

粗糙集的两相似性度量

林娟¹ 米据生¹ 解滨^{1,2}

(河北师范大学数学与信息科学学院 石家庄 050024)¹ (河北师范大学信息技术学院 石家庄 050024)²

摘要 粗糙集理论是一种新的处理模糊和不确定性知识的软计算工具。在近似空间中,首先基于集合的上下近似给出了一种粗糙集间的相似度量方法。然后通过定义一种基于粗糙隶属函数的包含度,给出了另外一种粗糙集间的相似度量方法,并分别研究了这两种相似度量方法的有关性质。最后讨论了这两种相似度量方法之间的关系。

关键词 粗糙集,上下近似,包含度,相似度量

中图分类号 O236 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.6.022

Two Similarity Measures between Rough Sets

LIN Juan¹ MI Ju-sheng¹ XIE Bin^{1,2}

(College of Mathematics and Information Science, Hebei Normal University, Shijiazhuang 050024, China)¹

(College of Information Technology, Hebei Normal University, Shijiazhuang 050024, China)²

Abstract Rough set theory is emerging as a powerful tool for dealing with vagueness and uncertainty of knowledge. First of all, one method of measuring the similarity between two rough sets was proposed in approximate space based on approximation sets. And then an inclusion degree was defined based on rough membership function, which was used to define another similarity measure between two rough sets. The properties of the two similarity measures were studied respectively. Finally, the relationship between the two similarity measures was also discussed.

Keywords Rough set, Lower and upper approximation, Inclusion degree, Similarity measure

1 引言

进入21世纪以来,如何对不确定性信息和数据进行更加有效的处理,从而发现不确定性信息中蕴涵的知识和规律,是一个重要的研究课题,不确定性知识发现的各种方法也应运而生。粗糙集理论是继概率论、模糊集、证据理论之后的又一种处理不确定性的数学工具。作为一种新的软计算方法,自波兰学者Pawlak教授于1982年提出以来,其由于思想新颖、方法独特、计算简便,已成为一种重要的智能信息处理技术^[9-12],受到越来越多的重视。

粗糙集理论是建立在分类机制的基础上的,它将分类理解为在特定空间上的等价关系,而等价关系又构成对该空间的划分。知识的不确定性主要是由给定论域里粗糙近似的边界引起的,当边界为空时,知识是完全确定的。边界越大,知识就越粗糙。在实际应用中,经常需要比较两种知识模式的一致性 or 相似性。因此,给定两个不确定性集合,考虑它们的相似性是非常重要的,不同的相似度量方法也不断被提出^[2,4,5,8]。

文献[4]给出了粗糙集之间的两种相似度量,其中基于上下近似的相似度量很简单,既没有分别考虑上下近似,也没有考虑上下近似所占权重。因此,这种相似度量在应用方面就

有很大的局限性。为了解决这些问题,本文定义了另外一种基于上下近似的相似度量。由于一种包含度可以定义一种相似度,基于粗糙隶属函数,文献[5]定义了一种包含度,从而给出了粗糙集间的一种相似度量。利用此包含度,本文给出了一种新的粗糙集间的相似度量,并将其与文献[5]给出的度量进行了比较。在最后一部分讨论了本文新定义的两相似度量之间的关系。

2 基本概念

等价关系可以将对象集分类。从认知的角度来看,人们需要通过分类去认识那些不能用分类精确表示的对象集,这种集合称为粗糙集。

下面先给出有关的基本概念,详细的讨论见文献[1,7,8]。

定义1 设 U 是对象集, R 是 U 上的等价关系。

(1)称 (U, R) 为近似空间,由 (U, R) 产生的等价类为

$$U/R = \{[x_i]_R \mid x_i \in U\}$$

其中, $[x_i]_R = \{x_j \mid (x_i, x_j) \in R\}$ 。

(2)对于任意的 $X \subseteq U$,记

$$\underline{R}(X) = \{x_i \mid [x_i]_R \subseteq X\}$$

$$\overline{R}(X) = \{x_i \mid [x_i]_R \cap X \neq \emptyset\}$$

到稿日期:2014-04-18 返修日期:2014-05-29 本文受国家自然科学基金(61170107,61300153,61300121),河北省高校创新团队领军人才培养计划项目(LJRC022),河北省自然科学基金(A2014205157)资助。

林娟(1989—),女,硕士生,主要研究方向为人工智能的数学基础,E-mail:qianluo.ok@163.com;米据生(1966—),男,教授,主要研究方向为粗糙集、概念格、近似推理和随机集等,E-mail:mijsh@263.net;解滨(1976—),男,副教授,主要研究方向为粗糙集、模糊集、近似推理等,E-mail:xiabin-hebtu@126.com。

称 $\underline{R}(X)$ 为 X 的下近似, $\bar{R}(X)$ 为 X 的上近似。

(3)若 $\underline{R}(X)=\bar{R}(X)$,称 X 是可定义的集合,否则称 X 为粗糙集。

定义2 设 U 为非空有限论域, R 为 U 上的等价关系。 $\forall x \in U, X \subseteq U$,元素 x 关于等价关系 R 隶属于集合 X 的隶属度为:

$$\mu_X^R(x) = \frac{|\{x\}_R \cap X|}{|\{x\}_R|}$$

其中, $|U|$ 表示集合 U 中所含元素的个数。

定义3 设 U 是一个非空论域,记 $P(U)$ 为 U 的所有子集的集合。若 $\forall X, Y \in P(U)$,有实数 $D(Y/X)$ 与之对应,且满足:

- (1) $0 \leq D(Y/X) \leq 1$;
- (2)当 $X \subseteq Y$ 时, $D(Y/X) = 1$;
- (3)若 $X \subseteq Y \subseteq Z$,则 $D(X/Z) \leq D(X/Y)$ 。

则称 D 为 $P(U)$ 上的包含度。

3 粗糙集的相似度量

检索是专家系统中的重要问题之一。检索实际上是一种识别,是在知识库的众多模型中筛选出一种模型与给出的特征最融洽。在信息不完全、不确定的情况下,就产生了相似度的概念。

定义4 设 U 是一个非空论域,记 $P(U)$ 为 U 的所有子集的集合, $S: P(U) \times P(U) \rightarrow [0, 1]$, $\forall X, Y, Z \in P(U)$,如果 S 满足:

- (1) $0 \leq S(X, Y) \leq 1$;
- (2) $S(X, Y) = S(Y, X)$;
- (3)如果 $X = Y$,则 $S(X, Y) = 1$;
- (4)如果 $X \subseteq Y \subseteq Z$,则 $S(X, Z) \leq S(X, Y)$ 且 $S(X, Z) \leq S(Y, Z)$ 。

则称 $S(X, Y)$ 是集合 X, Y 之间的相似度。

如果将(3)换成“当且仅当 $X = Y$ 时 $S(X, Y) = 1$ ”,则称 $S(X, Y)$ 是集合 X, Y 之间的弱相似度。

在实际应用中,经常需要比较两知识模式的一致性 or 相似性。因此,相似度量是一个很重要的度量方法,它涉及数据挖掘、机器学习、自然语言和信息检索等多个领域^[3]。本文将从以下两个方面定义粗糙集间的相似性度量。

3.1 基于上下近似的粗糙集相似度量

在文献[4]中,刘文军、赵丽萍给出了一种基于集合的上下近似的相似性度量,即

$$S(X, Y) = 1 - \frac{\max\{|\underline{R}(X) - \underline{R}(Y)|, |\underline{R}(Y) - \underline{R}(X)|, |\bar{R}(X) - \bar{R}(Y)|, |\bar{R}(Y) - \bar{R}(X)|\}}{|U|}$$

可以看出这个相似度量很粗糙,既没有分别考虑上下近似,也没有考虑上下近似所占权重。为了解决这些问题,下面定义一种新的相似度量。

定义5 设 (U, R) 为近似空间, $\forall X, Y \subseteq U, \alpha, \beta \in [0, 1]$, $\alpha + \beta = 1$ 。定义

$$S_1(X, Y) = 1 - \left(\frac{\alpha \max\{|\underline{R}(X) - \underline{R}(Y)|, |\underline{R}(Y) - \underline{R}(X)|\}}{|U|} + \frac{\beta \max\{|\bar{R}(X) - \bar{R}(Y)|, |\bar{R}(Y) - \bar{R}(X)|\}}{|U|} \right)$$

下面的定理说明 $S_1(X, Y)$ 是一种相似度量。

定理1 设 (U, R) 为近似空间,则 $\forall X, Y \subseteq U, S_1(X, Y)$ 为 X 与 Y 关于等价关系 R 的相似度。

证明:显然 $S_1(X, Y)$ 满足定义4中的(1)~(3),下面证明(4)。

若 $X \subseteq Y \subseteq Z$,则 $\underline{R}(X) \subseteq \underline{R}(Y) \subseteq \underline{R}(Z), \bar{R}(X) \subseteq \bar{R}(Y) \subseteq \bar{R}(Z)$ 。进而 $\max\{|\underline{R}(X) - \underline{R}(Y)|, |\underline{R}(Y) - \underline{R}(X)|\} \leq \max\{|\underline{R}(X) - \underline{R}(Z)|, |\underline{R}(Z) - \underline{R}(X)|\}, \max\{|\bar{R}(X) - \bar{R}(Y)|, |\bar{R}(Y) - \bar{R}(X)|\} \leq \max\{|\bar{R}(X) - \bar{R}(Z)|, |\bar{R}(Z) - \bar{R}(X)|\}$ 。所以 $S_1(X, Y) \geq S_1(X, Z)$ 。同理 $S_1(X, Z) \leq S_1(Y, Z)$ 。

用 $X_R Y$ 表示 X 与 Y 关于等价关系 R 粗相等,即上下近似分别相等。基于以上定理,有以下两条性质:

性质1 设 (U, R) 为近似空间, $\forall X, Y \subseteq U$,若 $X_R Y$,则 $S_1(X, Y) = 1$ 。

证明:由 S_1 的定义易证。

性质2 设 (U, R) 为近似空间, $\forall X, Y \subseteq U$,若 X, Y 是关于 R 的精确集,则

$$S_1(X, Y) = 1 - \frac{\max\{|X - Y|, |Y - X|\}}{|U|}$$

证明:若 X, Y 是关于 R 的精确集,则 $\underline{R}(X) = \bar{R}(X) = X, \underline{R}(Y) = \bar{R}(Y) = Y$ 。于是

$$\begin{aligned} S_1(X, Y) &= 1 - \frac{\alpha \max\{|X - Y|, |Y - X|\} + \beta \max\{|X - Y|, |Y - X|\}}{|U|} \\ &= 1 - \frac{\max\{|X - Y|, |Y - X|\}(\alpha + \beta)}{|U|} \\ &= 1 - \frac{\max\{|X - Y|, |Y - X|\}}{|U|} \end{aligned}$$

下面给出一个具体的例子。

例1 设 $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}, U/R = \{\{x_1, x_3, x_6\}, \{x_2, x_4\}, \{x_5\}\}$ 。取 $X = \{x_1, x_3\}, Y = \{x_1, x_2\}$,则 $\underline{R}(X) = \underline{R}(Y) = \emptyset, \bar{R}(X) = \{x_1, x_3, x_6\}, \bar{R}(Y) = \{\{x_1, x_3, x_6\}, \{x_2, x_4\}\}$ 。所以 $S_1(X, Y) = 1 - \left(\frac{\alpha \max\{0, 0\}}{|U|} + \frac{\beta \max\{0, 2\}}{|U|} \right) = 1 - \frac{\beta}{3}$ 。

若令 $\alpha = \frac{2}{3}, \beta = \frac{1}{3}$,则 $S_1(X, Y) = \frac{8}{9}$ 。

利用文献[4]中定义的相似度,计算可得:

$$S(X, Y) = 1 - \frac{\max\{0, 0, 0, 2\}}{|U|} = \frac{2}{3}$$

注:从例1可以看出本文定义的相似度量较文献[4]中的定义更加合理,因为本文定义的相似度不仅把上下近似分开讨论,还考虑到了上下近似的权重,更具有应用价值。

3.2 基于包含度的相似度量

相似度与包含度密切相关,由一种包含度就可以定义一种相似度。包含度有各种生成方法,不同的包含度就提供了不同的不确定推理。下面绍一种基于粗糙隶属函数的包含度。

定义6 设 U 是非空有限论域, R 为 U 上的等价关系, $P(U)$ 为 U 的所有子集的集合,对于任意的 $X, Y \in P(U)$,定义

$$D_1(Y/X) = \begin{cases} 1, & X = \emptyset \\ \frac{M(X \cap Y)}{M(X)}, & X \neq \emptyset \end{cases}$$

其中, $M(X) = \sum_{i=1}^n \mu_X^R(x_i)$,它表示每个元素 $x_i (x_i \in U, i = 1, 2, \dots, n)$ 属于集合 X 的隶属度之和。

定理 2 D_1 是 $P(U)$ 上的包含度。

证明：以下假设 $X \neq \emptyset$ 。

由定义 2 知 $\mu_{X \cap Y}^R(x_i) = \frac{[x_i]_R \cap (X \cap Y)}{[x_i]_R} \leq \mu_X^R(x_i) = \frac{[x_i]_R \cap X}{[x_i]_R}$, 即 $M(X \cap Y) \leq M(X)$ 。所以 $0 \leq D_1(Y/X) \leq 1$; 当 $X \subseteq Y$ 时, $M(X \cap Y) = M(X)$, 所以有 $D_1(Y/X) = 1$; 若 $X \subseteq Y \subseteq Z$, $M(Z) \geq M(Y)$, 则有 $D_1(X/Z) \leq D_1(X/Y)$ 。即证 D_1 为包含度。

基于这个包含度就可以得到一种相似度。

定理 3 设 (U, R) 为近似空间, $\forall X, Y \subseteq U$, 则 $S_2(X, Y) = D_1(Y/X) \wedge D_2(X/Y)$ 为 X 与 Y 关于等价关系 R 的相似度。

证明：显然 $S_2(X, Y)$ 满足定义 4 中的 (1) - (3), 下面证明 (4)。

若 $X \subseteq Y \subseteq Z$, 则 $\mu_X^R(x) \leq \mu_Y^R(x) \leq \mu_Z^R(x)$, 则 $S_2(X, Y) = \frac{\sum_{i=1}^n \mu_X^R(x_i)}{M(Y)}$, $S_2(X, Z) = \frac{M(X)}{M(Z)}$, 因此 $S_2(X, Y) \geq S_2(X, Z)$ 。同理 $S_2(Y, Z) \geq S_2(X, Z)$ 。

基于以上定理, 有以下两条性质:

性质 3 设 (U, R) 为近似空间, $U/R = \{Z_1, Z_2, \dots, Z_s\}$, $\forall X, Y \subseteq U$ 。若 $\forall Z_i \in U/R (i=1, 2, \dots, s)$ 都有 $|X \cap Z_i| = |Y \cap Z_i|$, 则 $S_2(X, Y) = 1$ 。

证明：因为 $\forall Z_i \in U/R$ 有 $|X \cap Z_i| = |Y \cap Z_i|$, 所以 $|X| = |Y|$ 。又 $\forall x \in U$, 有 $x \in Z_i$, 即 $[x]_R = Z_i$, 故

$$\begin{aligned} \mu_X^R(x) &= \frac{|[x]_R \cap X|}{|[x]_R|} = \frac{|Z_i \cap X|}{|[x]_R|} \\ &= \frac{|Z_i \cap Y|}{|[x]_R|} = \frac{|[x]_R \cap Y|}{|[x]_R|} = \mu_Y^R(x) \end{aligned}$$

所以 $S_2(X, Y) = 1$ 。

性质 4 设 (U, R) 为近似空间, $X, Y \subseteq U$, 若 X, Y 是关于 R 的精确集, 则 $S_2(X, Y) = D_1(Y/X) \wedge D_1(X/Y) = \frac{|X \cap Y|}{|X|} \wedge \frac{|X \cap Y|}{|Y|}$ 。

证明：由 S_2 的定义易证。

例 2(续例 1) 计算可得粗糙隶属函数:

$$\begin{aligned} \mu_X^R &= \frac{2}{x_1} + \frac{0}{x_2} + \frac{2}{x_3} + \frac{0}{x_4} + \frac{0}{x_5} + \frac{2}{x_6} \\ \mu_Y^R &= \frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \frac{1}{x_4} + \frac{0}{x_5} + \frac{1}{x_6} \end{aligned}$$

则 X 与 Y 的相似度为:

$$\begin{aligned} S_2(X, Y) &= \frac{\frac{1}{3} + 0 + \frac{1}{3} + 0 + 0 + \frac{1}{3}}{\frac{2}{3} + \frac{2}{3} + 0 + 0 + 0 + \frac{2}{3}} \wedge \\ &\quad \frac{\frac{1}{3} + 0 + \frac{1}{3} + 0 + 0 + \frac{1}{3}}{\frac{1}{3} + \frac{1}{2} + \frac{1}{3} + \frac{1}{2} + 0 + \frac{1}{3}} \\ &= \frac{1}{2} \wedge \frac{1}{2} = \frac{1}{2} \end{aligned}$$

文献[5]中定义的相似度为:

$$S_3(X, Y) = D_1(X \cap Y / X \cup Y) = \frac{M(X \cap Y)}{M(X \cup Y)}$$

计算可得

$$\begin{aligned} S_3(X, Y) &= \frac{M(X \cap Y)}{M(X \cup Y)} \\ &= \frac{\frac{1}{3} + 0 + \frac{1}{3} + 0 + 0 + \frac{1}{3}}{\frac{2}{3} + \frac{1}{2} + \frac{2}{3} + \frac{1}{2} + 0 + \frac{2}{3}} = \frac{1}{3} \end{aligned}$$

针对例 2, 可以直观地看出 X 与 Y 的相似度为 $\frac{1}{2}$, 所以本文定义的相似度量比文献[5]中定义的相似度量要合理。

从理论上比较 $S_2(X, Y)$ 和 $S_3(X, Y)$, 易知

$$S_2(X, Y) = \frac{M(X \cap Y)}{M(X)} \wedge \frac{M(X \cap Y)}{M(Y)}$$

$$S_3(X, Y) = \frac{M(X \cap Y)}{M(X \cup Y)}$$

因为 $\mu_{X \cup Y}^R(x) = \max\{\mu_X^R(x), \mu_Y^R(x)\}$, 所以 $M(X \cup Y) \geq \{M(X), M(Y)\}$ 。因此有 $S_3(X, Y) \leq S_2(X, Y)$ 。因此对于相似度较大的两个对象, $S_2(X, Y)$ 计算比较合理。

注: 从例 1 和例 2 可以看到, 在度量两个粗糙集间的相似度时, 采用不同的度量方式将得到不同的结果, 不同的相似度量公式各有优点, 具体问题具体分析。

3.3 两种粗糙集相似度量之间的关系

在 3.1 和 3.2 节中分别从上下近似和包含度两个角度考虑了粗糙集的相似度量, 下面考虑这两种相似度量之间的关系。

定理 4 设 (U, R) 为近似空间, $\forall X, Y \subseteq U$, 若 $S_2(X, Y) = 1$, 则 $S_1(X, Y) = 1$ 。

证明：若 $S_2(X, Y) = 1$, 则 $\forall x \in U$ 必有 $\mu_X^R(x) = \mu_Y^R(x)$, 也就是 $\frac{|[x]_R \cap X|}{|[x]_R|} = \frac{|[x]_R \cap Y|}{|[x]_R|}$ 。所以 $\forall x \in U$, 有 $|[x]_R \cap X| = |[x]_R \cap Y|$ 。由上下近似定义知 $\bar{R}(X) = \bar{R}(Y)$, $\underline{R}(X) = \underline{R}(Y)$, 即 $S_1(X, Y) = 1$ 。

注: 在近似空间 (U, R) 中, 任意 $X, Y \subseteq U$, 由 $S_1(X, Y) = 1$ 不能得出 $S_2(X, Y) = 1$ 。这是因为 $\forall x \in U$, 由 X, Y 的上下近似相等一般不能得出 $|[x]_R \cap X| = |[x]_R \cap Y|$ 。例如 $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, $U/R = \{\{x_1, x_2, x_3\}, \{x_4, x_6\}, \{x_5\}\}$ 。取 $X = \{x_1, x_2\}$, $Y = \{x_3\}$, 则有 $\underline{R}(X) = \underline{R}(Y) = \emptyset$, $\bar{R}(X) = \bar{R}(Y) = \{x_1, x_2, x_3\}$ 。但是 $|[x_1]_R \cap X| = 2 \neq |[x_1]_R \cap Y| = 1$ 。

结束语 本文从上下近似和包含度两个角度出发, 定义了两种粗糙集间的相似度量, 给出了度量两个粗糙集间相似程度的不同方法。在经典集合中, 两个集合 X 与 Y 的相似度满足 $S(X, Y) = 0 \Leftrightarrow X \cap Y = \emptyset$, $S(X, Y) = 1 \Leftrightarrow X = Y$ 。但在近似空间中, 两个集合 X 与 Y 的相似度 $S(X, Y) = 0$ 不再是 $X \cap Y = \emptyset$ 的充要条件, $S(X, Y) = 1$ 也不再是 $X = Y$ 的充要条件, 这是因为近似空间中单个元素并不能视为独立的个体, 而是把每个等价类视为独立个体, 这正是粗糙集的本质。

参考文献

- [1] Pawlak Z. Rough set[J]. International Journal of Computer and Information Sciences, 1982, 11(5): 341-356
- [2] He Xiao-xu, Shao Chen-xi, Xiong Yan. A new similarity measure based on shape information for invariant with multiple Distortions[J]. Neurocomputing, 2014, 129: 556-569
- [3] Reza khani P, Jang W-S, Lee S, et al. Project risk assessment model combining the fuzzy weighted average principle with a

- similarity measure[J]. KSCE journal of civil Engineering, 2014, 18(2):521-530
- [4] 刘文军, 赵利萍. 粗糙集的相似度量[J]. 数学理论与应用, 2012, 32, (3):35-42
Liu Wen-jun, Zhao Li-ping. The similarity degree of rough set [J]. Mathematical Theory and Applications, 2012, 32(3):35-42
- [5] 史占红, 连玉平. 基于包含度的粗糙集间的相似性度量[J]. 数学教学研究, 2008(2):53-54
Shi Zhan-hong, Lian Yu-ping. Similarity measure between rough sets based on inclusion degree[J]. Mathematics teaching research, 2008(2):53-54
- [6] 张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法[M]. 北京: 科学出版社, 2001
Zhang Wen-xiu, Wu Wei-zhi, Liang Ji-ye, et al. Rough set theory and method[M]. Beijing: Science Press, 2001
- [7] 张文修, 梁怡, 徐萍. 基于包含度的不确定推理[M]. 北京: 清华大学出版社, 2007
Zhang Wen-xiu, Liang Yi, Xu Ping. Uncertainty reasoning based on inclusion degree [M]. Beijing: Tsinghua University press, 2007
- [8] 徐久成, 沈均毅, 王国胤. Rough 集之间的相似度量[J]. 计算机科学, 2003, 30(10):55-60
Xu Jiu-cheng, Shen Jun-yi, Wang Guo-yin. Measure of similarity between rough sets[J]. Computer science, 2003, 30(10):55-60
- [9] Wang G Y, Guan L H, Hu F. Rough set extensions in incomplete information systems[J]. Frontiers of Electrical and Engineering in China, 2008, 3(4):399-405
- [10] Wang G Y, Peters J F, Skowron A, et al. Rough sets and knowledge technology[C]// Proc. of the RSKT 2006. LNCS 4062, Berlin: Springer-Verlag, 2006:1-32
- [11] Wang G Y, Li T R, Grzymala B J, et al. Rough sets and knowledge technology[C]// Proc. of the RSKT 2008. LNCS 5009, Berlin: Springer-Verlag, 2008:1-18
- [12] 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
Wang Guo-yin, Yao Yi-yu, Yu Hong. A Survey on Rough Set Theory and Applications [J]. Chinese Journal of Computers, 2009, 32(7):1229-1246

(上接第 81 页)

$$\text{apr}P_4(\{a, b\}) = \{b\}, \overline{\text{apr}P_4}(\{a, b\}) = B;$$

$$\text{apr}P_5(\{a\}) = \emptyset, \overline{\text{apr}P_5}(\{a\}) = B;$$

$$\text{apr}P_5(\{b\}) = \emptyset, \overline{\text{apr}P_5}(\{b\}) = B;$$

$$\text{apr}P_5(\{c\}) = \emptyset, \overline{\text{apr}P_5}(\{c\}) = B;$$

$$\text{apr}P_5(\{a, b\}) = \emptyset, \overline{\text{apr}P_5}(\{a, b\}) = B.$$

易验证引理 4 的 3 条性质均满足。

若分划族 $\mathcal{R} = \{P_1, P_2, P_3, P_4, P_5\}$, 则 $\overline{\mathcal{R}} = P_1$ 称为由 $\mathcal{R}$ 生成的分划。则 $\mathcal{S}_1 = \{P_1\}$, $\mathcal{S}_2 = \{P_1, P_2\}$, $\mathcal{S}_3 = \{P_1, P_3\}$, $\mathcal{S}_4 = \{P_1, P_4\}$, $\mathcal{S}_5 = \{P_1, P_5\}$, $\mathcal{S}_6 = \{P_2, P_3\}$, $\mathcal{S}_7 = \{P_2, P_4\}$, $\mathcal{S}_8 = \{P_3, P_4\}$ 均为 $\mathcal{R}$ 的协调集, 其中 $\mathcal{S}_1 = \{P_1\}$, $\mathcal{S}_6 = \{P_2, P_3\}$, $\mathcal{S}_7 = \{P_2, P_4\}$, $\mathcal{S}_8 = \{P_3, P_4\}$ 均为 $\mathcal{R}$ 的约简。

而 $\mathcal{S}_9 = \{P_2, P_5\}$, $\mathcal{S}_{10} = \{P_3, P_5\}$, $\mathcal{S}_{11} = \{P_4, P_5\}$ 均不为 $\mathcal{R}$ 的协调集。

再比如系统 $(L, \mathcal{R}, \mathcal{S}_1)$, $(L, \mathcal{R}, \mathcal{S}_6)$ 和 $(L, \mathcal{S}_7, \mathcal{S}_8)$ 均不协调, 而 $(L, \mathcal{S}_1, \mathcal{S}_6)$, $(L, \mathcal{S}_1, \mathcal{S}_7)$, $(L, \mathcal{S}_1, \mathcal{S}_8)$, $(L, \mathcal{S}_6, \mathcal{S}_7)$ 均协调, $\mathcal{S}_1$ 为 $\mathcal{S}_6, \mathcal{S}_7, \mathcal{S}_8$ 的相对约简, $\mathcal{S}_6$ 不为 $\mathcal{S}_7$ 的相对约简。

结束语 粗糙集理论(RST)自从 1982 年被波兰数学家 Z. Pawlak 提出用来研究不完整数据、不精确知识的表达和学习以来, 已被广泛应用于不同的领域进行数据分析。粗糙集的数据表现方式是信息系统, 本文把经典的信息系统看作是一种特殊的格结构, 通过在格上定义分划后, 借助粗糙集思想给出上下近似算子及可辨识集, 进而分析得到约简和相对约简的判定条件。这种表示从更深层次上揭示了知识的本质, 将知识约简的已有相关结果做了进一步推广。本文的讨论还是一个尝试, 相关研究还有待进一步深入。

参考文献

- [1] Pawlak Z. Rough set[J]. International Journal of Computer and Information Sciences, 1982, 11(5):341-356
- [2] Bryniarski E. Formal description of rough sets [C]// Rough Sets, Fuzzy Sets and Knowledge Discovery. Springer-Verlag, 1994:208-216
- [3] Pawlak Z. Rough sets: Theoretical Aspects of Reasoning about Data [M]. Boston: Kluwer Academic Publishers, 1991
- [4] Ganter B, Wille R. Formal Concept Analysis: Mathematical Foundations [M]. Berlin: Springer, 1999
- [5] Kryszkiewicz M. Rules in incomplete information systems[J]. Information Sciences, 1999, 113:271-292
- [6] Yao Y Y. Concept lattices in rough set theory[C]// Dick S, Kurgan L, Pedrycz W, et al., eds. Proceedings of 2004 Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS 2004). IEEE, June 2004:796-801
- [7] 王珏, 王任, 苗夺谦, 等. 基于 Rough set 理论的“数据浓缩”[J]. 计算机学报, 1998, 21(5):393-400
Wang J, Wang R, Miao D Q, et al. Data enriching based on rough set theory [J]. Chinese Journal of Computers, 1998, 21(5):393-400
- [8] 张文修, 梁怡, 吴伟志, 等. 信息系统与知识发现[M]. 北京: 科学出版社, 2003
Zhang W X, Liang Y, Wu W Z, et al. Information system and knowledge discovery [M]. Beijing: Science Press, 2003
- [9] 胡可云, 陆玉昌, 石纯一. 概念格及其应用进展[J]. 清华大学学报: 自然科学版, 2000, 40(9):77-81
Hu K Y, Lu Y C, Shi C Y. Advances in concept lattice and its application [J]. Journal of Tsinghua University, 2000, 40(9):77-81
- [10] 张文修, 魏玲, 祁建军. 概念格的属性约简理论与方法[J]. 中国科学(E), 2005, 35(6):628-639
Zhang W X, Wei L, Qi J J. Attribute reduction theory and approach to concept lattice [J]. Science in China(E), 2005, 35(6):628-639
- [11] 刘清. 邻域值信息表上的邻域逻辑及其数据推理[J]. 计算机学报, 2001, 24(4):405-410
Liu Q. Neighborhood logic and its data reasoning on neighborhood-valued information table[J]. Chinese Journal of Computers, 2001, 24(4):405-410
- [12] 宋笑雪, 解争龙, 张文修. 集值决策信息系统的知识约简与规则提取[J]. 计算机科学, 2007, 34(4):182-184
Song X X, Xie Z L, Zhang W X. Knowledge reduction and rule extraction in set-valued decision information system [J]. Computer Science, 2007, 34(4):182-184