

基于集对分析方法的不完备信息系统的扩充粗糙集模型^{*})

刘富春

(广东工业大学应用数学学院 广州 510090)

摘要 目前对不完备信息系统的扩充主要有基于容差关系、基于相似关系和基于限制容差关系等方法。集对分析方法是研究集合之间相互关系的一种新理论,它在处理个体之间确定性联系和不确定性联系方面已经得到了广泛的应用。本文运用集对分析方法,先分析了现有的扩充粗糙集模型所存在的局限性,然后通过引入个体之间的集对联系度概念,提出了两个基于集对分析方法的扩充粗糙集模型。它们既保留了限制容差关系扩充模型的优点,又丢弃了现有的扩充模型的局限性。最后,通过实例说明这两个基于集对分析方法的扩充模型对不完备信息系统的处理更加合理、可行、有效。

关键词 不完备信息系统,容差关系,相似关系,限制容差关系,集对分析方法

Extension of Rough Set under Incomplete Information System Based on Set-pair Analysis

LIU Fu-Chun

(Department of Applied Mathematics, Guangdong University of Technology, Guangzhou 510090)

Abstract Now there are several extensions for incomplete information system, such as extensions based on tolerance relation, similarity relation, or limited tolerance relation. Unfortunately, these extensions have their own limitations. In this paper, two new extensions of rough set based on set-pair analysis are presented by introduced the conception of relationship between elements. The new extensions not only reserve the advantage of limited tolerance relation, but also desert the limitations of the existing extensions. By an example, it is verified that the new extensions are more reasonable and effective than the existing extensions in practice.

Keywords Incomplete information system, Tolerance relation, Similarity relation, Limited tolerance relation, Set-pair analysis

1 引言

由 Pawlak 等人提出的粗糙集理论^[1],是一种新的处理不完整与不确定知识的软计算工具。其主要思想是在保持分类能力不变的前提下,将知识和数据理解为某种等价关系下对特定空间的划分,以描述知识的不确定性。由于粗糙集在处理不完整与不确定性方面具有很好的性能,它在数据挖掘、人工智能、机器学习和模式识别等领域已经取得了很大的成功^[2~6]。

但是,应用这种经典粗糙集理论的一个重要前提就是它的处理对象必须是完备信息系统。而在现实生活中,由于测量误差、数据获取能力不足等原因,使得大量的信息系统都是不完备的,因此需要对经典粗糙集理论进行适当的扩充^[2,3,6~9]。

目前,主要有基于容差关系^[2]、相似关系^[3]、量化容差关系^[3]和限制容差关系^[6]的扩充模型。尽管与其它扩充模型相比较,基于限制容差关系的扩充模型有更多优点,且更合理^[6],但是它和其它扩充模型一样,都存在以下共同的局限性:一方面,个体对象之间只要有一个属性值不同,就认为是完全不同的,必须被划分在不同的类中,导致知识库中的划分分类划分得太细、个数太多,特别不适用于对大型不完备信息系统的处理。另一方面,对于一个属性值都不确定的个体对

象,必须单独处理^[6],这也给不完备信息系统的处理带来不必要的麻烦。

集对分析方法^[7]是赵克勤教授近年来提出的、用于研究集合之间相互关系的一种新理论,其核心思想是把被研究的客观事物之确定性联系和不确定性联系作为一个系统来处理,现在它已经得到了广泛的应用^[7~9]。本文在第2节,利用集对分析方法,对现有的扩充模型所存在的局限性进行了分析。在第3节,通过引入个体之间的集对联系度概念,定义了一个集对 α 相似限制容差关系和一个优化集对 α 相似限制容差关系,提出了两个基于集对分析方法的新的扩充粗糙集模型。这两个扩充模型既保留了限制容差关系的扩充模型的优点,又克服了现有扩充模型的局限性,不仅可以通过对 α 的调节和控制,保证划分的准确性和灵活性,而且对于所有属性值都不确定的个体对象,不必单独处理。在第4节,用一个实例说明这两个基于集对分析方法的扩充模型对不完备信息系统的处理更加合理、可行和有效。

2 用集对分析方法对已有扩充模型的局限性的分析

集对分析方法是由赵克勤教授近年来提出的、用于研究集合之间相互关系的一种新理论,其核心思想是把被研究的客观事物之确定性联系和不确定性联系作为一个系统来分析

^{*})广东省自然科学基金项目(020146,031541)和广东工业大学青年基金项目(042027)。刘富春 副教授,主要研究领域为粗糙集理论、数理逻辑及其在计算机科学中的应用。

处理,现在已经得到了广泛的应用^[7-9]。下面用集对分析方法,对基于容差关系、相似关系和限制容差关系等现有的主要扩充模型所存在的局限性进行分析。

定义 1 设 (U, A) 是一个不完备信息系统,其中 U 是对象集, A 是属性集。 $\forall x, y \in U$,定义 x 与 y 的集对属性集如下:

- (1) $M(x, y) = \{a \in A | a(x) \neq * \wedge a(y) \neq * \wedge a(x) \neq a(y)\}$ (取值都明确但不相等的属性个数);
- (2) $N(x, y) = \{a \in A | a(x) \neq * \wedge a(y) \neq * \wedge a(x) = a(y)\}$ (取值都明确且相等的属性个数);
- (3) $Q(x, y) = \{a \in A | a(x) = * \vee a(y) = * \vee a(x) = a(y)\}$ (取值至少有一个不明确或者取值相等的属性个数);
- (4) $K(x, y) = \{a \in A | a(x) = * \vee a(y) = *\}$ (取值至少有一个不明确的属性个数);
- (5) $K_1(x, y) = \{a \in A | a(x) \neq * \wedge a(y) = *\}$ (x 取值明确但 y 取值不明确的属性个数);
- (6) $K_2(x, y) = \{a \in A | a(x) = * \wedge a(y) \neq *\}$ (x 取值不明确但 y 取值明确的属性个数);
- (7) $K_3(x, y) = \{a \in A | a(x) = * \wedge a(y) = *\}$ (取值都不明确的属性个数)。

定理 1 设 (U, A) 是一个不完备信息系统,其中 U 是对象集, A 是属性集,则 $\forall x, y \in U$,

- (1) $K(x, y) = K_1(x, y) \cup K_2(x, y) \cup K_3(x, y)$;
- (2) $A = M(x, y) \cup N(x, y) \cup K(x, y)$;
- (3) $Q(x, y) = N(x, y) \cup K(x, y)$;
- (4) $A - M(x, y) = Q(x, y)$ 。

证明:由定义 1 容易证得。

定义 2 设 (U, A) 是一个不完备信息系统,其中 U 是对象集, A 是属性集, $\forall x, y \in U$,定义 x 与 y 的集对联系度 $U(x, y)$ 为一个五元组:

$$U(x, y) = (s_0, s_1, s_2, s_3, s_4)$$

这里 $s_0 = |M(x, y)|/n$, $s_1 = |K_1(x, y)|/n$, $s_2 = |K_2(x, y)|/n$, $s_3 = |K_3(x, y)|/n$, $s_4 = |N(x, y)|/n$, 其中 $n = |A|$ 。

显然, $s_0 + s_1 + s_2 + s_3 + s_4 = 1$ 。

定义 3^[2] (基于容差关系的扩充粗糙集模型) 容差关系 T 是指: $\forall x, y \in U$,

$(x, y) \in T$ 当且仅当 $\forall a \in A (a(x) = * \vee a(y) = * \vee a(x) = a(y))$ 。

设 $T(x)$ 为个体 x 的容差类,则在容差关系下对象集 X 的下近似($\underline{TX}$)和上近似($\overline{TX}$)分别为:

$$\underline{TX} = \{x \in U | T(x) \subseteq X\}; \overline{TX} = \{x \in U | T(x) \cap X \neq \emptyset\}。$$

定理 2 设 (U, A) 是不完备信息系统, T 是容差关系。 $\forall x, y \in U, U(x, y) = (s_0, s_1, s_2, s_3, s_4)$, 则下面条件相互等价:

- (1) $(x, y) \in T$; (2) $A = Q(x, y)$; (3) $M(x, y) = \emptyset$;
- (4) $s_0 = 0$; (5) $s_1 + s_2 + s_3 + s_4 = 1$ 。

证明: (1) $\Leftrightarrow$ (2)。由定义 1 和定义 3 得。

$(x, y) \in T \Leftrightarrow \forall a \in A (a(x) = * \vee a(y) = * \vee a(x) = a(y)) \Leftrightarrow A = Q(x, y)$ 。

(2) $\Leftrightarrow$ (3) $\Leftrightarrow$ (4)。由定理 1 和定义 2 即得。

(4) $\Leftrightarrow$ (5)。由定义 2 和 $s_0 + s_1 + s_2 + s_3 + s_4 = 1$ 可得。

从定义 3 和定理 2 可看出,基于容差关系的扩充粗糙集

模型存在明显不足:不确定值“*”实际上被认为是和所有已知属性值都相等,因此会导致那些没有任何明确相同的已知属性值的个体(如 $x = (1, *, 3, *, 5, *, 7, *, 9, *)$ 和 $y = (*, 8, *, 6, *, 4, *, 2, *, 0)$),都被误判在同一个容差类中。

定义 4^[3] (基于相似关系的扩充粗糙集模型) 相似关系 S 是指: $\forall x, y \in U, (x, y) \in S$ 当且仅当 $\forall a \in A (a(x) = * \vee a(y) = a(y))$ 。

个体对象 x 的相似类为: $S(x) = \{y \in U | (y, x) \in S\}$; $S^{-1}(x) = \{y \in U | (x, y) \in S\}$ 。在相似关系下对象集 X 的下近似($\underline{SX}$)和上近似($\overline{SX}$)分别定义为:

$$\underline{SX} = \{x \in U | S^{-1}(x) \subseteq X\}; \overline{SX} = \{x \in U | S(x) \cap X \neq \emptyset\}。$$

定理 3 设 (U, A) 是一个不完备信息系统, S 是相似关系, $\forall x, y \in U, U(x, y) = (s_0, s_1, s_2, s_3, s_4)$, 则下面条件相互等价:

- (1) $(x, y) \in S$;
- (2) $A = N(x, y) \cup K_2(x, y) \cup K_3(x, y)$;
- (3) $M(x, y) = K_1(x, y) = \emptyset$;
- (4) $s_0 = s_1 = 0$;
- (5) $s_2 + s_3 + s_4 = 1$ 。

证明: (1) $\Leftrightarrow$ (3)。由定义 1 和定义 4 得, $(x, y) \in S \Leftrightarrow \forall a \in A (a(x) = * \vee a(y) = a(y))$

$$\Leftrightarrow \exists a \in A (a(x) \neq * \wedge a(x) \neq a(y))$$

$$\Leftrightarrow \exists a \in A ((a(x) \neq * \wedge a(x) \neq a(y) \wedge a(y) = *) \vee (a(x) \neq * \wedge a(x) \neq a(y) \wedge a(y) \neq *))$$

$$\Leftrightarrow \forall a \in A (\neg(a(x) \neq * \wedge a(y) = *) \wedge \neg(a(x) \neq * \wedge a(x) \neq a(y) \wedge a(y) \neq *))$$

$$\Leftrightarrow \forall a \in A (\neg(a(x) \neq * \wedge a(y) = *) \wedge \forall a \in A (\neg(a(x) \neq * \wedge a(x) \neq a(y) \wedge a(y) \neq *))$$

$$\Leftrightarrow K_1(x, y) = \emptyset \wedge M(x, y) = \emptyset。$$

(2) $\Leftrightarrow$ (3) $\Leftrightarrow$ (4)。由定理 1 和定义 2 即得。

(4) $\Leftrightarrow$ (5)。由定义 2 和 $s_0 + s_1 + s_2 + s_3 + s_4 = 1$ 可得。

从定义 4 和定理 3 不难发现,基于相似关系的扩充模型存在以下局限性:一些个体对象明显有大量相同的已知属性值,直观上就可以判断为相似(如 $x = (*, 1, 2, 3, 4, 5, 6, 7, 8, 9)$ 和 $y = (0, *, 2, 3, 4, 5, 6, 7, 8, 9)$),却不满足相似关系而被划分在不同的相似类中。

定义 5^[5] (基于限制容差关系的扩充模型) 限制容差关系 L 是指: $\forall x, y \in U, (x, y) \in L$ 当且仅当 $\forall a \in A (a(x) = a(y) = *) \vee (P(x) \cap P(y) \neq \emptyset \wedge \forall a \in A ((a(x) \neq *) \wedge (a(y) \neq *) \rightarrow (a(x) = a(y))))$ 。

其中 $P(x) = \{a \in A | a(x) \neq *\}$ 。设 $L(x)$ 是个体 x 的限制容差类,则在限制容差关系下对象集 X 的下近似($\underline{LX}$)和上近似($\overline{LX}$)分别定义为: $\underline{LX} = \{x \in U | L(x) \subseteq X\}$; $\overline{LX} = \{x \in U | L(x) \cap X \neq \emptyset\}$ 。

定理 4 设 (U, A) 是一个不完备信息系统, L 是限制容差关系, $\forall x, y \in U, U(x, y) = (s_0, s_1, s_2, s_3, s_4)$, 则下面条件相互等价:

- (1) $(x, y) \in L$;
- (2) $A = K_3(x, y) \vee (M(x, y) = \emptyset \wedge N(x, y) \neq \emptyset)$;
- (3) $s_3 = 1 \vee (s_4 > 0 \wedge s_0 = 0)$;
- (4) $s_3 = 1 \vee (s_4 > 0 \wedge s_1 + s_2 + s_3 + s_4 = 1)$ 。

证明: (1) $\Leftrightarrow$ (2)。由定义 1 和定义 5 得, $(x, y) \in L$

$$\Leftrightarrow \forall a \in A (a(x) = a(y) = *) \vee (P(x) \cap P(y) \neq \emptyset \wedge \forall a \in A ((a(x) \neq *) \wedge (a(y) \neq *) \rightarrow (a(x) = a(y))))$$

$$\Leftrightarrow A = K_3(x, y) \vee (\exists a \in A (a(x) \neq * \wedge a(y) \neq * \wedge a(x) = a(y)) \wedge \forall a \in A (\neg (a(x) \neq * \wedge a(y) \neq *) \vee a(x) = a(y)))$$

$$\Leftrightarrow A = K_3(x, y) \vee (N(x, y) \neq \emptyset \wedge \forall a \in A (a(x) = * \vee a(y) = * \vee a(x) = a(y)))$$

$$\Leftrightarrow A = K_3(x, y) \vee (N(x, y) \neq \emptyset \wedge A = Q(x, y))$$

$$\Leftrightarrow A = K_3(x, y) \vee (N(x, y) \neq \emptyset \wedge M(x, y) = \emptyset)。$$

$$(2) \Leftrightarrow (3) \Leftrightarrow (4)。由定理1、定义2和s_0 + s_1 + s_2 + s_3 + s_4 = 1即得。$$

限制容差关系介于容差关系和相似关系之间^[6],因此与其它扩充模型相比较,基于限制容差关系的扩充粗糙集模型有许多优点,且更加合理^[6],但是它和其它扩充模型一样,都存在以下共同的局限性:一方面,个体对象之间只要有一个属性值不同就认为是完全不同的,必须被划分在不同的容差类中。这种模型将会导致容差类的划分太细、个数太多,特别不适用于大型不完备信息系统中数据的处理。另一方面,对于一个属性值都不确定的个体对象(如(*, *, ..., *)),必须单独处理^[6],这给不完备信息系统的处理带来极大不便。

3 基于集对分析方法的扩充粗糙集模型

鉴于以上扩充粗糙集模型的局限性,下面我们应用集对分析的思想方法,提出两个新的扩充粗糙集模型,它们既丢弃了以上扩充模型的局限性,又保留了限制容差关系扩充模型的优点。

3.1 模型一:基于集对 α 相似限制容差关系的扩充粗糙集模型

定义6 设 (U, A) 是一个不完备信息系统, $0 \leq \alpha \leq 1$,则 U 上的集对 α 相似限制容差关系 $SP_1(\alpha)$ 是指二元关系:

$$SP_1(\alpha) = \{(x, y) \in U \times U \mid U(x, y) = (s_0, s_1, s_2, s_3, s_4) \wedge (s_3 = 1 \vee (s_4 > 0 \wedge s_1 + s_2 + s_3 + s_4 \geq \alpha))\}。$$

个体对象 x 的集对 α 相似限制容差类 $SP_1(\alpha)(x)$ 和对对象集 X 的下近似集 $\underline{SP}_1(\alpha)(X)$ 、上近似集 $\overline{SP}_1(\alpha)(X)$ 分别定义为:

$$\underline{SP}_1(\alpha)(X) = \{x \in U \mid SP_1(\alpha)(x) \subseteq X\};$$

$$\overline{SP}_1(\alpha)(X) = \{x \in U \mid SP_1(\alpha)(x) \cap X \neq \emptyset\}。$$

定理5 当 $\alpha=1$ 时, $SP_1(\alpha) = L$ 。

证明:由定义6和定理4得, $\alpha=1$ 时,

$$(x, y) \in SP_1(\alpha) \Leftrightarrow s_3 = 1 \vee (s_4 > 0 \wedge s_1 + s_2 + s_3 + s_4 \geq \alpha)$$

$$\Leftrightarrow s_3 = 1 \vee (s_4 > 0 \wedge s_1 + s_2 + s_3 + s_4 = 1) \Leftrightarrow (x, y) \in L。$$

由定理5可知,基于限制容差关系的扩充模型就是基于集对 α 相似限制容差关系的扩充粗糙集模型当 $\alpha=1$ 时的特殊情形。基于集对 α 相似限制容差关系的扩充粗糙集模型,是基于限制容差关系的扩充粗糙集模型的推广和改进,更适合不完备信息系统的处理。例如,个体对象 $x = (1, 2, 3, 4, \dots, 100)$ 和 $y = (2, 2, 3, 4, \dots, 100)$,在100个属性中有且仅有一个属性值是不同的。在基于限制容差关系的扩充模型中, x 和 y 被划分在不同的容差类中。但是在基于集对 α 相似限制容差关系的扩充粗糙集模型中,可以将 x 和 y 划分在相同的相似划分类中,并且相似程度仍然高达 $\alpha=0.99$ 。因此,这种模型既保留了基于限制容差关系的扩充粗糙集模型的优点,同时可以通过对 α 的调节,确保这种相似容差类划分的灵活性和准确性,更符合实际情况,特别适合大型不完备信息系统中数据的处理。

中数据的处理。

3.2 模型二:基于优化集对 α 相似限制容差关系的扩充粗糙集模型

尽管上述模型一方面对现有的扩充模型进行了较大的推广和改进,但是与其它模型一样,仍然存在以下不足:对于一个属性值都不确定的个体对象(如 $x = (*, *, \dots, *)$ 和 $y = (*, *, \dots, *)$)都属于同一个划分类,是不可分辨的。实际上,它们只是存在相同的可能性,而且如果能够将不确定属性值补齐的话,这些个体相同的可能性是非常小的。因此,将它们认为是不可分辨、近似相同而划分在同一个划分类就与实际情况不相符合。另一方面,如果按照这些模型,这些个体都必须单独处理^[6],这给不完备信息系统的处理带来极大不便。为此,在模型一的基础上,定义一个优化集对 α 相似限制容差关系,并由此提出一个基于优化集对 α 相似限制容差关系的扩充粗糙集模型,它既保留了模型一的优点,同时又克服了上面指出的不足。

定义7 设 (U, A) 是一个不完备信息系统, $0 \leq \alpha \leq 1$,则 U 上的优化集对 α 相似限制容差关系 $SP_2(\alpha)$ 是指二元关系:

$$SP_2(\alpha) = \{(x, y) \in U \times U \mid U(x, y) = (s_0, s_1, s_2, s_3, s_4) \wedge s_4 > 0 \wedge s_1 + s_2 + s_3 + s_4 \geq \alpha\} \cup I$$

这里 $I = \{(x, x) \mid x \in U\}$ 是 U 上的恒等关系。

设个体对象 x 的优化集对 α 相似限制容差类为 $SP_2(\alpha)(x)$,则对象集 X 的下近似集和上近似集定义为: $\underline{SP}_2(\alpha)(X) = \{x \in U \mid SP_2(\alpha)(x) \subseteq X\}$; $\overline{SP}_2(\alpha)(X) = \{x \in U \mid SP_2(\alpha)(x) \cap X \neq \emptyset\}$ 。

定理6 设 (U, A) 是一个不完备信息系统, $0 \leq \alpha \leq 1$, $SP_1(\alpha)$ 是 U 上的集对 α 相似限制容差关系, $SP_2(\alpha)$ 是 U 上的优化集对 α 相似限制容差关系,则 $SP_2(\alpha) \subseteq SP_1(\alpha)$ 。

证明: $\forall (x, y) \in SP_2(\alpha)$,则有下面两种情形:

情形1. $(x, y) \in I$,即 $x=y$ 。

(1)如果 $\forall a \in A$ 都有 $a(x) = *$,则 $s_3 = 1$ 。因此 $U(x, y) = U(x, x) = (s_0, s_1, s_2, s_3, s_4)$,其中 $s_3 = 1$ 。

由定义6知, $(x, y) \in SP_1(\alpha)$ 。

(2)如果 $\exists a \in A$,使得 $a(x) \neq *$,则 $U(x, y) = U(x, x) = (s_0, s_1, s_2, s_3, s_4)$,其中 $s_4 > 0$ 且 $s_1 + s_2 + s_3 + s_4 = 1 \geq \alpha$ 。由定义6,也有 $(x, y) \in SP_1(\alpha)$ 。

情形2. $x \neq y$,则由定义7得, $U(x, y) = (s_0, s_1, s_2, s_3, s_4)$,其中 $s_4 > 0$ 且 $s_1 + s_2 + s_3 + s_4 \geq \alpha$,由定义6知, $(x, y) \in SP_1(\alpha)$ 。

综合情形1和情形2,有 $SP_2(\alpha) \subseteq SP_1(\alpha)$ 。

定理7 设 (U, A) 是一个不完备信息系统, $0 \leq \alpha \leq 1$,则对象集 X 的集对 α 相似限制容差关系的上、下近似集和优化集对 α 相似限制容差关系的上、下近似集有下列关系式:

$$(1) \underline{SP}_1(\alpha)(X) \subseteq \underline{SP}_2(\alpha)(X)$$

$$(2) \overline{SP}_2(\alpha)(X) \subseteq \overline{SP}_1(\alpha)(X)$$

证明:(1)对 $\forall x \in \underline{SP}_1(\alpha)(X)$,有 $SP_1(\alpha)(x) \subseteq X$ 。根据定理6, $SP_2(\alpha) \subseteq SP_1(\alpha)$ 得, $\underline{SP}_2(\alpha)(x) \subseteq X$,即 $x \in \underline{SP}_2(\alpha)(X)$,所以 $\underline{SP}_1(\alpha)(X) \subseteq \underline{SP}_2(\alpha)(X)$ 。

(2)对 $\forall x \in \overline{SP}_2(\alpha)(X)$,都有 $SP_2(\alpha)(x) \cap X \neq \emptyset$ 。根据定理6, $SP_2(\alpha) \subseteq SP_1(\alpha)$ 得, $SP_1(\alpha)(x) \cap X \neq \emptyset$,即 $x \in \overline{SP}_1(\alpha)(X)$,所以 $\overline{SP}_2(\alpha)(X) \subseteq \overline{SP}_1(\alpha)(X)$ 。

4 实例分析

下面用一个不完备信息系统实例,来分析、比较本文建立

的两个基于集对分析方法的扩充粗糙集模型($\alpha=1, \alpha=0.8$)和基于限制容差关系的扩充粗糙集模型^[6]之间的性能,说明在不完备信息系统中,基于集对分析方法的扩充模型对不完整与不确定知识和数据的处理更加合理、可行、有效。

设不完备信息系统(U, A)如表 1, 其中对象集 $U=\{x_1, x_2, \dots, x_{10}\}$, 属性集 $A=\{a_1, a_2, \dots, a_{10}\}$ 。

表 1 一个不完备信息系统

$\begin{matrix} A \\ U \end{matrix}$	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
x_1	3	2	1	0	0	*	1	1	*	3
x_2	*	2	*	1	1	2	*	1	1	3
x_3	2	*	*	3	0	*	2	3	1	*
x_4	*	0	0	*	0	2	*	*	2	1
x_5	3	2	1	3	*	2	3	*	2	1
x_6	2	*	*	*	*	*	2	*	*	*
x_7	*	2	*	*	0	*	2	2	2	2
x_8	3	2	2	0	0	*	1	1	*	3
x_9	*	*	*	*	*	*	*	*	*	*
x_{10}	*	*	*	*	*	*	*	*	*	*

(1) 当 $\alpha=1$ 时, 由定理 5 得, 集对 α 相似限制容差类退化为限制容差类, 因此集对 α 相似限制容差类、优化集对 α 相似限制容差类和限制容差类分别为:

$$\begin{aligned} SP_1(\alpha)(x_1) &= SP_2(\alpha)(x_1) = L(x_1) = \{x_1\}; \\ SP_1(\alpha)(x_2) &= SP_2(\alpha)(x_2) = L(x_2) = \{x_2\}; \\ SP_1(\alpha)(x_3) &= SP_2(\alpha)(x_3) = L(x_3) = \{x_3, x_6\}; \\ SP_1(\alpha)(x_4) &= SP_2(\alpha)(x_4) = L(x_4) = \{x_4\}; \\ SP_1(\alpha)(x_5) &= SP_2(\alpha)(x_5) = L(x_5) = \{x_5\}; \\ SP_1(\alpha)(x_6) &= SP_2(\alpha)(x_6) = L(x_6) = \{x_3, x_6, x_7\}; \\ SP_1(\alpha)(x_7) &= SP_2(\alpha)(x_7) = L(x_7) = \{x_6, x_7\}; \\ SP_1(\alpha)(x_8) &= SP_2(\alpha)(x_8) = L(x_8) = \{x_8\}; \\ SP_1(\alpha)(x_9) &= L(x_9) = \{x_9, x_{10}\}; \\ \text{但是 } SP_2(\alpha)(x_9) &= \{x_9\}; \\ SP_1(\alpha)(x_{10}) &= L(x_{10}) = \{x_9, x_{10}\}; \\ \text{但是 } SP_2(\alpha)(x_{10}) &= \{x_{10}\}. \end{aligned}$$

尽管个体 x_1 与 x_8, x_3 与 x_4 很相似, 但是由于要求相似程度 $\alpha=1$, 因此它们还是属于不同的容差类。另外, 关系 $SP_1(\alpha)$ 和 L 都将个体 x_9 与 x_{10} 划分在同一个类中, 认为它们是不可分辨的, 并且必须与其它个体对象分开单独处理。事实上, 它们相同的可能性非常小, 因此, $SP_2(\alpha)$ 将它们划分在不同类中更符合实际情况, 而且它们也不必与其它个体对象分开处理。

(2) 当 $\alpha=0.8$ 时, 集对 α 相似限制容差类、优化集对 α 相似限制容差类和限制容差类分别为:

$$\begin{aligned} SP_1(\alpha)(x_1) &= SP_2(\alpha)(x_1) = \{x_1, x_2, x_8\}; \\ SP_1(\alpha)(x_2) &= SP_2(\alpha)(x_2) = \{x_1, x_2, x_8\}; \\ SP_1(\alpha)(x_3) &= SP_2(\alpha)(x_3) = \{x_3, x_4, x_6, x_7\}; \\ SP_1(\alpha)(x_4) &= SP_2(\alpha)(x_4) = \{x_3, x_4, x_5, x_7\}; \\ SP_1(\alpha)(x_5) &= SP_2(\alpha)(x_5) = \{x_4, x_5, x_7\}; \end{aligned}$$

$$\begin{aligned} SP_1(\alpha)(x_6) &= SP_2(\alpha)(x_6) = \{x_3, x_6, x_7\}; \\ SP_1(\alpha)(x_7) &= SP_2(\alpha)(x_7) = \{x_3, x_4, x_5, x_6, x_7\}; \\ SP_1(\alpha)(x_8) &= SP_2(\alpha)(x_8) = \{x_1, x_2, x_8\}; \\ SP_1(\alpha)(x_9) &= \{x_9, x_{10}\}; \\ \text{但是 } SP_2(\alpha)(x_9) &= \{x_9\}; \\ SP_1(\alpha)(x_{10}) &= \{x_9, x_{10}\}; \\ \text{但是 } SP_2(\alpha)(x_{10}) &= \{x_{10}\}. \end{aligned}$$

由于个体 x_2 与 x_8, x_4 与 x_7, x_5 与 x_7, x_6 与 x_7 在 10 个属性中仅有两个属性值不同, 因此当相似程度 $\alpha=0.8$ 时, 它们可以近似认为是不可分辨的。

从这个实例可以看出, 上面建立的两个基于集对分析方法的扩充粗糙集模型既保留了限制容差关系的扩充模型的优点, 同时可以结合实际情况, 根据需要来调节和控制相似程度 α 的取值, 以确保其准确性的基础上增加其灵活性。同时, 对于一个属性值都不确定的个体也不必与其它个体对象分开处理。因此, 在不完备信息系统, 特别是大型不完备信息系统中, 本文提出的基于集对分析方法的扩充模型对不完整与不确定知识的处理更加合理、可行、有效。

结论 本文利用集对分析方法, 先对现有不完备信息系统的扩充模型存在的局限性进行了分析。然后, 通过引入集对联系度的概念, 给出了一个集对 α 相似限制容差关系和一个优化集对 α 相似限制容差关系, 并由此提出了两个基于集对分析方法的新的扩充粗糙集模型。这两个扩充模型既保留了限制容差关系扩充模型的优点, 又克服了现有扩充模型的局限性。它们可以通过对 α 的调节和控制, 保证容差类划分的准确性和灵活性, 并且对于一个属性值都不确定的个体, 不必与其它个体分开单独处理, 为不完备信息系统的处理提供了一个更加合理、更加可行和更加有效的方法。

参考文献

- 1 Pawlak Z, Busse J G, et al. Rough sets. Communications of the ACM, 1995, 38(11): 89~95
- 2 Kryszkiewicz M. Rough set approach to incomplete information system. Information Sciences, 1998, 112: 39~49
- 3 Stefanowski J, Tsoukias A. On the extension of rough sets under incomplete information. In: S Zhong, A Skowron, S Ohsuga, eds. Proc. of the 7th Int'l Workshop on New Directions in Rough Sets, Data Mining, and Granular Soft Computing. Berlin: Springer-Verlag, 1999, 73~81
- 4 刘清. 粗糙集及粗糙推理. 北京: 科学出版社, 2001
- 5 Skowron A. Rough sets in KDD. Presented during PAKDD 2000, Kyoto, Japan Pacific-Asia Conference on Knowledge Discovery and Data Mining
- 6 王国胤. 粗糙集理论在不完备信息系统中的扩充. 计算机研究与发展, 2002, 39(10): 1238~1243
- 7 赵克勤. 集对分析及其初步应用. 杭州: 浙江科学出版社, 2000
- 8 刘富春. 基于限制容差关系的集对粗糙集模型. 计算机科学, 2005, 32(6)
- 9 黄兵, 周献中. 基于集对分析的不完备信息系统粗糙集模型. 计算机科学, 2002, 29(7): 1~3