

DOI:10.11992/tis.201603055
网络出版地址: <http://www.cnki.net/kcms/detail/23.1538.TP.20160513.0920.018.html>

一种基于概念格的集值信息系统中的知识获取方法

康向平^{1,2}, 苗夺谦^{1,2}
(1. 同济大学 计算机科学与技术系, 上海 201804; 2. 同济大学 嵌入式系统与服务计算教育部重点实验室, 上海 201804)

摘 要:以集值信息系统为研究背景,以概念格为理论基础,提出了一种基于概念格的集值信息系统中的知识获取方法。该模型首先将复杂的集值信息系统转化为形式上更加简单的单值背景,然后借助概念格理论,重点探讨了基于相容关系的粒化模型和集值系统中的格代数结构,该代数结构可以将论域中的所有覆盖以格的形式有机组织起来。此外,探讨了集值信息系统中的约简、核等问题。本文有助于拓展概念格的应用范围,同时也为集值信息系统的分析和处理提供了一种有益思路。
关键词:粗糙集;概念格;集值信息系统;相容关系;代数结构
中图分类号:TP18 **文献标志码:**A **文章编号:**1673-4785(2016)03-0287-07

中文引用格式:康向平, 苗夺谦.一种基于概念格的集值信息系统中的知识获取方法[J]. 智能系统学报, 2016, 11(3): 287-293.
英文引用格式:KANG Xiangping, MIAO Duoqian. A knowledge acquisition method based on concept lattice in set-valued information systems [J]. CAAI transactions on intelligent systems, 2016,11(3): 287-293.

A knowledge acquisition method based on concept lattice in set-valued information systems

KANG Xiangping^{1,2}, MIAO Duoqian^{1,2}
(1. Department of Computer Science and Technology, Tongji University, Shanghai 201804, China; 2. Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 201804, China)

Abstract: The paper takes set-valued information systems as research background, proposes a knowledge acquisition method on the basis of concept lattice. The model can transforms a complicated set-valued information system into a simpler one-valued context, and then by means of concept lattice, emphasizes the granularity model based on tolerance relation and the algebra structure in the set-valued information system, where the algebraic structure can organize all covers in the form of lattice structure, and it can be considered an important algebraic structure in the set-valued information system. Meanwhile, the paper also offers some simple solutions to common problems, such as reduction, core. In short, this paper not only helps to explore the application range of concept lattice, but also offers a useful idea for the analysis and processing of set-valued information systems.
Keywords: rough set; concept lattice; set-valued information systems; tolerance relations; algebraic structure

针对模糊、不确定等问题,波兰学者 Pawlak 基于粒化和近似思想于 1982 年提出了粗糙集^[1],该理论于 20 世纪 90 年代逐步走向成熟并随后引起了国内外学者的广泛关注^[2]。与其他类似理论相比较,其不需要借助复杂的先验知识便可以在数据集中有

效获取知识,相对来讲会更加客观和易用。近年来,粗糙集正快速从单一理论向多理论融合方向发展、从简单应用向复杂数据建模方向延伸,其相关研究成果已被广泛应用于数据挖掘、机器学习等领域^[3-5]。

概念格是与粗糙集同时代产生的一种数据分析工具^[6],其最早是由德国学者 Wille 基于人脑的概念思维提出来的。在概念格构筑的理论体系中,概念、格代数结构,伽罗瓦(Galois)连接等是核心要

收稿日期:2016-03-28. 网络出版日期:2016-05-13.
基金项目:国家自然科学基金项目(61273304, 61202170);国家博士后科学基金项目(2014M560352);高等学校博士学科点专项科研基金项目(20130072130004).
通信作者:康向平. E-mail:tongji_kangxp@sina.com.

素。其中,概念可以分别从内涵和外延两个不同角度进行刻画,具有丰富的语义,有助于深化人们对现实生活中概念的概要认识和准确理解;代数格结构作为一种基于序关系建立起来的概念层次模型,客观上反映了概念之间的泛化/特化关系;伽罗瓦连接则为生成概念和代数结构的核心理论基础。近年来,关于概念格的基础理论研究和应用拓展已经取得了一系列重要成果^[7-10]。

虽然上述理论方法上存在一定差异,但集合论和二元关系都是它们的建模基础,同时它们均面向相类似的数据集,这就意味着它们之间可能存在着某种紧密的关联。在充分挖掘它们各自理论优势的基础上,探讨二者之间的融合理论必将有助于粗糙集和概念格的进一步拓展。近年来,许多学者探讨了它们之间的融合理论^[11]。一些学者把概念格中的概念思维、格代数结构、伽罗瓦连接等要素引入到粗糙集中,探讨了信息系统中的格代数结构、约简、规则获取等^[12-13];一些学者把形式背景视为一类特殊的信息系统,然后借助粗糙集中的近似算子和区分矩阵,提出了概念格中的粗糙概念分析、属性约简,以及上近似格和下近似格等^[14-16];一些学者从不同角度对比了两种理论,揭示了它们之间的紧密联系^[17-20],这样做不仅有助于人们从不同角度深化对单一理论的认识和理解,同时也有助于两种理论之间的深度融合。

在关于经典信息系统的描述中,对象在属性下的取值是唯一的。然而,在一些复杂的信息系统,对象在属性下的取值有时可能不是一个值,而是由若干个元素组成的一个非空集合,即取值不是一个值而是一个集合。当取值是一个集合时,若其代表一个取值范围,那么此时取值就存在一定的不确定性。例如,在地震预测中,不同专家可能得出不一样的预测结果,也许这些预测结果都是合理的,但却只有一种预测是正确的。据此,人们将经典信息系统进一步泛化为集值信息系统。本质上,信息的不完备性也可以理解为上述不确定性的一种特殊情况^[21-22]。例如,在不完备信息系统中,针对取值缺失的情况,人们通常会用一个集合替代某个缺失值,即依据先验知识或算法将缺失值限定在某个范围,并将该范围视为缺失值。在此意义下,不完备信息系统本质上是一种特殊的集值信息系统。针对上述不确定性,以往人们更多采用的是一些间接的知识获取方法,即数据建模是建立在对数据预处理的基础之上。此类方法虽然简单易用,但数据预处理可能会丢失大量的有用信息或由于人为因素增加一些额外的偏

差信息,存在一定的局限性。针对上述情况,许多学者尝试发展直接面向集值信息系统的知识获取方法^[23-28],诸如基于相容关系、优势关系等的粗糙集模型。直接处理方法不经过数据预处理,能直接对原始数据进行处理,可以有效避免由于数据预处理不当所导致的偏差逐层传导扩大。

针对上述不确定性问题,同时考虑到概念格在二元关系分析中的良好数学基础,本文引入了概念格,并重点探讨了基于相容关系的粒化模型和集值系统中的格代数结构,该代数结构可以将论域中的所有覆盖以格的形式有机组织起来。此外,本文还重点探讨了集值信息系统中的约简、核等问题。总的来讲,本文结论一方面有助于拓展概念格的应用范围,另一方面也为集值信息系统的分析和处理提供了一种有益思路。

1 概念格

定义 1 一个形式背景通常由一个三元组 (G,M,I) 来表示,其中 G 和 M 分别是非空有限的对象集和属性集, $I \subseteq G \times M$ 。一个典型的形式背景如表 1 所示,其中 website1, website2, ..., website6 分别表示一些网站的编号; Finance, Culture, ..., Shopping 则表示上述网站可能涉及的一些主题;若某个网站含有某种主题,用“×”进行标记。

表 1 一个典型的形式背景
Table 1 A typical formal context

	Finance	Culture	Military	History	Shopping
website 1	×	×	×		
website 2	×	×	×	×	
website 3	×	×	×	×	
website 4		×	×	×	×
website 5		×	×	×	×
website 6			×	×	×

定义 2 定义 1 在形式背景 (G,M,I) 中,对于任意 $X \in P(G)$, 定义

$$X^* = \{m \in M \mid (x,m) \in I, \forall x \in X\}$$

相应地,对于任意 $B \in P(M)$, 定义

$$B^* = \{x \in G \mid (x,m) \in I, \forall m \in B\}$$

其中 $P(G)$ 和 $P(M)$ 分别是 G 和 M 的幂集。如果 $X^* = B$ 且 $B^* = X$, 则称 (X,B) 是一个形式概念,其中 X 和 B 分别是外延和内涵。若将任意概念 (X_1,B_1) 和 (X_2,B_2) 之间的序关系“ \leq ”定义为

$$(X_1,B_1) \leq (X_2,B_2) \Leftrightarrow X_1 \subseteq X_2 \Leftrightarrow B_1 \supseteq B_2$$

在此意义下,偏序集 $(B(G,M,I), \leq)$ 本质上是

一个完备格,其中 $B(G,M,I)$ 是所有概念组的集合。

引理 1 在形式背景 (G,M,I) 中,对于任意 $B \subseteq M$,有

$$B^{**} = \cap \{L \in \text{Int}(G,M,I) \mid B \subseteq L\}$$

式中 $\text{Int}(G,M,I)$ 表示所有概念内涵组成的集合。

定义 3 在表 1 所示的背景中,若 $a_1 = \text{Finance}$ 、 $a_2 = \text{Culture}$ 、 $a_3 = \text{Military}$ 、 $a_4 = \text{History}$ 、 $a_5 = \text{Shopping}$, 且将 website1, website2, ..., website6 简记为 1, 2, ..., 6,则表 1 中的概念格如图 1 所示,图 1 可以直观反映信息在不同层次上的聚类。例如,概念 $(456,a_3a_4a_5)$ 揭示 website4、website5 和 website6 等网站均含有主题 Military、History 和 Shopping。

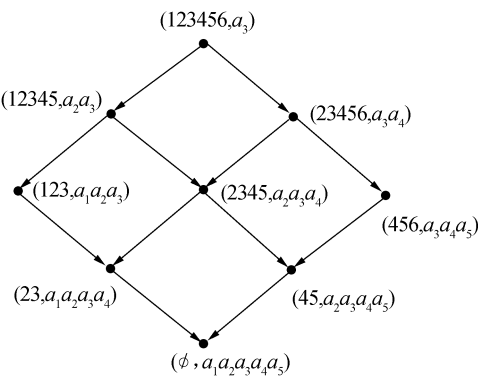


图 1 从表 1 导出的概念格结构

Fig.1 The concept lattice derived from table 1

定义 4 关于概念格的详细介绍请参阅文献 [29]。

2 集值信息系统和单值形式背景

一个集值信息系统一般由四元组 (U,AT,V,F) 进行表示。其中 U 是一个非空对象集,称为论域; AT 是一个非空属性集; $V = \cup_{m \in AT} V_m$, 其中 V_m 是属性 m 的值域; $F = \{f_m:m \in AT\}$, 其中 $f_m:U \rightarrow P(V_m)$ 。一个典型的集值信息系统如表 2。

表 2 一个典型的集值信息系统

Table 2 A typical set-valued information system					
	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>
<i>u</i> ₁	<i>v</i> ₁	<i>v</i> ₁	<i>v</i> ₁	<i>v</i> ₁	<i>v</i> ₁ , <i>v</i> ₃
<i>u</i> ₂	<i>v</i> ₁	<i>v</i> ₁	<i>v</i> ₁	<i>v</i> ₁	<i>v</i> ₁ , <i>v</i> ₃
<i>u</i> ₃	<i>v</i> ₁ , <i>v</i> ₂	<i>v</i> ₁	<i>v</i> ₁ , <i>v</i> ₂	<i>v</i> ₂	<i>v</i> ₂
<i>u</i> ₄	<i>v</i> ₁ , <i>v</i> ₂	<i>v</i> ₁	<i>v</i> ₁ , <i>v</i> ₂	<i>v</i> ₂	<i>v</i> ₂
<i>u</i> ₅	<i>v</i> ₂ , <i>v</i> ₃	<i>v</i> ₁ , <i>v</i> ₂	<i>v</i> ₂	<i>v</i> ₁	<i>v</i> ₁ , <i>v</i> ₂
<i>u</i> ₆	<i>v</i> ₂ , <i>v</i> ₃	<i>v</i> ₁ , <i>v</i> ₂	<i>v</i> ₂	<i>v</i> ₁	<i>v</i> ₁ , <i>v</i> ₂
<i>u</i> ₇	<i>v</i> ₃	<i>v</i> ₂	<i>v</i> ₂	<i>v</i> ₁ , <i>v</i> ₂	<i>v</i> ₃
<i>u</i> ₈	<i>v</i> ₃	<i>v</i> ₂	<i>v</i> ₂	<i>v</i> ₁ , <i>v</i> ₂	<i>v</i> ₃

单值背景中的经典算子往往并不适用于多值的信息系统。即若采用定义 1 中的算子,必须将信息系统转化为单值形式背景。由此,针对集值信息系统,本文采用了如下转化策略。

定义 3 设 (U,AT,V,F) 是一个集值信息系统,称 $(U \times U,AT,J)$ 是 (U,AT,V,F) 的衍生形式背景,其中 J 被描述为:对于任意 $(x,y) \in U^2,m \in AT$,有

$$((x,y),m) \in J \Leftrightarrow f(x,m) \cap f(y,m) \neq \emptyset$$

由于衍生背景不再保留原信息系统中属性值自身的信息,仅体现相同属性下不同对象之间的关系,因而在形式上较原信息系统更加简单明了。一个由表 2 所示集值信息系统衍生的形式背景如表 3。

表 3 一个衍生形式背景

Table 3 A derivative formal context					
	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>
(u_1, u_1)	×	×	×	×	×
(u_1, u_2)	×	×	×	×	×
(u_1, u_3)	×	×	×		
(u_1, u_4)	×	×	×		
(u_1, u_5)		×		×	×
(u_1, u_6)		×		×	×
⋮	⋮	⋮	⋮	⋮	⋮
(u_8, u_3)			×	×	
(u_8, u_4)			×	×	
(u_8, u_5)	×	×	×	×	
(u_8, u_6)	×	×	×	×	
(u_8, u_7)	×	×	×	×	×
(u_8, u_8)	×	×	×	×	×

3 集值信息系统中的代数结构和粒化模型

概念格是一类重要的格代数结构,其在形式概念分析理论中扮演了重要的角色。据此,本文尝试基于形式概念分析理论重点探讨集值信息系统中的格代数结构、信息粒化等问题。其中,本文提到的粒化模型是基于定义 3 中的相容关系构造的。

定义 4^[23] 设 (U,AT,V,F) 是一个集值信息系统,对于任意 $B \subseteq AT$,论域 U 中的相容关系定义为

$$R_B = \{(x,y) \in U^2 \mid \forall m \in B, f(x,m) \cap f(y,m) \neq \emptyset\}$$

定理 1 设 $(U \times U,AT,J)$ 是 (U,AT,V,F) 的衍生背景,对于任意 $B \subseteq AT$,有 $B^* = R_B$ 。

证明 由定义 1、定义 2 和定义 3 即得,证毕。

定义 5 设 R 是论域 U 中的一个相容关系,

$H \subseteq U$ 。若 $H \times H \subseteq R$, 且不存在 $N \times N \subseteq R$ 满足 $H \times H \subset N \times N$, 称 H 是一个极大相容类。 U 中所有极大相容类组成的集合记为 U/R 。 设 R_1 和 R_2 是 U 中的相容关系, 对于任意 $X \in U/R_1$, 若存在 $Y \in U/R_2$ 且满足 $X \subseteq Y$, 则称 U/R_2 相对于 U/R_1 是一种更粗的粒化结果, 记为 $U/R_1 \subseteq U/R_2$ 。

引理 2 在定义 4 中, 有下述结论成立:

1) 若 R_1 和 R_2 是论域 U 中的相容关系, 则 $R_1 \subseteq R_2 \Leftrightarrow U/R_1 \subseteq U/R_2$;

2) 在 (U, AT, V, F) 中, 设 $B, D \subseteq AT$, 则 $B \subseteq D \Rightarrow R_D \subseteq R_B$ 。

定义 6 若 R 是论域 U 中的一个相容关系, 则称 (U, U, R) 是由 R 决定的中间背景。

定理 2 在上述背景 (U, U, R) 中, 有下述结论成立:

$$R = \cup \{X \times X \mid X \in U/R\}$$

$$X \in U/R \Leftrightarrow (X, X) \in B(U, U, R)$$

证明 由定义 1 和定义 4 即得, 证毕。

在表 2 中, 表 4 是由 $R_{\{d,e\}}$ 决定的中间背景, 相应的概念格如图 2 所示。显然, 依据定理 2, 可以判定 $\{u_1, u_2, u_5, u_6\}, \{u_1, u_2, u_7, u_8\}, \{u_3, u_4\}$ 均是 U/R_B 中的极大相容类。在图 2 中, “ u_i ” 简化表示为 “ i ”。

表 4 一个中间背景

Table 4 A intermediate context

	u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8
u_1	×	×			×	×	×	×
u_2	×	×			×	×	×	×
u_3			×	×				
u_4			×	×				
u_5	×	×			×	×		
u_6	×	×			×	×		
u_7	×	×					×	×
u_8	×	×					×	×

依据定理 2, 本文在集值信息系统中进一步定义如下 α 和 β 映射。

定义 7 设 $\nu(R)$ 和 $\nu(L)$ 分别是由论域 U 中所有相容关系和覆盖组成的集合。映射 $\alpha: \nu(R) \rightarrow \nu(L)$ 定义为

$$\alpha(R) = \{X \mid (X, X) \in B(U, U, R)\}$$

映射 $\beta: \nu(L) \rightarrow \nu(R)$ 定义为

$$\beta(C) = \cup \{X \times X \mid X \in C\}$$

式中 $R \in \nu(R)$ 和 $C \in \nu(L)$ 。

定义 8 设 $(U \times U, AT, J)$ 是 (U, AT, V, F) 的衍

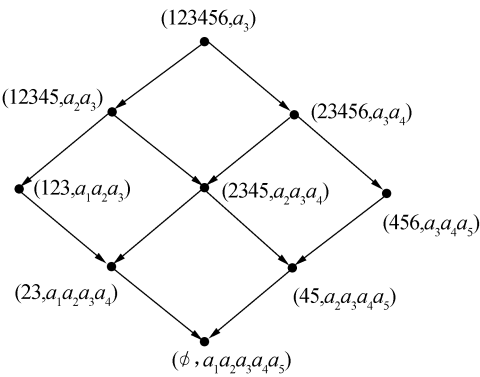


图 2 从表 4 导出的概念格结构

Fig. 2 A concept lattice derived from table 4

生背景, 映射 $+: \nu(L) \rightarrow P(AT)$ 被描述为

$$C^+ = (\beta(C))^* = \{m \in AT \mid ((x, y), m) \in J, \forall (x, y) \in \beta(C)\}$$

映射 $+: P(AT) \rightarrow \nu(L)$ 被描述为

$$B^+ = \alpha(B^*) = \{X \mid (X, X) \in B(U, U, B^*)\}$$

式中 $B \subseteq AT, C \in \nu(L)$ 。若 $C^+ = B$ 且 $B^+ = C$, 则称 (C, B) 是 (U, AT, V, F) 中的相容概念。对于任意相容概念 (C_1, B_1) 和 (C_2, B_2) , 它们之间的偏序关系 “ \leq ” 定义为

$$(C_1, B_1) \leq (C_2, B_2) \Leftrightarrow C_1 \subseteq C_2 \Leftrightarrow B_2 \subseteq B_1$$

偏序集 $(B(U, AT, V, F), \leq)$ 是一个完备格, 其中 $B(U, AT, V, F)$ 是由所有相容概念构成的集合。该结论与经典概念格中的相关结论类似, 具体证明详见文献[29], 在此不再详述。事实上, 依据上述定义, 对于任意覆盖 U/R , 有

$$(U/R)^+ = R^* = \{m \in AT \mid ((x, y), m) \in J, \forall (x, y) \in R\}$$

对于任意属性子集 $B \subseteq AT$, 有

$$B^+ = \alpha(R_B) = \{X \mid (X, X) \in B(U, U, R_B)\} = U/R_B$$

定理 3 设 (C, B) 是一个相容概念, 则 $C = U/R_B$ 。

证明 由于 (C, B) 是一个相容概念, 所以由定义 7 即得 $B^+ = C$ 。又由于 $B^+ = U/R_B$, 故 $C = U/R_B$ 成立, 证毕。

定理 4 在上述定义中, 设 R, R_1 和 R_2 是论域 U 中的相容关系, $B, D \subseteq AT$ 。有下述结论成立:

- 1) $U/R_1 \subseteq U/R_2 \Rightarrow (U/R_2)^+ \subseteq (U/R_1)^+$
- 2) $B \subseteq D \Rightarrow D^+ \subseteq B^+$
- 3) $U/R \subseteq (U/R)^{++}; B \subseteq B^{++}$
- 4) $(U/R)^+ = (U/R)^{+++}; B^+ = B^{+++}$

证明 1) 当 $U/R_1 \subseteq U/R_2$ 时, 有 $\beta(U/R_1) \subseteq \beta(U/R_2)$, 进而由定义 7 即得结论成立。

2) 由于 $B \subseteq D \Rightarrow R_D \subseteq R_B \Rightarrow U/R_D \subseteq U/R_B \Rightarrow D^+ \subseteq B^+$

B^+ ,故 $B \subseteq D \Rightarrow D^+ \subseteq B^+$ 成立。

3) 当 $(U/R)^+ = B$ 时,有 $(U/R)^{++} = B^+ = U/R_B$ 成立。又 $(U/R)^+ = B = \{m \in AT \mid R \subseteq \{m\}^*\}$, 故 $m \in B \Leftrightarrow R \subseteq \{m\}^* \Leftrightarrow R \subseteq R_m$, 从而有 $R \subseteq R_B$ 。显然,此时由 $R \subseteq R_B$ 即得 $U/R \subseteq U/R_B$, 进而 $U/R \subseteq (U/R)^{++}$ 成立。由于 $B^{++} = (U/R_B)^+ = B \cup \{m \in (AT - B) \mid R_B \subseteq \{m\}^*\}$ 故结论 $B \subseteq B^{++}$ 成立。

4) 由上述结论可知 $U/R \subseteq (U/R)^{++} \Rightarrow \overline{(U/R)^{+++}} \subseteq (U/R)^+$, 此外, 又由结论 3) 即得 $(U/R)^+ \subseteq ((U/R)^+)^{++} = (U/R)^{+++}$, 故 $(U/R)^+ = (U/R)^{+++}$ 成立;同理可证 $B^+ = B^{+++}$, 证毕。

定理 5 对于任意 $B \subseteq AT$, (B^+, B^{++}) 是一个相容概念,且 U/R_B 是 (B^+, B^{++}) 的外延。

证明 由 $B^{+++} = B^+$ 即得 (B^+, B^{++}) 是一个相容概念,进而由 $B^+ = U/R_B$ 可知 U/R_B 是相容概念 (B^+, B^{++}) 的外延。证毕。

由上述定理可知, $(B(U, AT, V, F), \leq)$ 可以将所有覆盖的集合 $\{U/R_B \mid B \subseteq AT\}$ 以格的形式有机地组织起来,由此,本文将 其 视 为 集 值 信 息 系 统 (U, AT, V, F) 中的重要代数结构。

定理 6 设 $\text{Int}(S)$ 是 $S = (U, AT, V, F)$ 中所有相容概念内涵的集合,则对于任意 $B \subseteq AT$, 有

$$B^{++} = \{L \subseteq \text{Int}(S) \mid B \subseteq L\}$$

证明 与参考文献[29]中有关结论的证明过程相类似,在此不再详述。证毕。

例如,基于定义 7 中的算子,可以从表 3 导出一个如图 3 的格结构,该结构即为表 2 所示集值信息系统中的相容概念格。图 3 中的 $X_1/X_2/\cdots/X_l$ 表示覆盖 $\{X_1, X_2, \cdots, X_l\}$, 任意“ u_i ”表简化表示为“ i ”。

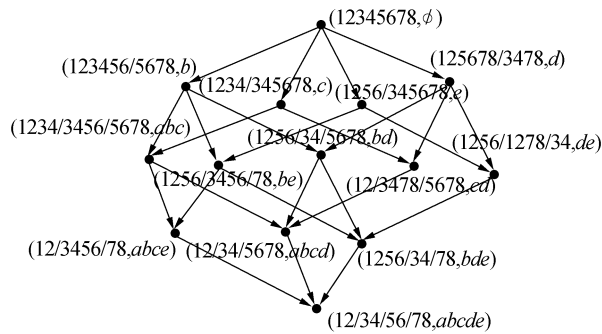


图 3 一个相容概念格
Fig.3 A tolerance concept lattice

4 集值信息系统中的约简、核、依赖等知识的获取

定理 1 约简、核、依赖等作为粗糙集研究的重要组成部分,也是本文研究的重点。针对集值信息系统中的相关问题,本文给出一种基于定理 6 的求解方法。此节内容是通过借鉴文献[12]得到的。

在 $S = (U, AT, V, F)$ 中,设 $m \in B \subseteq AT$, 若 $R_B \neq R_{B-m}$, 称 m 是 B 中的必要属性; B 中所有必要属性的集合称为 B 的核,记为 $\text{Core}(B)$; 设 $C \subseteq B \subseteq AT$, 若 C 是满足 $R_B = R_C$ 的最小属性子集,称 C 是 B 的一个约简; 设 $B, D \subseteq AT$, 若 $R_B \subseteq R_D$, 称 $B \rightarrow D$ 是一个依赖。

定理 7 设 $m \in B \subseteq AT$, 若满足下述条件:
 $\{L \in \text{Int}(S) \mid (B - m) \subseteq L\} \neq \{L \in \text{Int}(S) \mid B \subseteq L\}$

定理 4 则 $m \in \text{Core}(B)$ 成立。

证明 假设 $m \notin \text{Core}(B)$ 成立。有如下结论:
 $R_{B-m} = R_B \Leftrightarrow U/R_{B-m} = U/R_B \Leftrightarrow (B - m)^{++} = B^{++}$
进而由 $(B - m)^{++} = B^{++}$ 即得

$$\{L \in \text{Int}(S) \mid (B - m)^{++} \subseteq L\} = \{L \in \text{Int}(S) \mid B^{++} \subseteq L\}$$

又 $(B - m) \subseteq L \Leftrightarrow (B - m)^{++} \subseteq L$ 且 $B \subseteq L \Leftrightarrow B^{++} \subseteq L$, 故 $\{L \in \text{Int}(S) \mid (B - m) \subseteq L\} = \{L \in \text{Int}(S) \mid B \subseteq L\}$ 与条件相矛盾! 所以 $m \in \text{Core}(B)$ 成立, 证毕。

定理 8 设 $C \subseteq B \subseteq AT$, 若 C 是满足下述条件的最小属性子集:

$$\{L \in \text{Int}(S) \mid B \subseteq L\} = \{L \in \text{Int}(S) \mid C \subseteq L\}$$

则 C 是 B 的一个约简。

证明 基于定理 6, 由上述条件即得 $B^{++} = C^{++}$, 进而有

$B^{++} = C^{++} \Leftrightarrow B^+ = C^+ \Leftrightarrow U/R_B = U/R_C \Leftrightarrow R_B = R_C$
又由于 C 是满足上述条件的 B 的最小属性子集, 故 C 是 B 的一个约简, 证毕。

在表 2 中所示的集值信息系统中, 当 $B = \{a, b, c, d\}$ 时, 依据上述定理, 并参照表 5 中的包含关系, 我们不难发现 $\{a, d\}$ 和 $\{b, c, d\}$ 是 B 的约简, $\text{Core}(B) = \{d\}$; 而当 $B = \{a, b, c\}$ 时, $\{a\}$ 和 $\{b, c\}$ 是 B 的约简, $\text{Core}(B) = \emptyset$ 。为便于表示, 表 5 中的任意属性子集 $\{m_1, m_2, \cdots, m_n\}$ 简化为 $m_1 m_2 \cdots m_n$ 。

定理 9 设 $B, D \subseteq AT$, 若满足条件:
 $\{L \in \text{Int}(S) \mid B \subseteq L\} \subseteq \{L \in \text{Int}(S) \mid D \subseteq L\}$
则 $B \rightarrow D$ 是一个依赖。

证明 由条件可知:
 $\cap \{L \in \text{Int}(S) \mid D \subseteq L\} \subseteq \cap \{L \in \text{Int}(S) \mid B \subseteq L\}$
进而由定理 6 即得 $D^{++} \subseteq B^{++}$ 。由于

$D^{++} \subseteq B^{++} \Leftrightarrow B^+ \subseteq \bar{D}^+ \Leftrightarrow U/R_B \subseteq U/R_D \Leftrightarrow R_B \subseteq R_D$
故 $B \rightarrow D$ 是一个依赖。证毕。

一个信息系统中往往包含着大量的冗余依赖,例如,当 $D \subseteq B$ 时, $B \rightarrow D$ 是绝对冗余的; 当 $B_1 \subseteq B_2$ 且

$D_2 \subseteq D_1$ 时, 由 $B_1 \rightarrow D_1$ 可以推导出 $B_2 \rightarrow D_2$, 即 $B_2 \rightarrow D_2$ 是相对冗余的。在表 2 中, 去掉上述冗余依赖后, 一个规模较小的依赖集如表 6。在表 6 中, $m_1 m_2 \cdots m_n$ 表示集合 $\{m_1, m_2, \cdots, m_n\}$ 。

表 5 一些属性子集与内涵之间的包含关系
Table 5 Inclusion relation between some attributes sets and intents

属性子集	相容概念内涵													
	\emptyset	b	c	d	e	bd	be	cd	de	abc	bde	$abcd$	$abce$	$abcde$
$abcd$												\subseteq		\subseteq
a										\subseteq		\subseteq	\subseteq	\subseteq
b		\subseteq				\subseteq	\subseteq			\subseteq	\subseteq	\subseteq	\subseteq	\subseteq
c			\subseteq					\subseteq		\subseteq		\subseteq	\subseteq	\subseteq
d				\subseteq		\subseteq		\subseteq	\subseteq		\subseteq	\subseteq		\subseteq
ab										\subseteq		\subseteq	\subseteq	\subseteq
ac										\subseteq		\subseteq	\subseteq	\subseteq
ad												\subseteq		\subseteq
bc										\subseteq		\subseteq	\subseteq	\subseteq
bd						\subseteq					\subseteq	\subseteq		\subseteq
cd								\subseteq				\subseteq		\subseteq
abc										\subseteq		\subseteq	\subseteq	\subseteq
abd												\subseteq		\subseteq
acd												\subseteq		\subseteq
bcd												\subseteq		\subseteq
\emptyset	\subseteq	\subseteq	\subseteq	\subseteq	\subseteq	\subseteq	\subseteq	\subseteq	\subseteq	\subseteq	\subseteq	\subseteq	\subseteq	\subseteq

表 6 一个依赖集
Table 6 A dependency set

$a \rightarrow abc$	$ad \rightarrow abcd$	$bcd \rightarrow abcd$	$ae \rightarrow abce$
$ade \rightarrow abcde$	$cde \rightarrow abcde$	$ce \rightarrow abce$	$bc \rightarrow abc$

5 结束语

我们知道,二元关系是粗糙集的核心要素,而概念格本质上又是一种以二元关系为研究对象的数学工具,因此将概念格融入到粗糙集研究中,必将有助于拓展粗糙集的分析能力。本文尝试将概念格中的概念思维、格代数结构、伽罗瓦连接等引入到集值信息系统中,重点探讨了基于相容关系的粒化模型和集值系统中的格代数结构,该代数结构可以将论域中的所有覆盖以格的形式有机组织起来。此外,本文还探讨了集值信息系统中的约简、核等问题。理论推理和实例验证揭示了本文结论的合理性和有效性。本文从概念格视角提出了集值信息系统中的知识获取方法,有助于人们深入理解粗糙集理论,也有助于揭示两种理论之间的紧密联系。相关研究内容仍将是我们的下一阶段的研究重点。

参考文献:

[1] PAWLAK Z. Rough sets[J]. International journal of computer & information sciences, 1982, 11(5): 341-356.
[2] PAWLAK Z. Rough sets: theoretical aspects of reasoning about data[M]. Dordrecht: Kluwer Academic Publishers, 1991.
[3] 王国胤, 张清华, 胡军. 粒计算研究综述[J]. 智能系统学报, 2007, 2(6): 8-26.
WANG Guoyin, ZHANG Qinghua, HU Jun. An overview of granular computing[J]. CAAI transactions on intelligent systems, 2007, 2(6): 8-26.
[4] 王国胤, 姚一豫, 于洪. 粗糙集理论及应用研究综述[J]. 计算机学报, 2009, 32(7): 1229-1246.
WANG Guoyin, YAO Yiyu, YU Hong. A survey on rough set theory and applications[J]. Chinese journal of computers, 2009, 32(7): 1229-1246.
[5] 伞冶, 叶玉玲. 粗糙集理论及其在智能系统中的应用[J]. 智能系统学报, 2007, 4(2): 40-47.
SAN Ye, YE Yuling. Rough set theory and its application in the intelligent systems[J]. CAAI transactions on intelligent systems, 2007, 4(2): 40-47.
[6] WILLE R. Restructuring lattice theory: an approach based on hierarchies of concepts[M]//RIVAL I. Ordered Sets.

- Netherlands: Springer, 1982: 445-470.
- [7] LI Jinhai, MEI Changlin, WANG Lidong, et al. On inference rules in decision formal contexts[J]. International journal of computational intelligence systems, 2015, 8(1): 175-186.
- [8] LI Jinhai, MEI Changlin, WANG Junhong, et al. Rule-preserved object compression in formal decision contexts using concept lattices[J]. Knowledge-based systems, 2014, 71: 435-445.
- [9] QI Jianjun, QIAN Ting, WEI Ling. The connections between three-way and classical concept lattices[J]. Knowledge-based systems, 2016, 91: 143-151.
- [10] ZHAI Yanhui, LI Deyu, QU Kaishe. Decision implication canonical basis: a logical perspective[J]. Journal of computer and system sciences, 2015, 81(1): 208-218.
- [11] 张文修, 姚一豫, 梁怡. 粗糙集与概念格[M]. 西安: 西安交通大学出版社, 2006.
- [12] KANG Xianping, LI Deyu, WANG Suge, et al. Rough set model based on formal concept analysis[J]. Information sciences, 2013, 222: 611-625.
- [13] 曲开社, 翟岩慧, 梁吉业, 等. 形式概念分析对粗糙集理论的表示及扩展[J]. 软件学报, 2007, 18(9): 2174-2182.
- QU Kaishe, ZHAI Yanhui, LIANG Jiye, et al. Representation and extension of rough set theory based on formal concept analysis[J]. Journal of software, 2007, 18(9): 2174-2182.
- [14] 仇国芳, 张志霞, 张伟. 基于粗糙集方法的概念格理论研究综述[J]. 模糊系统与数学, 2014, 28(1): 168-177.
- QIU Guofang, ZHANG Zhixia, ZHANG Wei. A survey for study on concept lattice theory via rough set[J]. Fuzzy systems and mathematics, 2014, 28(1): 168-177.
- [15] WAN Qing, WEI Ling. Approximate concepts acquisition based on formal contexts[J]. Knowledge-based systems, 2015, 75: 78-86.
- [16] KENT R E. Rough concept analysis: a synthesis of rough sets and formal concept analysis[J]. Fundamenta informaticae, 1996, 27(2-3): 169-181.
- [17] CHEN Jinkun, LI Jinjin, LIN Yaojin, et al. Relations of reduction between covering generalized rough sets and concept lattices[J]. Information sciences, 2015, 304: 16-27.
- [18] LI Jinhai, REN Yue, MEI Changlin, et al. A comparative study of multigranulation rough sets and concept lattices via rule acquisition[J]. Knowledge-based systems, 2016, 91: 152-164.
- [19] SHAO Mingwen, LEUNG Y. Relations between granular reduct and dominance reduct in formal contexts[J]. Knowledge-based systems, 2014, 65: 1-11.
- [20] TAN Anhui, LI Jinjin, LIN Guoping. Connections between covering-based rough sets and concept lattices[J]. International journal of approximate reasoning, 2015, 56: 43-58.
- [21] KRYSZKIEWICZ M. Rough set approach to incomplete information systems[J]. Information sciences, 1998, 112(1/2/3/4): 39-49.
- [22] KRYSZKIEWICZ M. Rules in incomplete information systems[J]. Information sciences, 1999, 113(3/4): 271-292.
- [23] 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.
- WANG Guoyin. Extension of rough set under incomplete information systems[J]. Journal of computer research and development, 2002, 39(10): 1238-1243.
- [24] LEUNG Y, LI Deyu. Maximal consistent block technique for rule acquisition in incomplete information systems[J]. Information sciences, 2003, 153: 85-106.
- [25] ZHANG Wenxiu, MI Jusheng. Incomplete information system and its optimal selections[J]. Computers & mathematics with applications, 2004, 48(5/6): 691-698.
- [26] GUAN Yanyong, WANG Hongkai. Set-valued information systems[J]. Information sciences, 2006, 176(17): 2507-2525.
- [27] QIAN Yuhua, DANG Chuangyin, LIANG Jiye, et al. Set-valued ordered information systems[J]. Information sciences, 2009, 179(16): 2809-2832.
- [28] QIAN Y H, LIANG J Y, SONG P, et al. On dominance relations in disjunctive set-valued ordered information systems[J]. International journal of information technology & decision making, 2010, 9(1): 9-33.
- [29] GANTER B, WILLE R. Formal concept analysis: mathematical foundations[M]. Berlin Heidelberg: Springer-Verlag, 1999.

作者简介:



康向平, 男, 1982 年生, 博士后, 主要研究方向为概念格、粗糙集、粒计算。



苗夺谦, 男, 1964 年生, 教授, 博士生导师, 中国人工智能学会理事、中国计算机学会杰出会员、中国自动化学会智能自动化专委会委员、上海市计算机学会理事、上海市人工智能学会理事, 主要研究方向为粗糙集、粒计算、Web 智能、数据挖掘和机器学习。主持完成国家级、省部级自然科学基金与科技攻关项目多项, 参与完成 973 计划项目、863 计划项目、国家自然科学基金重大研究计划项目、上海市科委重大科技攻关项目等多项。作为项目组成员, 曾获国家教委科技进步三等奖、教育部科技进步一等奖、上海市技术发明一等奖、重庆市自然科学一等奖等。发表学术论文 260 余篇, 其中被 SCI 或 EI 检索 150 余篇。编著教材及著作 12 部, 获得专利授权 9 项。