 的 m 个属性子集, $0 < \beta \leq 1$,则 β 可变粒度粗糙集与乐观和悲观多粒度粗糙集相比有如下的性质

$$\underline{\sum_{i=1}^m A_i}^P(X) \subseteq \underline{\sum_{i=1}^m A_i}^\beta(X) \subseteq \underline{\sum_{i=1}^m A_i}^O(X)$$

$$\overline{\sum_{i=1}^m A_i}^O(X) \subseteq \overline{\sum_{i=1}^m A_i}^\beta(X) \subseteq \overline{\sum_{i=1}^m A_i}^P(X)$$

证明:由定义3、定义4和定义5易知定理6成立。

定理7 设 $IS = \langle U, AT, V, f \rangle$,令 $A_1, A_2, \dots, A_m$ 是 AT 的 m 个属性子集, $0 < \beta \leq 1$,则 β 可变粒度粗糙集有如下的性质

$$(1) \underline{\sum_{i=1}^m A_i}^\beta(X) \subseteq X \subseteq \overline{\sum_{i=1}^m A_i}^\beta(X)$$

$$(2) \underline{\sum_{i=1}^m A_i}^\beta(\emptyset) = \overline{\sum_{i=1}^m A_i}^\beta(\emptyset) = \emptyset$$

$$\underline{\sum_{i=1}^m A_i}^\beta(U) = \overline{\sum_{i=1}^m A_i}^\beta(U) = U$$

$$(3) X \subseteq Y \Rightarrow \underline{\sum_{i=1}^m A_i}^\beta(X) \subseteq \underline{\sum_{i=1}^m A_i}^\beta(Y)$$

$$\overline{\sum_{i=1}^m A_i}^\beta(X) \subseteq \overline{\sum_{i=1}^m A_i}^\beta(Y)$$

$$(4) \beta_1 \leq \beta_2 \Rightarrow \underline{\sum_{i=1}^m A_i}^{\beta_1}(X) \supseteq \underline{\sum_{i=1}^m A_i}^{\beta_2}(X)$$

$$\overline{\sum_{i=1}^m A_i}^{\beta_1}(X) \supseteq \overline{\sum_{i=1}^m A_i}^{\beta_2}(X)$$

例1 表1是一个多专家面试学生的决策信息表, $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}$,条件属性集 $C = \{a_1, a_2, a_3, a_4\}$,表示不同学科的专家(老师)以本领域的知识给出的评价,决策属性 $D = \{d\}$ 表示学生是否通过面试。可以令条件属性集 $A = \{A_1, A_2, A_3, A_4\} = \{\{a_1\}, \{a_2\}, \{a_3\}, \{a_4\}\}$,由 $U/IND(D) = \{D_1, D_2\} = \{\{x_2, x_4, x_6, x_8\}, \{x_1, x_3, x_5, x_7, x_9\}\}$ 。根据定义3可求得决策类的乐观多粒度粗糙集的下、上近似集如下。

$$\underline{\sum_{i=1}^4 A_i}^O(D_1) = \{x_2, x_4, x_8\}, \underline{\sum_{i=1}^4 A_i}^O(D_2)$$

$$= \{x_1, x_3, x_5, x_7, x_9\}$$

$$\overline{\sum_{i=1}^4 A_i}^O(D_1) = \{x_2, x_4, x_6, x_8\}, \overline{\sum_{i=1}^4 A_i}^O(D_2)$$

$$= \{x_1, x_3, x_5, x_6, x_7, x_9\}$$

表1 决策信息表

U	a ₁	a ₂	a ₃	a ₄	d
x ₁	Common	God	God	God	Yes
x ₂	Bad	Bad	Bad	Bad	No
x ₃	Bad	God	God	God	Yes
x ₄	Bad	Common	Common	Bad	No
x ₅	God	Common	Common	God	Yes
x ₆	Bad	Common	Common	Common	No
x ₇	Bad	God	Bad	Common	Yes
x ₈	Bad	Bad	Common	Common	No
x ₉	Bad	God	Bad	Common	Yes

根据定义4可求得决策类的悲观多粒度粗糙集的下、上近似集如下。

$$\underline{\sum_{i=1}^4 A_i}^P(D_1) = \emptyset$$

$$\underline{\sum_{i=1}^4 A_i}^P(D_2) = \{x_1\}$$

$$\overline{\sum_{i=1}^4 A_i}^P(D_1) = \{x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}$$

$$\overline{\sum_{i=1}^4 A_i}^P(D_2) = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}$$

令变粒度的 $\beta = 0.5$,根据定义5可求得决策类的变粒度粗糙集的下、上近似集如下。

$$\underline{\sum_{i=1}^4 A_i}^\beta(D_1) = \{x_2\}$$

$$\underline{\sum_{i=1}^4 A_i}^\beta(D_2) = \{x_1, x_3, x_5\}$$

$$\overline{\sum_{i=1}^4 A_i}^\beta(D_1) = \{x_2, x_4, x_6, x_7, x_8, x_9\}$$

$$\overline{\sum_{i=1}^4 A_i}^\beta(D_2) = \{x_1, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}$$

注2: (1)在决策类 D_2 的乐观多粒度下近似集中, $[x_7]_{\{a_2\}} = \{x_1, x_3, x_7, x_9\} \subseteq D_2$,而 $[x_7]_{\{a_1\}} \not\subseteq D_2$, $[x_7]_{\{a_3\}} \subseteq D_2$, $[x_7]_{\{a_4\}} \not\subseteq D_2$ 。即对象 x_7 只有知识粒 $[x_7]_{\{a_2\}} \subseteq D_2$,或者说专家 a_2 认为其通过面试。尽管在其它的知识粒度下都不是,但仍然将其归于下近似集中,这种决策过于乐观;而 β 可变粒度粗糙集认为要有多数的粒度(β 通常在两类决策时取 $1/2$,在三类决策时取 $1/3$)认为对象通过面试,故 x_7 不是决策类 D_2 的变粒度粗糙集的下近似。

(2)在 D_2 的悲观多粒度下近似集中,只有 $[x_1]_{\{a_1\}}$ 、 $[x_1]_{\{a_2\}}$ 、 $[x_1]_{\{a_3\}}$ 和 $[x_1]_{\{a_4\}}$ 同时 $\subseteq D_2$,此时 x_1 才是悲观多粒度的下近似,这个定义过于严格也过于悲观。而 β 可变粒度粗糙集认为要有多数的粒度满足即可。故虽然 $[x_3]_{\{a_1\}} = \{x_2, x_3, x_4, x_6, x_7, x_8\} \not\subseteq D_2$,但是 $[x_3]_{\{a_2\}} = \{x_1, x_3, x_9\}$, $[x_3]_{\{a_3\}} = \{x_1, x_3\}$, $[x_3]_{\{a_4\}} = \{x_1, x_3, x_5\}$ 都 $\subseteq D_2$,所以 x_3 是

(下转第247页)

in information system[M]//Slowinski R, ed. Intelligent Decision Support Handbook of Applications and Advances of the Rough Sets Theory. Dordrecht: Kluwer Academic Publishers, 1992: 331-362

- [6] Hu Xiao-hua, Cercone N. Learning in relational databases: A rough set approach[J]. Computational Intelligence, 1995, 11(2): 323-338
- [7] 刘少辉, 盛秋戡, 史忠植. 一种新的快速计算正域的方法[J]. 计算机研究与发展, 2003, 40(5): 637-642
- [8] 徐章艳, 刘作鹏, 杨炳儒, 等. 一个复杂度为 $\max(O(|C||U|), O(|C|^2|U/C|))$ 的快速属性约简算法[J]. 计算机学报, 2006, 29

(3): 391-399

- [9] 史君华, 胡学钢. 一种基于粗集的决策表属性约简新算法[J]. 计算机科学, 2006, 33(11): 83-85
- [10] 田卫东, 周创德, 胡学钢, 等. 基于简化分辨矩阵的粗糙集属性约简算法[J]. 计算机科学, 2008, 35(3): 209-212
- [11] 王兵, 陈善本. 一种基于差别矩阵的属性约简完备算法[J]. 上海交通大学学报, 2004, 38(1): 43-46
- [12] 蒋瑜, 王燮, 叶振. 基于差别矩阵的 Rough 集属性约简算法[J]. 系统仿真学报, 2008, 20(14): 3717-3725
- [13] <http://www.ics.uci.edu/~mllearn/databases/>
- [14] <http://www.cs.uregina.ca/~roughset>

(上接第 222 页)

决策类 D_2 的可变粒度粗糙集的下近似。

3.2 变粒度粗糙集的量度

从粗糙集近似的角度来看, 知识的粗糙性(不确定性)是由于边界域的存在而造成的, 边界域越大, 则粗糙程度就越大。下面从粗糙精度、属性依赖度和概念的可精确近似程度等方面来度量变粒度粗糙集。

定义 6 设 $IS = \langle U, AT, V, f \rangle$, 令 $A_1, A_2, \dots, A_m$ 是 AT 的 m 个属性子集, $A = \{A_1, A_2, \dots, A_m\}$, $0 < \beta \leq 1$, $\forall X \subseteq U (X \neq \emptyset)$ 关于 A 的乐观多粒度粗糙集、悲观多粒度粗糙集和可变粒度粗糙集的粗糙度因子分别定义为 α_O 、 α_P 和 α_β 。

$$\alpha_O(\sum_{i=1}^m A_i, X) = \frac{|\sum_{i=1}^m A_i^O(X)|}{|\sum_{i=1}^m A_i^O(X)|}$$

$$\alpha_P(\sum_{i=1}^m A_i, X) = \frac{|\sum_{i=1}^m A_i^P(X)|}{|\sum_{i=1}^m A_i^P(X)|}$$

$$\alpha_\beta(\sum_{i=1}^m A_i, X) = \frac{|\sum_{i=1}^m A_i^\beta(X)|}{|\sum_{i=1}^m A_i^\beta(X)|}$$

定义 7 设 $DIS = \langle U, C \cup \{d\}, V, f \rangle$, 令 $A, A_2, \dots, A_m$ 是 C 的 m 个属性子集, $A = \{A_1, A_2, \dots, A_m\}$, $0 < \beta \leq 1$, $D = \{D_1, D_2, \dots, D_i\}$ 是由决策属性集 $\{d\}$ 在论域 U 上导出的划分, 则决策类 D 对属性子集 A 的粗糙分类能力, 也即对 A 的属性依赖程度记为 γ , 且乐观多粒度、悲观多粒度和可变粒度粗糙集的依赖度因子分别记为 γ_O 、 γ_P 和 γ_β 。

$$\gamma_O(\sum_{i=1}^m A_i, D) = \frac{\sum\{|\sum_{i=1}^m A_i^O(D_i)| : D_i \in D\}}{|U|}$$

$$\gamma_P(\sum_{i=1}^m A_i, D) = \frac{\sum\{|\sum_{i=1}^m A_i^P(D_i)| : D_i \in D\}}{|U|}$$

$$\gamma_\beta(\sum_{i=1}^m A_i, D) = \frac{\sum\{|\sum_{i=1}^m A_i^\beta(D_i)| : D_i \in D\}}{|U|}$$

定义 8 设 $IS = \langle U, AT, V, f \rangle$, 令 $A_1, A_2, \dots, A_m$ 是 AT 的 m 个属性子集, $A = \{A_1, A_2, \dots, A_m\}$, $0 < \beta \leq 1$, $\forall X \subseteq U (X \neq \emptyset)$ 在属性子集 A 下可以被精确近似的程度定义为 π , 且乐观多粒度、悲观多粒度和可变粒度粗糙集的精确近似程度因子分别记为 π_O 、 π_P 和 π_β 。

$$\pi_O(\sum_{i=1}^m A_i, X) = \frac{|\sum_{i=1}^m A_i^O(X)|}{|X|}$$

$$\pi_P(\sum_{i=1}^m A_i, X) = \frac{|\sum_{i=1}^m A_i^P(X)|}{|X|}$$

$$\pi_\beta(\sum_{i=1}^m A_i, X) = \frac{|\sum_{i=1}^m A_i^\beta(X)|}{|X|}$$

性质 1 设 $IS = \langle U, AT, V, f \rangle$, 令 $A_1, A_2, \dots, A_m$ 是 AT 的 m 个属性子集, $A = \{A_1, A_2, \dots, A_m\}$, $0 < \beta \leq 1$, $\forall X \subseteq U (X \neq \emptyset)$ 关于属性子集 A 的变粒度粗糙精度因子 α_β 、可精确近似程度 π_β 与多粒度粗糙集对应的度量因子有如下的关系成立。

$$\alpha_P(\sum_{i=1}^m A_i, X) \leq \alpha_\beta(\sum_{i=1}^m A_i, X) \leq \alpha_O(\sum_{i=1}^m A_i, X)$$

$$\pi_P(\sum_{i=1}^m A_i, X) \leq \pi_\beta(\sum_{i=1}^m A_i, X) \leq \pi_O(\sum_{i=1}^m A_i, X)$$

性质 2 设 $DIS = \langle U, C \cup \{d\}, V, f \rangle$, 令 $A_1, A_2, \dots, A_m$ 是 C 的 m 个属性子集, $A = \{A_1, A_2, \dots, A_m\}$, $0 < \beta \leq 1$, $D = \{D_1, D_2, \dots, D_i\}$ 是由决策属性集 $\{d\}$ 在论域 U 上导出的划分, 则决策类 D 对属性子集 A 的依赖度因子 γ_β 与多粒度粗糙集对应的度量因子有如下的关系成立。

$$\gamma_P(\sum_{i=1}^m A_i, X) \leq \gamma_\beta(\sum_{i=1}^m A_i, X) \leq \gamma_O(\sum_{i=1}^m A_i, X)$$

说明: 性质(1)和性质(2)由定理 6 很容易得到。

结束语 多粒度粗糙集模型是 Qian 从粒计算的角度出发提出的一种新的粗糙集模型, 笔者在深入地研究了乐观和悲观多粒度粗糙集性质的基础上, 提出了可变粒度粗糙集模型, 该模型是 Qian 的两种多粒度粗糙集模型的泛化。笔者将进一步研究可变粒度粗糙集的知识获取和约简等问题。

参考文献

- [1] Pawlak Z. Rough sets[J]. International Journal of Computer and Information Sciences, 1984, 11(5): 341-356
- [2] Kryszkiewicz M. Rough set approach to incomplete information systems[J]. Information Sciences, 1998, 112(1-4): 39-49
- [3] 王国胤. Rough 集理论在不完备信息系统中的扩充[J]. 计算机研究与发展, 2002, 39(10): 1238-1243
- [4] Stefanowski J, Tsoukias A. Incomplete information tables and rough classification[J]. Computational Intelligence, 2001, 17(3): 545-566
- [5] 苗夺谦, 王国胤, 刘清, 等. 粒计算: 过去、未来和展望[M]. 北京: 科学出版社, 2007
- [6] Qian Y H, Liang J Y, Yao Y Y C, et al. MGRS: a multigranulation rough set[J]. Information Sciences, 2010, 180(5): 949-970
- [7] Qian Y H, Liang J Y, Dang C Y. Incomplete multigranulation rough set[J]. IEEE Transactions on Systems, Man and Cybernetics, Part A, 2010, 40(2): 420-431
- [8] Qian Y H, Liang J Y, Wu W. Pessimistic rough decision[C]// Second International Workshop on Rough Sets Theory. Zhoushan, P. R. China, October 2010: 440-449