形式概念分析在不完备信息系统中的知识获取

李 想[1] 王素格[2,3] 李德玉[2,3] 康向平[4] 翟岩慧[2,3]
(山西大学数学科学学院 太原 030006)[1] (山西大学计算机与信息技术学院 太原 030006)[2]
(山西大学计算智能与中文信息处理教育部重点实验室 太原 030006)[3]
(同济大学电子与信息工程学院 上海 201804)[4]

摘 要 在实际应用中,信息系统通常不满足完备性,直接采用经典粗糙集无法进行有效求解。针对该问题,在粗糙集中引入形式概念分析,通过探讨二者之间的关系,在不完备信息系统中提出了两大理论融合下的知识获取模型。在该模型中,首先将不完备信息系统转换为单值形式背景,提出了相容概念和相容概念格,然后基于相容概念对不完备信息系统中的一些常见问题(如上下近似算子、核、约简等)的求解进行了研究,最后探讨了相容概念在不完备决策表中的应用。该模型不仅对两种理论之间的融合研究进行了有益探索,而且为不完备信息系统中的一些基本问题求解提供了一种新的思路。

关键词 形式概念分析,粗糙集,相容关系,不完备信息系统
中图法分类号 TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2014.07.052

Knowledge Acquisition in Incomplete Information System Based on Formal Concept Analysis

LI Xiang[1] WANG Su-ge[2,3] LI De-yu[2,3] KANG Xiang-ping[4] ZHAI Yan-hui[2,3]
(School of Mathematics Science, Shanxi University, Taiyuan 030006, China)[1]
(School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China)[2]
(Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan 030006, China)[3]
(Electronic and Information Engineering, Tongji University, Shanghai 201804, China)[4]

Abstract In practical applications, some common problems in information system can't be solved effectively based on the classical rough set due to its incompleteness. To solve this problem, the paper introduced formal concept analysis into rough set, proposed a knowledge acquisition model in incomplete information systems by discussing the relationship between formal concept analysis and rough set theory. First by converting incomplete information system into a one-valued context, consistent concepts and consistent concept lattice were proposed. Then some common problems in incomplete information systems were studied, such as upper and lower approximations, cores, reducts, etc. Finally, the application of consistent concepts in a decision table was discussed. The proposed model not only explores the fusion of two theories greatly, but provides a new idea for solving some basic problems of incomplete information system.

Keywords Formal concept analysis, Rough set, Consistent relation, Incomplete information system

1 引言

粗糙集理论是1982年由波兰学者Pawlak[1]提出的一种处理不确定性信息的数学方法,粗糙集理论由于不需要关于数据额外的信息,并且能够分析处理不精确性和不完全的信息和知识,因此在知识发现、数据挖掘和模式识别等领域有广泛的应用[2-6]。1982年德国学者Wille[7]提出了形式概念分析理论,形式概念分析不仅具有良好的数学性质,还能够直观展现概念间泛化和特化关系,目前已作为一种有效的工具在数据挖掘、软件工程、信息检索等诸多领域有广泛的应用[8-12]。

粗糙集和形式概念分析作为数据建模和数据分析的互补性工具,尽管在研究方法上存有差异,但因有相同的研究背景和目标,必然也存在一些共性。研究二者理论的异同,各取所长,并且构建两者间的联系,国内外学者对此也已经做了众多的研究[13-18]。例如,Kent将粗糙集和概念格融合提出了粗糙形式概念分析理论[13];Yao在数据分析中对形式概念分析和粗糙集进行了对比研究[14];Lai和Zhang基于形式概念分析

到稿日期:2013-09-23 返修日期:2013-12-27 本文受国家自然科学基金资助项目(61175067,61272095,60970014),山西省科技攻关项目(20110321027-02),山西省回国留学人员科研项目(2013-014)资助。

李 想(1988-),女,硕士生,主要研究方向为形式概念分析;王素格(1964-),女,教授,主要研究方向为自然语言处理、智能检索等,E-mail:wsg@sxu.edu.cn(通信作者);李德玉(1965-),男,教授,博士生导师,主要研究方向为人工智能等;康向平(1982-),男,博士后,主要研究方向为数据挖掘;翟岩慧(1981-),男,博士生,主要研究方向为概念格。

和粗糙集理论对模糊背景下的概念格进行了研究，并且验证了基于粗糙集的概念格比基于形式概念分析的概念格表达知识的能力要弱[15]；Wang 揭示了概念的外延和粗糙集理论的等价类的一般关系，并且研究了概念格中形式背景的约简和粗糙集理论中属性约简的关系[16]；Kang 把形式概念分析引入到粗糙集中，提出了一种基于形式概念分析的粗糙集模型，为粗糙集提供了一种新的基于形式概念分析理论的解释[17]；Wei 和 Qi 分别在粗糙集及形式概念分析视角下研究了约简理论，并对二者进行了比较研究[18]等等。

本文提出了一种基于形式概念分析的不完备信息系统中的知识获取方法，该方法是对文献[17]中所提方法的拓展和延伸。在文献[17]中，作者仅讨论了完备条件下的离散信息系统。然而在实际应用中，信息系统通常是不完备的，基于此，本文在不完备信息系统中引入了形式概念分析，探讨了形式概念分析在不完备信息系统中的应用。

本文第 2 节概要介绍了形式概念分析和粗糙集的基本概念；第 3 节将不完备信息系统转换为单值形式背景，提出了相容概念和相容概念格；第 4 节探讨了基于相容概念的知识获取方法；第 5 节给出了相容概念在不完备决策表中的应用。

2 基本概念

2.1 形式概念分析

定义 1[19] 形式背景为一个三元组 $K=(G,M,I)$，其中 G 表示由所有对象构成的集合，M 表示由所有属性构成的集合，$I\subseteq G\times M$ 表示由 G 和 M 中元素之间的关系构成的关系集合，对于任意 $g\in G$ 和 $m\in M$，$(g,m)\in I$ 表示“对象 g 具有属性 m”。

定义 2[19] 在 $K=(G,M,I)$ 中，设 $A\subseteq G,B\subseteq M$，记

$A'=\{m\in M|(g,m)\in I,\forall g\in A\}$

$B'=\{g\in G|(g,m)\in I,\forall m\in B\}$

如果 $A'=B$ 且 $B'=A$，则称 (A,B) 为一个形式概念，简称概念，其中 A 为外延，B 为内涵。由所有概念组成的集合记为 $\mathscr{B}(K)$。对于任意两个概念 (A_1,B_1) 和 (A_2,B_2)，二者之间的偏序关系定义如下：

$(A_1,B_1)\preceq(A_2,B_2)\Leftrightarrow A_1\subseteq A_2(\Leftrightarrow B_1\supseteq B_2)$

引理 1[19] 对任意对象子集 $A,A_1,A_2\subseteq G$ 和任意属性子集 $B,B_1,B_2\subseteq M$，有下列结论成立：

(1) $A_1\subseteq A_2\Rightarrow A_2'\subseteq A_1'$；(2) $B_1\subseteq B_2\Rightarrow B_2'\subseteq B_1'$；

(3) $A\subseteq A''$；$B\subseteq B''$；(4) $A'=A'''$；$B'=B'''$。

引理 2[19] 偏序集 $(\mathscr{B}(K),\preceq)$ 是一个完备格，相应的下确界和上确界为：

$$\bigwedge_{t\in T}(A_t,B_t)=\left(\bigcap_{t\in T}A_t,\left(\bigcup_{t\in T}B_t\right)''\right)$$

$$\bigvee_{t\in T}(A_t,B_t)=\left(\left(\bigcup_{t\in T}A_t\right)'',\bigcap_{t\in T}B_t\right)$$

其中，$(A_t,B_t)\in\mathscr{B}(K)$，$T$ 是指标集。称完备格 $(\mathscr{B}(K),\preceq)$ 是由形式背景 K 导出的概念格，简记为 $\underline{\mathscr{B}}(K)$。

2.2 粗糙集

定义 3[20] 信息系统 S 表示为一个四元组 (U,AT,V,f)，其中，U 是由有限对象构成的非空集合，AT 是由有限属性构成的非空集合，$V=\bigcup_{a\in AT}V_a$ 是属性值域，V_a 是属性 a 的值域，f 是从 $U\times AT$ 到 V 的映射，即对任意的 $x\in U,a\in AT$，有 $f(x,a)\in V_a$；如果至少存在一个属性 $a\in AT,V_a$ 包含空值，则称 S 是一个不完备信息系统。

例如，表 1 是一个关于小汽车的不完备信息系统，其中 a_1 表示 Price，a_2 表示 Mileage，a_3 表示 Size，a_4 表示 MaxSpeed，* 表示对应的属性值域的某个值。

表 1 一个关于小汽车的不完备信息系统

	a_1	a_2	a_3	a_4
u_1	high	high	full	low
u_2	low	*	full	low
u_3	*	*	compact	high
u_4	high	*	full	high
u_5	*	*	full	high
u_6	low	high	full	*

在不完备信息系统 $S=(U,AT,V.f)$ 中，每一个属性子集 $B\subseteq AT$ 都决定了 U 上的一个二元关系 $R(B)$，设 $B,C\subseteq AT$，如果 $m\in B$ 且 $R(B)=R(B-m)$，则称 m 在 B 中是不必要的，否则称 m 在 B 中是必要的；如果每一个 $m\in B$ 都为 B 中必要的，则称 B 是独立的，否则称 B 是依赖的，由所有独立的属性子集构成的集合记为 IND_S；如果 $C\subseteq B,C$ 是独立的且 $R(B)=R(C)$，则称 C 是 B 的一个约简，B 的所有约简记为 $Red_S(B)$；由 B 中所有必要属性组成的集合称为 B 的核，记为 $Core_S(B)$；如果 $R(B)\subseteq R(C)$，则称 $B\rightarrow C$ 是一个函数依赖。

3 不完备信息系统中的相容概念及格结构

一个不完备信息系统本质上是一个不完备多值形式背景[19]，而形式概念分析处理的数据集通常是典型单值形式背景。基于此，对任意 $x,y\in U,m\in AT$，本文通过如下规则：

$(x,y)I_Sm\Leftrightarrow f(x,m)=f(y,m)$ or $f(x,m)=*$ or $f(y,m)=*$

将不完备信息系统

$S=(U,AT,V,f)$

变换成一个单值形式背景

$K_S=(U\times U,AT,I_S)$

例如，表 1 所列的不完备信息系统转换后的单值形式背景如表 2 所列。

定理 1 在形式背景 $K_S=(U\times U,AT,I_S)$ 中，设 $B,B_1,B_2\subseteq AT$，则下列结论成立：

(1) B' 是论域 U 上的相容关系；

(2) 如果 $B_1''=B_2''$，则 $B_1'=B_2'$，即 B_1' 和 B_2' 是论域 U 上的同一个相容关系；

(3) 对于任意属性子集 B，总存在属性子集 $B''\in\mathscr{Int}(S)$，满足：在论域 U 上，B 与 B'' 下的相容关系是相同的，其中 $\mathscr{Int}(S)$ 是 K_S 中所有内涵的集合。

证明：(1) 显然 B' 为论域 U 上的一个二元关系，且满足自反性和对称性，即 B' 是论域 U 上的相容关系。

(2) 由引理 1 得 $B_1''=B_2''\Leftrightarrow B_1'''=B_2'''\Leftrightarrow B_1'=B_2'$。

(3) 由引理 1 得 $B'=B'''$，即在论域 U 上，B 与 B'' 下的相容关系是相同的。证毕。

由定理 1 可知，任何属性集 $B\subseteq AT$ 与其内涵 $B''\in\mathscr{Int}(S)$ 下的相容关系是相同的，换句话说，$\mathscr{Int}(S)$ 中包含多少元素，那么在论域 U 中就有多少个互不相同的相容关系（在本文所定义的转换规则下，所有可能得到的相容关系）。

表 2　变换后的单值形式背景(由表 1 导出)

	a_1	a_2	a_3	a_4
(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_2,u_2)	×	×	×	×
(u_2,u_3)	×	×		
(u_2,u_4)		×	×	
(u_2,u_5)	×	×	×	
(u_2,u_6)	×	×	×	×
(u_3,u_3)	×	×	×	×
(u_3,u_4)	×	×		×
(u_3,u_5)	×	×		×
(u_3,u_6)	×	×		×
(u_4,u_4)	×	×	×	×
(u_4,u_5)	×	×	×	×
(u_4,u_6)		×	×	×
(u_5,u_5)	×	×	×	×
(u_5,u_6)	×	×	×	×
(u_6,u_6)	×	×	×	×
(u_2,u_1)		×	×	×
(u_3,u_1)	×	×		
(u_4,u_1)	×	×	×	
(u_5,u_1)	×	×	×	
(u_6,u_1)		×	×	×
(u_3,u_2)	×	×		
(u_4,u_2)		×	×	
(u_5,u_2)	×	×	×	
(u_6,u_2)	×	×	×	×
(u_4,u_3)	×	×		×
(u_5,u_3)	×	×		×
(u_6,u_3)	×	×		×
(u_5,u_4)	×	×	×	×
(u_6,u_4)		×	×	×
(u_6,u_5)	×	×	×	×

下面给出不完备信息系统中相容概念的定义。

定义 4[21]　已知 R 是论域 U 上的一个相容关系，设 $Q\subseteq U$，如果 Q 是满足下列条件的最大子集，且 $x,y\in Q$，则 $(x,y)\in R$，称 Q 是论域 U 中 R 关系下的极大相容块。所有 R 关系下的极大相容块的集合记为：

$U/R=\{Q_1,Q_2,\cdots,Q_n\}$

事实上，U/R 是论域 U 的一个覆盖，且 U/R 与相容关系 R 之间是一一对应的，即如果已知 U/R，则可推导出 $R=\{(u,v)\in Q_i\times Q_j\mid Q_i,Q_j\in U/R\}$；如果已知相容关系 R，则可推导出 U/R。

在形式背景 $K_S=(U\times U,AT,I_S)$ 中，假设 $R\subseteq U\times U$，$B\subseteq AT$，(R,B) 是 K_S 中的一个概念，由于 U/R 与 R 之间相互唯一决定，因此可将 (R,B) 变换为 $(U/R,B)$，也可将 $(U/R,B)$ 变换为 (R,B)。

定义 5　称变换后的概念 $(U/R,B)$ 为相容概念，所有相容概念的集合记为 $\mathscr{B}(S)$。相应的概念格称为相容概念格，记为 $\underset{\sim}{\mathscr{B}}(S)$。

显然，U/R 为论域 U 中内涵 B 下的覆盖(由极大相容块构成)，内涵 B 是指与覆盖 U/R(由极大相容块构成)相对应的最大属性子集。

生成相容概念格的概要步骤如图 1 所示。

例如，基于图 1 所示的概要步骤，由表 2 得到的相容概念格如图 2 所示，其中，$(\{\{u_1,u_3,u_4,u_5\},\{u_2,u_3,u_5,u_6\}\},\{a_1,a_2\})$ 简记为 $\{(1345,2356),a_1a_2\}$，其余类似。

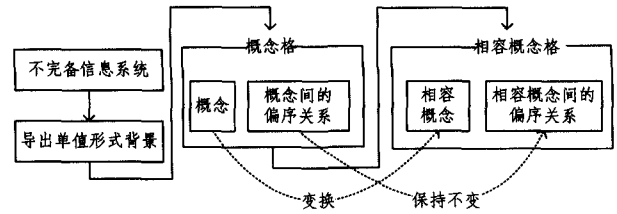

图 1　生成相容概念格的概要步骤

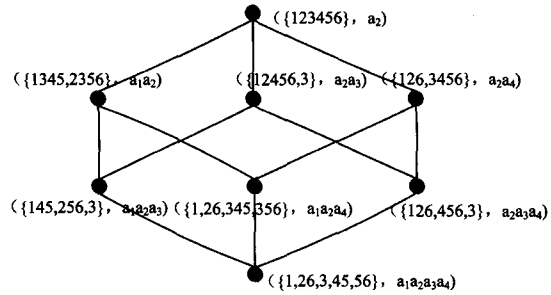

图 2　表 1 的相容概念格

设 U/R_1，U/R_2 是论域的覆盖，如果对任意 $P\in U/R_1$，总存在 $Q\in U/R_2$ 满足 $P\subseteq Q$，则称 U/R_2 粗于 U/R_1，或称 U/R_1 细于 U/R_2，记为 $U/R_1\ \widetilde{\subseteq}\ U/R_2$。如果 $U/R_1\ \widetilde{\subseteq}\ U/R_2$ 且 $U/R_2\ \widetilde{\subseteq}\ U/R_1$，则记为 $U/R_1=U/R_2$。

定理 2　在 $K_S=(U\times U,AT,I_S)$ 中，设 $R,R_1,R_2\subseteq U\times U$，$B,B_1,B_2\subseteq AT$，则下列各式成立：

(1) $U/R_1\ \widetilde{\subseteq}\ U/R_2\Rightarrow R_2'\subseteq R_1'$；

(2) $B_1\subseteq B_2\Rightarrow U/B_2'\ \widetilde{\subseteq}\ U/B_1'$；

(3) $U/R\ \widetilde{\subseteq}\ U/R''$；

(4) $U/B'=U/B'''$。

证明：由引理 1 即得。证毕。

定理 3　在经变换后的完备格 $\underset{\sim}{\mathscr{B}}(S)=(\mathscr{B}(S),\preceq)$ 中，相应的下确界和上确界为：

$$\bigwedge_{t\in T}(U/R_t,B_t)=(U/\bigcap_{t\in T}R_t,(\bigcup_{t\in T}B_t)'')$$

$$\bigvee_{t\in T}(U/R_t,B_t)=(U/(\bigcup_{t\in T}R_t)'',\bigcap_{t\in T}B_t)$$

其中，$(U/R_t,B_t)\in\mathscr{B}(S)$，$T$ 是指标集。

证明：与文献[19]中的证明过程类似。

相容概念格清晰地展现了不同粗细程度下的所有覆盖(由极大相容块构成)的层次结构模型。在相容概念格中，依据定理 3 的求解方法，我们可以算出任意两个或多个相容概念的上确界和下确界，即基于形式概念分析，可以求解同时粗于两个或多个覆盖(由极大相容块构成)的最细覆盖(由极大相容块构成)，也可以求解同时细于两个或多个覆盖(由极大相容块构成)的最粗覆盖(由极大相容块构成)。

例如，在图 2 中，同时粗于覆盖{145,256,3}和覆盖{1,26,345,356}的最细覆盖是{1345,2356}，同时细于覆盖{145,256,3}和覆盖{1,26,345,356}的最粗覆盖是{1,26,3,45,56}。

4　基于相容概念的知识获取

引理 3[19]　在形式背景 $K=(G,M,I)$ 中，设 $B\subseteq M$，如果 $\mathscr{M}$ 是 K 中所有内涵的集合，则

$B''=\bigcap\{T\in\mathcal{M}\mid B\subseteq T\}$

定理 4 在不完备信息系统 $S=(U,AT,V,f)$ 中，设 $B\subseteq AT$，如果 $(U_B,\vartheta_S^B)\in\mathcal{B}(S)$，则 $\mathcal{U}_B=U/B'$。其中，$\vartheta_S^B=\bigcap\{T\in\mathcal{M}(S)\mid B\subseteq T\}$，$(U_B,\vartheta_S^B)\in\mathcal{B}(S)$。

证明：由引理 1 可知 $(B')'=B''$ 且 $(B'')'=B'''=B'$，显然 (B',B'') 是一个概念，(B',B'') 经变换后为 $(U/B',B'')\in\mathcal{B}(S)$，又由引理 3 可知 $\vartheta_S^B=B''$，所以 $\mathcal{U}_B=U/B'$。证毕。

由定理 1 可知，论域 U 中属性集 B 下的覆盖（由极大相容块构成）与内涵 B'' 下的覆盖（由极大相容块构成）相同。如果能够得到内涵 B'' 下的覆盖，那么就可得到属性集 B 下的覆盖（由极大相容块构成）。基于此，由定理 4 可知，如果给定所有相容概念，那么属性集 B 下的覆盖（由极大相容块构成）即为 U/B'，其中 $(U/B',B'')\in\mathcal{B}(S)$。

依据定理 4，可求解任意对象子集在某属性子集的上、下近似算子。设 $X\subseteq U,B\subseteq AT,X$ 关于 R 的下近似定义为：

$$\underline{R}(X)=\bigcup_{P\subseteq X\text{ and }P\in\mathcal{U}_B}P$$

相应地，X 关于 R 的上近似定义为：

$$\overline{R}(X)=\bigcup_{P\cap X\neq\phi\text{ and }P\in\mathcal{U}_B}P$$

其中，$R=B'$，$(\mathcal{U}_B,\vartheta_S^B)\in\mathcal{B}(S)$。

下面讨论如何基于相容概念在不完备信息系统中求解约简、独立、核等粗糙集中一些最基本的问题。

定理 5 在不完备信息系统 $S=(U,AT,V,f)$ 中，有下列结论成立：

(1) $IND_S=\{B\subseteq AT\mid\forall a\in B,\vartheta_S^{B-\{a\}}\neq\vartheta_S^B\}$

(2) $Core_S(B)=\{a\in B\mid\vartheta_S^{B-\{a\}}\neq\vartheta_S^B\}$

(3) $C\in Red_S(B)$ 当且仅当 $\vartheta_S^C=\vartheta_S^B$，且不存在 $C_1\subset C$ 满足 $\vartheta_S^{C_1}=\vartheta_S^B$。

证明：由引理 1、引理 3 得，$\vartheta_S^{B-\{a\}}\neq\vartheta_S^B\Leftrightarrow(B-\{a\})''\neq B''\Leftrightarrow(B-\{a\})'''\neq B'''\Leftrightarrow(B-\{a\})'\neq B'\Leftrightarrow R(B-\{a\})\neq R(B)$，显然，结论(1)－(3)是成立的。证毕。

例如，表 1 中所有的相容内涵如表 3 所列，依据定理 5，参照表 3，求解可以得到 $IND_S=\{a_1,a_3,a_4\}$，$Core_S(B)=\{a_1,a_4\}$，$Red_S(B)=\{a_1a_4\}$，其中 $B=a_1a_2a_4$。显然，经比对，本文提出的方法与粗糙集中的求解方法所得到的计算结果是一致的。

表 3 所有相容内涵（表 1 中）

T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8
a_2	a_1a_2	a_2a_3	a_2a_4	$a_1a_2a_3$	$a_1a_2a_4$	$a_2a_3a_4$	$a_1a_2a_3a_4$

定理 6 在不完备信息系统 $S=(U,AT,V,f)$ 中，设 $B,C\subseteq AT$，下面结论成立：

(1) $B\rightarrow C$ 当且仅当 $C\subseteq\vartheta_S^B$；

(2) $B\rightarrow C$ 当且仅当 $\mathcal{U}_B\widetilde{\subseteq}\mathcal{U}_C$。

证明：(1) 由引理 1、引理 3、定理 1 得 $B\rightarrow C\Leftrightarrow R(B)\subseteq R(C)\Leftrightarrow B'\subseteq C'\Leftrightarrow C''\subseteq B''\Leftrightarrow C\subseteq B''\Leftrightarrow C\subseteq\vartheta_S^B$。

(2) 由引理 1、定理 1 和定理 2 得 $B\rightarrow C\Leftrightarrow R(B)\subseteq R(C)\Leftrightarrow B'\subseteq C'\Leftrightarrow C''\subseteq B''\Leftrightarrow U/B''\widetilde{\subseteq}U/C'''\Leftrightarrow U/B'\widetilde{\subseteq}U/C'\Leftrightarrow\mathcal{U}_B\widetilde{\subseteq}\mathcal{U}_C$。证毕。

5 决策表中的相容概念

决策表是一类特殊而重要的信息系统，并且以其可以形式化地描述与表达多数决策问题等优点在决策应用中起着重要的作用。下面将重点讨论不完备信息系统中相容概念在不完备决策表中的应用。

定义 6 不完备信息系统 $S=\{U,AT,V,f\}$ 称为一个不完备决策表，若 $AT=M\cup N,M\cap N=\phi$。其中，M 称为条件属性集，N 称为决策属性集。事实上，不完备决策表由 $S_M=(U,M,V_M,f_M)$ 和 $S_N=(U,N,V_N,f_N)$ 两部分构成，相应地，称 $(U/B',B)\in\mathcal{B}(S_M)$ 为条件相容概念，称 $(U/D',D)\in\mathcal{B}(S_N)$ 为决策相容概念，其中 $B\subseteq M,D\subseteq N$。

例如，表 4 是一个不完备决策表，其中，$M=\{a_1,a_2,\cdots,a_4\}$，$N=\{d_1,d_2\}$。下文中，将 $U=\{u_1,u_2,\cdots,u_6\}$ 简写为 $U=\{1,2,\cdots,6\}$。

定义 7 设 $(U/B',B)\in\mathcal{B}(S_M)$，$(U/D',D)\in\mathcal{B}(S_N)$，称 $(U/B',B)\xrightarrow{Rough}(U/D',D)$ 是由一组决策规则构成的决策组。其中包含的决策规则定义如下：令 $X\in U/B'$ 和 $Y\in U/D'$，$des_B(X)$ 表示对相容类 X 的描述，$des_D(Y)$ 表示对相容类 Y 的描述。决策规则定义为：

$r:des_B(X)\Rightarrow des_D(Y)$

规则的确定性因子定义如下：

$\mu(X,Y)=|Y\cap X|/|X|$，$0<\mu(X,Y)\leqslant 1$

当 $\mu(X,Y)=1$ 时，γ 是确定的决策规则；当 $0<\mu(X,Y)<1$ 时，r 是不确定的决策规则。

表 4 一个关于小汽车的不完备决策表

	a_1	a_2	a_3	a_4	d_1	d_2
u_1	high	high	full	low	good	good
u_2	low	*	full	low	good	good
u_3	*	*	compact	high	*	good
u_4	high	*	full	high	excel	poor
u_5	*	*	full	high	excel	poor
u_6	low	high	full	*	*	poor

例如，在表 4 所列的决策表中，决策组 $(\{1,26,3,45,56\},a_1a_2a_3a_4)\xrightarrow{Rough}(\{123,456\},d_1d_2)$ 包含如下决策规则，其中确定性决策规则：

$(a_1,high)\wedge(a_2,high)\wedge(a_3,full)\wedge(a_4,low)\Rightarrow(d_1,good\text{ or }*)\wedge(d_2,good)$

$(a_1,*)\wedge(a_2,*)\wedge(a_3,compact)\wedge(a_4,high)\Rightarrow(d_1,good\text{ or }*)\wedge(d_2,good)$

$(a_1,high\text{ or }*)\wedge(a_2,*)\wedge(a_3,full)\wedge(a_4,high)\Rightarrow(d_1,excel\text{ or }*)\wedge(d_2,poor)$

$(a_1,low\text{ or }*)\wedge(a_2,high\text{ or }*)\wedge(a_3,full)\wedge(a_4,high\text{ or }*)\Rightarrow(d_1,excel\text{ or }*)\wedge(d_2,poor)$

不确定性决策规则（确定性因子均为 0.5）：

$(a_1,low)\wedge(a_2,high\text{ or }*)\wedge(a_3,full)\wedge(a_4,low\text{ or }*)\Rightarrow(d_1,good\text{ or }*)\wedge(d_2,good)$

$(a_1,low)\wedge(a_2,high\text{ or }*)\wedge(a_3,full)\wedge(a_4,low\text{ or }*)\Rightarrow(d_1,excel\text{ or }*)\wedge(d_2,poor)$

如果决策组 $(U/B',B)\xrightarrow{Rough}(U/D',D)$ 中的所有决策规则均为确定性决策规则，即 $U/B'\widetilde{\subseteq}U/D'$，则称 $B\rightarrow D$ 为 S 上的概念决策依赖。事实上，概念决策依赖就是建立条件相容概念内涵与决策相容概念内涵之间的依赖关系。例如，在

（下转第 260 页）

[19] Lin Hong-kang, Fan Cheng-Xian. The dual form of P-reasoning and idengtification of unknown attribute[J]. International Journal of Digital Content Technology and its Applications, 2012, 6(1): 121-131

[20] Robinson G D, Harry N. Evaluation of two applications of special mixing models to image fusian[J]. Remote Sensenvirion, 2000(71): 272-281

[21] Figue J, Grabisch M, Charbonnle M P. A method for still image interpretation relying on a multi algorithems fusion scheme, Application to human face characterization[J]. Fuzzy sets and Systems, 1999(103): 317-337

[22] 董志荣. 再论信息融合[J]. 舰船论证参考, 2002(4): 8-12

[23] 董志荣. 论信息融合[J]. 情报指挥控制系统与仿真技术, 2001(7): 27-36

[24] 杨宝强, 等. 信息融合技术研究及其应用[J]. 空军工程大学学报, 2005, 5(2): 41-44

[25] Waltz, Llinas J. Multisensor data fusion[M]// Artech House, LNC, 1990: 47-92

(上接第 253 页)

表 4 所列的不完备决策表中，所有概念决策依赖有：$a_2a_4 \to d_1$，$a_1a_2a_4 \to d_1$，$a_2a_3a_4 \to d_1$，$a_1a_2a_3a_4 \to d_1$。

在决策组中，不同的条件属性(或属性集)相对于决策属性集 D 的重要性程度可能不同。为了度量决策组 $(U/B', B) \xrightarrow{Rough} (U/D', D)$ 中某些条件属性(或属性集) $B_1 \subseteq B$ 相对于决策属性集 D 的重要程度，设 $(\mathcal{U}_B, \mathcal{P}_S^B) \in \mathcal{B}(S)$，$R_1 = (\mathcal{P}_S^B)'$，$R_2 = (\mathcal{P}_S^{B-B_1})'$，我们定义了如下所示的基于相容概念的属性重要性因子：

$$\rho_{BD}(B_1) = \frac{1}{|U|}\left(\left|\bigcup_{Y \in \mathcal{U}_D} \underline{R_1}Y\right| - \left|\bigcup_{Y \in \mathcal{U}_D} \underline{R_2}Y\right|\right)$$

其中，

$$\underline{R_1}Y = \bigcup_{X \subseteq Y \text{ and } X \in \mathcal{U}_B} X, \quad \underline{R_2}Y = \bigcup_{X \subseteq Y \text{ and } X \in \mathcal{U}_{B-B_1}} X$$

显然，$\rho_{BD}(B_1)$ 越大，B_1 相对于 D 则越重要。例如，在表 4 中，我们可以得到决策组

$$(\{126, 3456\}, a_2a_4) \xrightarrow{Rough} (\{1236, 3456\}, d_1)$$

则 a_2 和 a_4 相对于 d_1 的重要性因子分别为 0 和 1，显然，相对于决策属性 d_1，条件属性 a_4 的重要性明显要强于 a_2。

事实上，鉴于 a_2 相对于 d_1 的重要性为零，即使将决策组 $(\{126, 3456\}, a_2a_4) \xrightarrow{Rough} (\{1236, 3456\}, d_1)$ 中的条件属性 a_2 去掉，对该决策组中的确定性和不确定性决策规则(决策及相应的确定性因子)也没有任何影响。

结束语 本文把形式概念分析引入到离散的不完备信息系统中，探讨了形式概念分析在不完备信息系统中的知识获取。通过构建不完备信息系统中的相容概念格，对不完备信息系统中的一些常见问题(如上下近似算子、核、约简等)的求解进行了研究，并且讨论了相容概念在不完备决策表中的应用。把形式概念分析引入到不完备信息系统中，有助于我们从形式概念分析的角度更好地理解并解决粗糙集中的一些问题。对于一些特殊的粗糙集模型(如变精度粗糙集模型、模糊粗糙集模型和概率粗糙集模型等)，我们将继续研究如何与形式概念分析进行更为广泛的融合。

参 考 文 献

[1] Pawlak Z. Rough sets [J]. International Journal of Computer and Information Sciences, 1982, 11: 341-356

[2] Chouchoulas A, Shen Q. Rough set-aided keyword reduction for text categorization [J]. Applied Artificial Intelligence, 2001, 15(9): 843-873

[3] Swiniarski R W, Skowron A. Rough set methods in feature selection and recognition [J]. Pattern Recognition Letters, 2003, 24: 833-849

[4] Hu Yi-chung. Rough sets for pattern classification using pairwise-comparison-based tables [J]. Applied Mathematical Modelling, 2013, 37: 7330-7337

[5] 李永敏，朱善军，陈湘辉，等. 基于粗糙集模型的数据挖掘模型[J]. 清华大学学报：自然科学版，1999(1)：110-113

[6] 张颖淳，苏伯洪，曹娟. 基于粗糙集的属性约简在数据挖掘中的应用研究[J]. 计算机科学，2013，40(8)：223-226

[7] Wille R. Restructuring lattice theory: an approach based on hierarchies of concepts[C]// Rival I, ed. Ordered Sets. Dordrecht Reidel, 1982: 445-470

[8] Tonella. Using a concept lattice of decomposition slices for program understanding and impact analysis [J]. IEEE Transactions on Software Engineering, 2003, 29(6): 495-509

[9] Arevalo G, Mens T. Analyzing object-oriented application frameworks using concept analysis [C]// LNCS. 2002, 2426: 53-63

[10] Kaytoue M, Duplessis S, Kuznetsov S O, et al. Mining gene expression data with pattern structures in formal concept analysis [J]. Information Sciences, 2011, 181(10): 1989-2001

[11] 谢志鹏，刘宗田. 概念格与关联规则发现[J]. 计算机研究与发展，2000，37(12)：1415-1421

[12] 沈夏炯，刘宗田. 形式概念分析方法与软件过程改进[J]. 计算机科学，2003(7)：103-105

[13] Kent R E. Rough concept analysis [J]. Fundamenta Informaticae, 1996, 27: 169-181

[14] Yao Yi-yu. A comparative study of formal concept analysis and rough set theory in data analysis [C]// Proceedings of the 3rd International Conference on Rough Sets and Current Trends in Computing, RSCTC'04. LNCS3066, 2004: 59-68

[15] Lai Hong-liang, Zhang De-xue. Concept lattices of fuzzy contexts: formal concept analysis vs. rough set theory [J]. International Journal of Approximate Reasoning, 2009, 50(5): 695-707

[16] Wang Hong, Zhang Wen-xiu. Relationships between concept lattice and rough set [J]. Lecture Notes in Computer Science, 2006, 4029: 438-547

[17] Kang Xiang-ping, Li De-yu, Wang Su-ge. Rough set model based on formal concept analysis [J]. Information Sciences, 2013, 222: 611-625

[18] Wei Ling, Qi Jian-jun. Relation between concept lattice reduct and rough set reduct [J]. Knowledge-Based Systems, 2010, 23(8): 934-938

[19] Ganter B, Wille R. Formal Concept Analysis: Mathematical Foundations [M]. Berlin, Springer-Verlag, 1999

[20] 张文修，梁吉业，李德玉，等. 粗糙集理论与方法[M]. 北京：科学出版社，2004

[21] Leung Yee, Li De-yu. Maximal consistent block technique for rule acquisition in incomplete information systems [J]. Information Sciences, 2003, 153: 85-106