

基于集值的 Rough 集扩充模型

吴 鹏^{1,2}, 杨 勇¹, 张阿红¹

WU Peng^{1,2}, YANG Yong¹, ZHANG A-hong¹

1. 西北师范大学 数学与信息科学学院, 兰州 730070

2. 浙江林学院 信息工程学院, 浙江 临安 311300

1. College of Mathematics and Information Science, Northwest Normal University, Lanzhou 730070, China

2. School of Information Engineering, Zhejiang Forestry College, Lin'an, Zhejiang 311300, China

E-mail: okalpha@gmail.com

WU Peng, YANG Yong, ZHANG A-hong. Extension model based on set-valued Rough Set. Computer Engineering and Applications, 2008, 44(32): 134-136.

Abstract: Based on tolerance relation and dominance relation in literature [7], a model of k -degree limitation tolerance relation was proposed. Meanwhile the properties were investigated. Finally, a practical case study has proved the feasibility of the model, the result of classification was more suitable to the fact.

Key words: Rough Set; set-valued information system; limited tolerance relation

摘 要: 基于文献[7]中的相容关系和优势关系, 提出了一种 k 度限制相容关系模型, 同时研究其性质。最后通过实例验证该模型具有很好的灵活性, 分类结果也更加切合实际。

关键词: Rough 集; 集值信息系统; 限制相容关系

DOI: 10.3778/j.issn.1002-8331.2008.32.040 **文章编号:** 1002-8331(2008)32-0134-03 **文献标识码:** A **中图分类号:** TP18

1 引言

波兰数学家 Pawlak^[1]在 1982 年提出的 Rough 集理论是在完备信息系统范围内进行研究的, 然而, 在现实生活中, 由于数据采集中的种种原因(比如测量误差、数据丢失等等), 人们得到的信息往往是不完备的, 从而导致经典的 Rough 集模型不能精确描述现实中的不完备信息系统中的知识(或概念)。

对于不完备的信息系统, 许多人都提出了经典 Rough 集的扩充模型。目前已有的扩充模型有: Kryszkiewicz^[2]提出的基于容差关系的 Rough 集模型; Stefanowski 等^[3-4]提出的基于非对称相似关系和量化容差关系下的 Rough 集模型; 王国胤^[5]提出的基于限制容差关系下的 Rough 集模型; 须文波^[6]提出的 k 等价度容差关系 Rough 集模型等。最近, 宋笑雪等^[7-8]在不完备信息系统中提出的基于集值的相容关系和优势关系, 从集合的角度研究不完备信息系统对象的分类, 为知识的约简与规则的提取给出了一种新的途径。

通过对文献[7]中的这两种模型的分析, 发现基于集值的相容关系条件过于宽松, 使得易于将根本没有相同已知属性信息的个体误分到同一个相容类中, 而基于集值的优势关系却可能将具有很多相同已知属性信息的个体分到不同优势类中。基于

上述划分的局限性, 本文提出了一种基于集值的 k 度相容关系 Rough 集模型, 这种新模型刚好介于相容关系和优势关系这两个极端情况之间, 它可以通过选取适当的相容度 k 的值, 使得分类结果在不完备信息系统中更加符合实际。

2 预备知识

定义 1^[9] 称 (U, A, F) 是集值信息系统, 其中 U 是有限非空的对象集, A 是有限非空属性集, $F = \{f_l : l \leq m\}$ 是 U 与 A 的关系集, $f_l : U \rightarrow P_0(V_l)$ ($l \leq m$), V_l 是属性 a_l 的值域, $P_0(V_l)$ 表示 V_l 的非空子集全体。

例 1^[7] 表 1 是一个不完备信息系统, 其属性域为:

表 1 不完备信息系统(1)

Car	Price	Mailage	Size	Max-Speed
1	high	low	full	low
2	low	*	full	low
3	*	*	compact	low
4	high	*	full	high
5	*	*	full	high
6	low	high	full	*

基金项目: 国家自然科学基金(the National Natural Science Foundation of China under Grant No.10671158); 甘肃省教育厅科研基金(the Research Foundation of Department of Education of Gansu Province, China under Grant No.0701-16)。

作者简介: 吴鹏(1980-), 男, 讲师, 研究方向: 粗糙集理论; 杨勇(1967-), 男, 博士, 副教授, 研究方向: 粗糙集理论; 张阿红(1984-), 女, 硕士研究生, 研究方向: 数据挖掘, 粗糙集。

收稿日期: 2007-12-10 **修回日期:** 2008-02-26

$$V_{Price}=\{high, low\} \quad V_{Mileage}=\{high, low\}$$

$$V_{Size}=\{full, compact\} \quad V_{Max-Speed}=\{high, low\}$$

表 1 所示的不完备信息系统可以转化为表 2 所示的集值信息系统。

表 2 集值表示下的不完备信息系统(1)

Car	Price	Milage	Size	Max-Speed
1	{high}	{low}	{full}	{low}
2	{low}	{low, high}	{full}	{low}
3	{low, high}	{low, high}	{compact}	{low}
4	{high}	{low, high}	{full}	{high}
5	{low, high}	{low, high}	{full}	{high}
6	{low}	{high}	{full}	{low, high}

在集值信息系统中,宋笑雪等给出了相容关系和优势关系,其定义如下:

定义 2^[7] 设 (U, A, F) 是一个集值信息系统,任意属性子集 $B \subseteq A$,定义二元关系: $R_B^{\cap} = \{(x, y) \in U \times U \mid f_a(x) \cap f_a(y) \neq \emptyset (\forall a \in B)\}$,并记 $[x]_B^{\cap} = \{y \in U \mid (x, y) \in R_B^{\cap}\} = \{y \in U \mid f_a(x) \cap f_a(y) \neq \emptyset (\forall a \in B)\}$ 。

显然 R_B^{\cap} 是自反的和对称的,但未必是等价关系, $[x]_B^{\cap}$ 称之为相容关系,称为相容类。

对任意属性子集 $B \subseteq A$,集合 $X \subseteq U$ 在相容关系 R_B^{\cap} 下关于属性集 B 的下近似集 $\underline{R}_B^{\cap}(X)$ 和上近似集 $\overline{R}_B^{\cap}(X)$ 是指:

$$\underline{R}_B^{\cap}(X) = \{x \in U \mid [x]_B^{\cap} \subseteq X\}, \overline{R}_B^{\cap}(X) = \{x \in U \mid [x]_B^{\cap} \cap X \neq \emptyset\}$$

定义 3^[7] 设 (U, A, F) 是一个集值信息系统,任意属性子集 $B \subseteq A$,定义二元关系: $R_B^{\supseteq} = \{(x, y) \in U \times U \mid f_a(x) \supseteq f_a(y) (\forall a \in B)\}$,并记 $[x]_B^{\supseteq} = \{y \in U \mid (x, y) \in R_B^{\supseteq}\} = \{y \in U \mid f_a(x) \supseteq f_a(y) (\forall a \in B)\}$ 。

显然 R_B^{\supseteq} 是自反的和传递的,但未必是等价关系, $[x]_B^{\supseteq}$ 称之为优势关系,称为优势类。

对任意属性子集 $B \subseteq A$,集合 $X \subseteq U$ 在相容关系 R_B^{\supseteq} 下关于属性集 B 的下近似集 $\underline{R}_B^{\supseteq}(X)$ 和上近似集 $\overline{R}_B^{\supseteq}(X)$ 是指:

$$\underline{R}_B^{\supseteq}(X) = \{x \in U \mid [x]_B^{\supseteq} \subseteq X\}, \overline{R}_B^{\supseteq}(X) = \{x \in U \mid [x]_B^{\supseteq} \cap X \neq \emptyset\}$$

3 限制相容关系

在相容关系下,集值信息系统中只要两个属性集值相交非空,就可以认为这两个对象包含在同一个等价类中,这种分类方式过于宽松;而在优势关系下,只有当一个属性集值是另一个属性集值的子集,两个对象才能够划分在同一个等价类中,这种分类方式又过于严格,鉴于这种情况,本文研究并提出了一种 k 度限制相容关系。

设 (U, A, F) 是一个集值信息系统,令 $k = \frac{|f_a(x) \cap f_a(y)|}{|f_a(y)|}$,显然, $0 \leq k \leq 1$,当 $|f_a(y)|=0$ 时,规定 $k=1$ 。

定义 4 限制容差关系定义为: $0 < k \leq 1, R_{kB}^{\cap} = \{(x, y) \in U \times U \mid \frac{|f_a(x) \cap f_a(y)|}{|f_a(y)|} \geq k (\forall a \in B)\}$ 。

显然,当 $k=0$ 时,此关系为一般相容关系, $k=1$ 时,此关系就为优势关系,因此限制相容关系是相容关系和优势关系的推

广。另外, R_{kB}^{\cap} 具有自反性和对称性,不具有传递性,所以不是等价关系。

设 (U, A, F) 是一个集值信息系统,任意属性子集 $B \subseteq A, 0 < k \leq 1$,称 $[x]_{kB}^{\cap} = \{y \in U \mid (x, y) \in R_{kB}^{\cap}\}$ 为 k 度相容类,对任意的 $X \subseteq U$,分别称

$$\underline{R}_{kB}^{\cap}(X) = \{x \in U \mid [x]_{kB}^{\cap} \subseteq X\}, \overline{R}_{kB}^{\cap}(X) = \{x \in U \mid [x]_{kB}^{\cap} \cap X \neq \emptyset\}$$

为 X 的下近似和上近似。

定理 1 设 (U, A, F) 为集值信息系统,对任意的 $X \subseteq U, 0 < k \leq 1$ 有: $\underline{R}_B^{\cap}(X) \subseteq \underline{R}_{kB}^{\cap}(X), \overline{R}_B^{\cap}(X) \subseteq \overline{R}_{kB}^{\cap}(X)$ 。

定理 2 设 (U, A, F) 为集值信息系统,对任意的 $X, Y \subseteq U, 0 < k \leq 1$ 则有

- (1) $\underline{R}_{kB}^{\cap} X \subseteq X \subseteq \overline{R}_{kB}^{\cap} X$;
- (2) $\underline{R}_{kB}^{\cap} \emptyset = \overline{R}_{kB}^{\cap} \emptyset = \emptyset$;
- (3) $\underline{R}_{kB}^{\cap}(X \cap Y) = \underline{R}_{kB}^{\cap} X \cap \underline{R}_{kB}^{\cap} Y$;
- (4) $\overline{R}_{kB}^{\cap}(X \cup Y) = \overline{R}_{kB}^{\cap} X \cup \overline{R}_{kB}^{\cap} Y$;
- (5) $X \subseteq Y \Rightarrow \underline{R}_{kB}^{\cap} X \subseteq \underline{R}_{kB}^{\cap} Y$;
- (6) $X \subseteq Y \Rightarrow \overline{R}_{kB}^{\cap} X \subseteq \overline{R}_{kB}^{\cap} Y$ 。

在 k 度相容关系下分类时,通过相容度 k 可以控制两个属性集值之间相同属性在整个属性集中的比重,可以根据实际情况选取适当的 k 值来获得好的分类效果。 k 取值越大,分类条件就越严格。对于不确定元素较多的集值信息系统, k 值可适当调小,反之,对于不确定元素较少的集值信息系统, k 值可适当调大。针对具体的集值信息系统,可以灵活地调整相容度,故本文提出的相容度为 k 的集值 Rough 集模型更切合实际。

4 性能分析

为了验证 k 度相容关系的分类效果,构造一个不完备信息系统,如表 3 所示。该表对应的用集值表示的不完备信息系统如表 4 所示。

表 3 不完备信息系统(2)

A	c ₁	c ₂	c ₃	c ₄
a ₁	1	*	3	*
a ₂	*	1	3	4
a ₃	2	2	1	*
a ₄	1	*	3	4
a ₅	*	2	*	1
a ₆	*	1	2	3
a ₇	2	1	*	3
a ₈	1	*	*	2
a ₉	1	2	1	*
a ₁₀	*	*	1	2
a ₁₁	2	1	*	*
a ₁₂	*	2	1	*

表 4 集值表示下的不完备信息系统(2)

A	c ₁	c ₂	c ₃	c ₄
a ₁	{1}	{1,2}	{3}	{1,2,3,4}
a ₂	{1,2}	{1}	{3}	{4}
a ₃	{2}	{2}	{1}	{1,2,3,4}
a ₄	{1}	{1,2}	{3}	{4}
a ₅	{1,2}	{2}	{1,2,3}	{1}
a ₆	{1,2}	{1}	{2}	{3}
a ₇	{2}	{1}	{1,2,3}	{3}
a ₈	{1}	{1,2}	{1,2,3}	{2}
a ₉	{1}	{2}	{1}	{1,2,3,4}
a ₁₀	{1,2}	{1,2}	{1}	{2}
a ₁₁	{2}	{1}	{1,2,3}	{1,2,3,4}
a ₁₂	{1,2}	{2}	{1}	{1,2,3,4}

根据表 4 所示的集值信息系统,相容关系的分析结果如下:

$$[a_1]_4^{\cap} = \{1, 2, 4, 5, 8\}, [a_2]_4^{\cap} = \{1, 2, 4, 11\}, [a_3]_4^{\cap} = \{3, 5, 10, 12\},$$

$$[a_4]_4^{\cap} = \{1, 2, 3\}, [a_5]_4^{\cap} = \{1, 3, 5, 9, 12\}, [a_6]_4^{\cap} = \{6, 7, 11\},$$

$$[a_7]_4^{\cap} = \{6, 7, 11\}, [a_8]_4^{\cap} = \{1, 8, 9, 10, 12\}, [a_9]_4^{\cap} = \{5, 8, 9, 10, 12\},$$

$$[a_{10}]_4^{\cap} = \{3, 8, 9, 10, 11, 12\}, [a_{11}]_4^{\cap} = \{2, 6, 7, 10, 11\},$$

$$[a_{12}]_k^{\cap} = \{3, 5, 8, 9, 10, 12\}$$

利用优势关系分析结果如下:

$$[a_1]_k^{\supset} = \{1, 4\}, [a_2]_k^{\supset} = \{2\}, [a_3]_k^{\supset} = \{3\}, [a_4]_k^{\supset} = \{4\}, [a_5]_k^{\supset} = \{5\},$$

$$[a_6]_k^{\supset} = \{6\}, [a_7]_k^{\supset} = \{7\}, [a_8]_k^{\supset} = \{8\}, [a_9]_k^{\supset} = \{9\}, [a_{10}]_k^{\supset} = \{10\},$$

$$[a_{11}]_k^{\supset} = \{7, 11\}, [a_{12}]_k^{\supset} = \{3, 8, 12\}$$

取相容度 $k=0.3$ 的限制相容关系分析结果如下:

$$[a_1]_{kA}^{\cap} = \{1, 2, 4, 5, 8\}, [a_2]_{kA}^{\cap} = \{2, 4\}, [a_3]_{kA}^{\cap} = \{3, 5, 10, 12\},$$

$$[a_4]_{kA}^{\cap} = \{2, 4\}, [a_5]_{kA}^{\cap} = \{5\}, [a_6]_{kA}^{\cap} = \{6, 7\}, [a_7]_{kA}^{\cap} = \{7\}, [a_8]_{kA}^{\cap} = \{8, 10\},$$

$$[a_9]_{kA}^{\cap} = \{5, 8, 9, 10, 12\}, [a_{10}]_{kA}^{\cap} = \{3, 8, 9, 10, 12\},$$

$$[a_{11}]_{kA}^{\cap} = \{2, 6, 7, 11\}, [a_{12}]_{kA}^{\cap} = \{3, 5, 8, 9, 10, 12\}$$

相容度 $k=0.5$ 的限制相容关系分析结果如下:

$$[a_1]_{kA}^{\cap} = \{1, 4, 8\}, [a_2]_{kA}^{\cap} = \{2, 4\}, [a_3]_{kA}^{\cap} = \{3, 10, 12\}, [a_4]_{kA}^{\cap} = \{2, 4\},$$

$$[a_5]_{kA}^{\cap} = \{5\}, [a_6]_{kA}^{\cap} = \{6\}, [a_7]_{kA}^{\cap} = \{7\}, [a_8]_{kA}^{\cap} = \{8, 10\},$$

$$[a_9]_{kA}^{\cap} = \{9, 10, 12\}, [a_{10}]_{kA}^{\cap} = \{3, 9, 10, 12\}, [a_{11}]_{kA}^{\cap} = \{2, 6, 7, 11\},$$

$$[a_{12}]_{kA}^{\cap} = \{3, 9, 10, 12\}$$

由此可以看出,本文模型不但避免了优势关系对集值信息系统要求严格的缺陷,而且还改进了相容关系的不足,比如对象 a_{10} 和 a_{11} ,没有一个集值属性相同,而在相容关系下,却认为二者是不可分辨的,在本文模型取适当的相容度($k=0.3$ 或 $k=0.5$),可以将二者分辨开来;而对象 a_{11} 和 a_7 ,有一半集值属性相同,在优势关系下却认为二者是可分辨的,而采用本文模型进行分析,二者是不可分辨的,可以划归为同一个等价类。

(上接 108 页)

的地址是第 8 步中应用服务器的应用逻辑转换之后的一个字符串,将它直接放入 Request-URI 的头中即可。

5 结束语

在 Parlay/OSA MPCC SCF 向 SIP 协议的映射中,两者的呼叫模型以及层次都有所不同,映射中需要从概念、方法和参数等几个方面进行考虑。在本文中采用分层的体系结构,构建了 SIP 协议映射子系统的实现模型。在该模型中,每一层都构建在为下一层提供服务的基础上,并在层与层之间定义互操作接口,使得该系统易于升级和移植。本文最后所举实例比较简单,还有很多复杂的 Parlay API 并未涉及,有待于日后的改进和补充。值得提出的是:本文中,SIP 协议映射子系统中 SIP 服务器模式是根据应用的实际功能来设定的。为了进一步实现对底层网络协议细节的屏蔽,如何让 SCF 根据应用需求自动地转换 SIP 服务器的模式,将成为以后工作和研究的重点。

参考文献:

- [1] 3GPP TS 29.198-1 Open Service Access (OSA) Application Programming Interface(API) Part 1:Overview(Release 6)[S].2005.
- [2] Rosenberg J,Schulzrinne H,Camarillo G,et al.RFC 3261 SIP:Session

5 结束语

将不完备信息系统转化为集值信息系统,为处理不完备信息系统提供了另外一种新的思想,本文在对集值信息系统原有两种关系(相容关系和优势关系)分析理解的基础上,提出了一种改进的相容关系的 Rough 集模型,即 k 度限制相容关系模型,这种模型吸取了原有的两种关系模型的优点,使分类结果更加切合不完备信息系统的实际。

参考文献:

- [1] Pawlak Z.Rough sets[J].International Journal of Computer and Information Sciences,1982,11:341-356.
- [2] Kryszkiewicz M.Rough set approach to incomplete information systems[J].Information Sciences,1998,112(1/4):39-49.
- [3] Stefanowski J,Tsoukiàs A.On the extension of rough sets under incomplete information[C]/Zhong N,Skowron A,Ohsuga S.Lecture Notes in Computer Science:Proceedings of the 7th International Workshop on New Directions in Rough Sets,Data Mining,and Granular-Soft Computing,Yamaguchi,Japan,November 9-11,1999.Berlin:Springer-Verlag,1999,1711:73-81.
- [4] Stefanowski J,Tsoukiàs A.Incomplete information tables and rough classification[J].Computational Intelligence,2001,17(3):546-564.
- [5] 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.
- [6] 须文波,李翔飞.不完备信息系统的 Rough 模型扩充[J].计算机工程与应用,2007,43(5):169-171.
- [7] 宋笑雪,李鸿儒,张文修.集值决策信息系统的知识约简与属性特征[J].计算机科学,2006,33(7):179-181.
- [8] 宋笑雪,张文修.基于集值决策属性的集值信息系统[J].计算机工程与应用,2007,43(17):8-10.
- [9] 张文修,梁怡,吴伟志.信息系统与知识发现[M].北京:科学出版社,2001.
- [10] Initiation Protocol[S].June 2002.
- [3] 3GPP TR 29.998-04-4 API Mapping for OSA Part 4:Call control service mapping subpart 4:Multiparty call control ISC(Release 6)[S].2004.
- [4] 3GPP TS 29.198-4-4 Open Service Access(OSA) Application Programming Interface(API) Part4:Call control sub-part 3:Multi-party call control Service Capability Feature(SCF)(Release 6)[S].2005.
- [5] Glitho R H,Poulin A.A high level service creation environment forparlay in a SIP environment[C]//IEEE International Conference on Volume 4,28 April-2 May 2002:2008-2013.
- [6] 3GPP TS 29.198-2 Open Service Access (OSA) Application Programming Interface(API) Part 2:Common data definitions(Release 6)[S].2005.
- [7] 万晓榆,姚平香,张洪,等.下一代网络的业务生成技术[M].北京:北京邮电大学出版社,2005.
- [8] 樊自甫,万晓榆.NGN 基于 Parlay API 业务生成接口的研究与设计[J].计算机工程与应用,2006,42(3):127-130.
- [9] 樊自甫,万晓榆.Parlay API 中 GCC SCF 向 SIP 协议的映射研究[J].计算机工程与应用,2007,43(17):133-136.
- [10] 管有庆.Parlay CC API 与 SIP 的逻辑[J].南京邮电学院学报,2004,24(2).
- [11] 吴德强,管有庆,沈苏彬.Parlay 的呼叫控制服务在 SIP 环境的映射[J].南京邮电学院学报,2004,24(2).