

文章编号:1001-9081(2009)02-0539-03

基于优势关系的信息系统与形式概念分析

王俊红^{1,2}, 梁吉业^{1,2}, 曲开社^{1,2}

(1. 计算智能与中文信息处理教育部重点实验室, 太原 030006; 2. 山西大学 计算机与信息技术学院, 太原 030006)
(cswjh@126.com)

摘要:对形式概念分析在基于优势关系的信息系统中的应用进行了讨论,证明了由基于优势关系的信息系统导出的形式背景中的对象粒与优势类是等价的,同时对对象粒协调集与优势协调集是等价的。

关键词:形式概念分析;概念格;信息系统;优势关系

中图分类号: TP18 **文献标志码:** A

Information system based on dominance relations and formal concept analysis

WANG Jun-hong^{1,2}, LIANG Ji-ye^{1,2}, QU Kai-she^{1,2}

(1. Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Taiyuan Shanxi 030006, China;
2. School of Computer and Information Technology, Shanxi University, Taiyuan Shanxi 030006, China)

Abstract: The implication of formal concept analysis in the information system based on dominance relations was discussed. The object granular in the formal context extract formed the information system based on dominance relations and the dominance class, and the consistent set of object granular and the consistent set of dominance class were proved to be equivalent.

Key words: formal concept analysis; concept lattice; information system; dominance relation

0 引言

信息系统中有很大一部分为连续值信息系统,这种信息系统中普遍蕴涵着序的特征,如大与小、高与矮、多与少、优与劣等,此时用的不是等价关系而是优势关系,研究基于优势关系的信息系统的知识获取具有重要的意义^[1-3]。

形式概念分析^[4]用于概念的发现、排序和显示。形式概念分析理论中的核心数据结构是概念格结构模型。概念格是根据数据集中对象与属性之间的二元关系建立的一种概念层次结构,生动简洁地体现了概念之间的泛化和特化关系,概念的外延被理解为属于这个概念的所有对象的集合,而内涵被认为是所有这些对象所共有的特征(或属性)集合,其相应的Hasse图则实现了对数据的可视化。因此,概念格被认为是进行数据分析的有力工具^[5-8]。

本文讨论了由基于优势关系的信息系统中导出形式背景的方法,证明了由基于优势关系的信息系统导出的形式背景中的对象粒与优势类是等价的,对象粒协调集与优势协调集是等价的,最后给出了一种基于概念格的约简方法。

1 基本概念

下面介绍基于优势关系的信息系统的一些基本概念^[9]。

一个基于优势关系的信息系统由一个四元组构成: $S = (U, A, V, f)$, 其中: $U = \{x_1, x_2, \dots, x_n\}$ 为对象集; $A = (a_1, a_2, \dots, a_n)$ 为属性集; $V = \cup V_a$ 为属性集值域,并且属性值 V_a

有偏好次序; $f: U \times A \rightarrow V$ 是一个信息函数,表示对每一个 $a \in A, x \in U, f(x, a) \in V_a$ 。

定义 1 设 $S = (U, A, V, f)$ 是一个基于优势关系的信息系统, $B \subseteq A$, 定义:

$$R_B^< = \{(x_i, x_j) \in U^2 \mid f(x_i, a) \leq f(x_j, a), (\forall a \in B)\}$$

$$R_B^> = \{(x_i, x_j) \in U^2 \mid f(x_i, a) \geq f(x_j, a), (\forall a \in B)\}$$

为 S 在属性集 B 上的优势关系和劣势关系。

定义 2 设 $S = (U, A, V, f)$ 是一个基于优势关系的信息系统, $B \subseteq A$, 定义:

$$[x_i]_B^< = \{x_j \mid (x_i, x_j) \in R_B^<\}$$

$$[x_i]_B^> = \{x_j \mid (x_i, x_j) \in R_B^>\}$$

$[x_i]_B^<$ 表示在属性集 B 条件下, 优于对象 x_i 的所有对象集合, 称为 x_i 的优势类; $[x_i]_B^>$ 表示在属性集 B 条件下, 劣于对象 x_i 的所有对象集合, 称为 x_i 的劣势类。

下面介绍形式概念分析的一些基本概念^[10]。

定义 3 形式背景是一个三元组 (U, D, I) , 其中, $U = \{x_1, x_2, \dots, x_n\}$ 是对象的集合; $D = (d_1, d_2, \dots, d_n)$ 是属性的集合; $I \subseteq U \times D$ 是 U 和 D 之间的二元关系, $(x, d) \in I$ 意味着对象 x 具有属性 d 。

在形式背景 (U, D, I) 中, 对于集合 $X \subseteq U$ 和 $B \subseteq D$ 定义:

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

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

对于 $x \in U$ 和 $d \in D$, 记 $x^* = \{x\}^*, d^* = \{d\}^*$ 。设 B

收稿日期: 2008-08-11; 修回日期: 2008-09-23。 基金项目: 国家自然科学基金资助项目(60773133); 国家 863 计划项目(2007AA01Z165); 高等学校博士学科点专项科研基金资助项目(20050108604); 教育部科学技术研究重点项目(206017); 山西省重点实验室开放基金资助项目(200603023); 山西省自然科学基金资助项目(2008011038, 2007011040)。

作者简介: 王俊红(1979-), 女, 山西曲沃人, 讲师, 博士研究生, 主要研究方向: 概念格、粗糙集; 梁吉业(1962-), 男, 山西晋城人, 教授, 博士生导师, 主要研究方向: 计算智能、数据挖掘、粒计算; 曲开社(1954-), 男, 山西运城人, 教授, 主要研究方向: 概念格、数据挖掘。

$\subseteq D, x^* \cap B$ 也记为 x^{*B} 。

定义 4 设 (U, D, I) 是一个形式背景, 对于 $X \subseteq U, B \subseteq D$, 如果二元组 (X, B) 满足 $X^* = B$ 并且 $B^* = X$, 则称 (X, Y) 是一个概念。其中 X 称为概念的外延, Y 称为概念的内涵。

形式背景 (U, D, I) 的形式概念全体构成一个完备格称为概念格, 表示为 $L(U, D, I)$, 其中概念格的交和并运算定义如下:

$$(X_1, B_1) \wedge (X_2, B_2) = (X_1 \cap X_2, (B_1 \cup B_2)^{**})$$

$$(X_1, B_1) \vee (X_2, B_2) = ((X_1 \cup X_2)^{**}, B_1 \cap B_2)$$

在概念格中, 由对象 $x \in U$ 所对应的概念 (x^{**}, x^*) 称为对象概念, 任意一个形式概念可以表示为它的外延的对象概念的并, 即:

$$(X, B) = \bigvee_{x \in X} (x^{**}, x^*)$$

所有的对象概念 $\{(x^{**}, x^*) \mid x \in U\}$ 的集合形成了概念格的基础, 即概念格中所有对象概念的集合反映了概念格结构的信息粒度。称基本外延集 $\{(x^{**}, x^*) \mid x \in U\}$ 是概念格的对象信息粒, 简称为对象粒。

定理 1^[11] 设 (U, D, I) 是一个形式背景, 则 $y \in x^{**}$ 的充要条件是 $y^* \supseteq x^*$ 。

2 基于优势关系的信息系统与形式概念分析

经典的概念格仅对 0 和 1 的形式背景进行研究, 而通常的信息系统的值域具有多值性, 因此要使用概念格对多值背景进行研究, 需要首先将多值背景转换为单值背景, 这种将多值背景转换为单值背景的处理过程称为概念定标^[10]。概念定标的基本思想是通过概念标尺从多值背景中导出单值背景。在概念定标的过程中首先通过另一个称为概念标尺的背景来对每个属性进行“翻译”, 选用不同的标尺会得到不同转换结果, 常用的标尺有额定标尺、顺序标尺、中序标尺、双序标尺等。对具有优势关系的偏序信息系统我们采用顺序标尺来处理。

下面给出基于优势关系的多值背景的概念标尺的形式化定义。

定义 5 设 $S = (U, A, V, f)$ 是一个基于优势关系的多值信息系统, 属性 $m \in A$ 的一个概念标尺是一个形式背景 $S_m = (U_m, D_m, I_m)$, 其中 $U_m = D_m = V_m$, 对于 $x \in U_m, a \in D_m$, 有 $xI_m a$ 当且仅当 $a \geq x$ 。

如设 $m \in M, V_m = \{1, 2, 3, 4\}$, 则属性 $m \in A$ 的概念标尺定义为:

概念标尺				
U	1	2	3	4
1	1	1	1	1
2	0	1	1	1
3	0	0	1	1
4	0	0	0	1

例 1 给出信息系统 (U, A, V, f) (见表 1), 其中 $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ 为对象集, $A = (a_1, a_2, a_3)$ 为属性集, $V_i = \{1, 2, 3\} (a_i \in A), f_i: U \rightarrow V_i, a_i \in A$ 为对象与属性之间

的关系集。通过 a_1, a_2, a_3 的概念标尺可导出形式背景如表 2 所示。

表 1 例 1 的信息系统

U	a_1	a_2	a_3
x_1	1	2	1
x_2	3	2	2
x_3	1	1	2
x_4	2	1	3
x_5	3	3	2
x_6	3	2	3

表 2 表 1 的形式背景

U	a_1			a_2			a_3		
	a_{11}	a_{12}	a_{13}	a_{21}	a_{22}	a_{23}	a_{31}	a_{32}	a_{33}
x_1	1	0	0	1	1	0	1	0	0
x_2	1	1	1	1	1	0	1	1	0
x_3	1	0	0	1	0	0	1	1	0
x_4	1	1	0	1	0	0	1	1	1
x_5	1	1	1	1	1	1	1	1	0
x_6	1	1	1	1	1	0	1	1	1

其中 $a_{ij}(i, j = 1, 2, 3)$ 表示 a_i 的第 j 个取值。

在转化后的形式背景中可以看出, 属性 a_{11}, a_{21}, a_{31} 在所有对象上的取值都为 1, 为了保证形式背景的正则性, 转化的形式背景中删除这些属性列, 即设 $D = \{a_{12}, a_{13}, a_{22}, a_{23}, a_{32}, a_{33}\}$, 同时将转化后的形式背景称为由基于优势关系的信息系统导出的形式背景。

定义 6 设 (U, D, I) 是一个由基于优势关系的信息系统导出的形式背景, $B \subseteq D$, 定义:

$$R_B^< = \{(x_i, x_j) \in U^2 \mid x_i^* \cap B \subseteq x_j^* \cap B\}$$

$$R_B^> = \{(x_i, x_j) \in U^2 \mid x_i^* \cap B \supseteq x_j^* \cap B\}$$

为 (U, A, I) 在属性集 B 上的优势关系和劣势关系。

若 $(x_i, x_j) \in R_B^<$, 则表示对象 x_j 在属性集 B 上优于对象 x_i ; 若 $(x_i, x_j) \in R_B^>$, 则表示对象 x_j 在属性集 B 上劣于对象 x_i 。

定义 7 设 (U, D, I) 是一个由基于优势关系的信息系统导出的形式背景, $B \subseteq D$, 定义:

$$[x_i]_B^< = \{x_j \mid (x_i, x_j) \in R_B^<\}$$

$$[x_i]_B^> = \{x_j \mid (x_i, x_j) \in R_B^>\}$$

$[x_i]_B^<$ 表示在属性集 B 条件下, 优于对象 x_i 的所有对象集合, 称为 x_i 的优势类; $[x_i]_B^>$ 表示在属性集 B 条件下, 劣于对象 x_i 的所有对象集合, 称为 x_i 的劣势类。

定理 2 设 $S = (U, A, V, f)$ 是一个基于优势关系的信息系统, (U, D, I) 是由其导出的基于优势关系的形式背景, 则对 $\forall x \in U$, 有 $x^{*D \cdot D} = [x]_A^<$ 。

证明 $\forall y \in [x]_A^<$, 由定义 6、7, 有 $y^{*D} \supseteq x^{*D}$, 再由定理 1, 则有 $y \in x^{*D \cdot D}$ 。因此 $[x]_A^< \subseteq x^{*D \cdot D}$ 。设 $y \in x^{*D \cdot D}$, 则由定理 1, 有 $y^{*D} \supseteq x^{*D}$, 则 $y \in [x]_A^<$ 。因此 $x^{*D \cdot D} \subseteq [x]_A^<$ 。

由定理 2 可以得出, 基于优势关系的信息系统中的优势类与相应形式背景中的对象粒是等价的。

由表 1 可以得到:

$$[x_1]_A^< = \{x_1, x_2, x_5, x_6\}$$

$$[x_2]_A^< = \{x_2, x_5, x_6\}$$

$$[x_3]_A^< = \{x_2, x_3, x_4, x_5, x_6\}$$

$$[x_4]_A^< = \{x_4, x_6\}$$

$$[x_5]_A^< = \{x_5\}$$

$$[x_6]_A^< = \{x_6\}$$

由表2可以得到:

$$x_1^{*D*D} = \{x_1, x_2, x_5, x_6\}$$

$$x_2^{*D*D} = \{x_2, x_5, x_6\}$$

$$x_3^{*D*D} = \{x_2, x_3, x_4, x_5, x_6\}$$

$$x_4^{*D*D} = \{x_4, x_6\}$$

$$x_5^{*D*D} = \{x_5\}$$

$$x_6^{*D*D} = \{x_6\}$$

$$\text{显然有 } x^{*D*D} = [x]_A^<.$$

在基于优势关系的信息系统中的多个属性并不是同等重要的,因而有必要研究信息冗余问题,建立优势与劣势关系上的属性约简方法,合理简化信息决策问题。

定义8^[9] 设 $S = (U, A, V, f)$ 是一个基于优势关系的信息系统,给出优势关系 $R_B^<$ 及劣势关系 $R_B^>$, 若 $R_B^< = R_A^<$ 或 $R_B^> = R_A^>$, 称 B 为优势协调集或劣势协调集。若 B 是优势协调集或劣势协调集,且 B 的任意真子集都不是优势协调集或劣势协调集,称 B 为优势关系或劣势关系下的属性约简集,简称优势约简集或劣势约简集。

同时可以证明,在一个基于优势关系的信息系统中,优势约简集和劣势约简集是等价的。

定义9 设形式背景 (U, D, I) 是由基于优势关系的信息系统 $S = (U, A, V, f)$ 导出的形式背景,对属性集 $B \subseteq D$, 如果对任意 $x \in U$ 都有 $x^{*B*B} = x^{*D*D}$, 则 B 被称为优势关系下的对象粒协调集。如果 B 的任意子集都不是优势关系下的对象粒协调集,则 B 称为该形式背景的优势关系下的对象粒约简。

定义10 设 (U, D, I) 是一个基于优势关系的形式背景,定义:

$$M^*(x, y) = y^{*D} - x^{*D}$$

称 $M^*(x, y)$ 为 (U, D, I) 的优势关系下的对象粒辨识属性集,称 $\{M^*(x, y) : (x, y) \in U \times U\}$ 为 (U, D, I) 的优势关系下的对象粒辨识矩阵。

推论1 在形式背景 (U, D, I) 中,设 $x_i, x_j \in U$, 若 $x_i \in [x_j]_D^<$, 则 $M^*(x_j, x_i) = \emptyset$ 。

对于表1的信息系统,可以计算它的优势约简集是 $\{a_2, a_3\}$, 同时根据定义10可以得到表2的形式背景的对象粒辨识矩阵。

表3 对象粒辨识矩阵

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	\emptyset	\emptyset	a_{22}	a_{22}	\emptyset	\emptyset
x_2	$a_{12} \ a_{13} \ a_{32}$	\emptyset	$a_{12} \ a_{13} \ a_{22}$	$a_{13} \ a_{22}$	\emptyset	\emptyset
x_3	a_{32}	\emptyset	\emptyset	\emptyset	\emptyset	\emptyset
x_4	$a_{12} \ a_{32} \ a_{33}$	$a_{12} \ a_{32} \ a_{33}$	$a_{12} \ a_{32} \ a_{33}$	\emptyset	a_{33}	\emptyset
x_5	$a_{12} \ a_{13} \ a_{23} \ a_{32}$	a_{23}	$a_{12} \ a_{13} \ a_{23} \ a_{22} \ a_{32}$	$a_{13} \ a_{22} \ a_{33}$	\emptyset	a_{23}
x_6	$a_{12} \ a_{13} \ a_{32} \ a_{33}$	a_{33}	$a_{12} \ a_{13} \ a_{22} \ a_{33}$	$a_{13} \ a_{22}$	a_{33}	\emptyset

(上接第538页)

- [11] CORTES C, VAPNIK V. Support vector networks[J]. Machine Learning, 1995, 20(3): 273–297.
- [12] CASTANE R, MARTI P, VELASCO M, et al. Resource management for control tasks based on the transient dynamics of closed-loop systems[C]// Proceedings of the 18th Euromicro Conference on Real-Time Systems (ECRTS06). Washington, DC: IEEE Computer Society, 2006: 171–182.
- [13] JIN HONG, WANG HUI, WANG HONGAN, et al. Optimization design of controller periods using evolution strategy[J]. LNAI, 2005,

取 $B = \{a_{22}, a_{23}, a_{32}, a_{33}\}$, 可以验证 B 是表2的优势关系下的对象粒协调集,也是该形式背景的优势关系下的对象粒约简。因此优势关系下的对象粒约简与优势约简集是等价的。

3 结语

基于优势关系的信息系统是一类在现实生活中普遍存在的信息系统,本文对形式概念分析在基于优势关系的信息系统中的应用进行了研究,证明了由基于优势关系的信息系统导出的形式背景中的对象粒与优势类是等价的,对象粒协调集与优势协调集是等价的,这些结论进一步丰富了形式概念分析理论,对基于优势关系的信息系统的知识获取的研究有着积极的意义。

参考文献:

- [1] POLKOWSKI L, SKOWRON A. Rough sets in knowledge discovery 1- Methodology and applications [M]. Wurzburg: Physica-Verlag, 1998.
- [2] KAZIMIERZ Z. Rough approximation of a preference relation by a multi-attribute dominance for deterministic, stochastic and fuzzy decision problems[J]. European Journal of Operational Research, 2004, 159(1): 196–206.
- [3] SALVATORE G, BENEDETTO M, ROMAN S. Rough sets methodology for sorting problems in predence of multiple attributes and criteria[J]. European Journal of Operational Research, 2002, 138(2): 247–259.
- [4] WILLE R. Restructuring lattice theory: an approach based on hierarchies of concepts[C]// Ordered Sets. Dordrecht: Reidel, 1982: 445–470.
- [5] 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.
- [6] 曲开社, 翟岩慧. 偏序集、包含度与形式概念分析[J]. 计算机学报, 2006, 29(2): 219–226.
- [7] 曲开社, 翟岩慧, 梁吉业, 等. 形式概念分析对粗糙集理论表示及扩展[J]. 软件学报, 2007, 18(9): 214–218.
- [8] 梁吉业, 王俊红. 基于概念格的规则产生集挖掘算法[J]. 计算机研究与发展, 2004, 41(8): 1339–1344.
- [9] 张文修, 仇国芳. 基于粗糙集的不确定决策[M]. 北京: 清华大学出版社, 2005.
- [10] GANTER B, WILLE R. Formal concept analysis: Mathematical foundations[M]. Berlin: Springer-Verlag, 1999.
- [11] 宋笑雪, 张文修. 形式概念分析与集值信息系统[J]. 计算机科学, 2007, 34(11): 129–131.
- [12] WU WEI ZHI, LEUNG Y, MI J S. Granular computing and knowledge reduction in formal contexts[C]// IEEE Transactions on Knowledge and Data Engineering. Washington, DC: IEEE Computer Society, 2008.
- [13] 3801(1): 1100–1105.
- [14] SETO D, LEHOCZKY J P, SHA L, et al. On task schedulability in real-time control systems[C] // Proceeding of 17th IEEE RTSS. Washington, DC: IEEE press, 1996: 13–21.
- [15] HWANG C, HONG D H, KYUNG H S. Support vector interval regression machine for crisp input and output data[J]. Fuzzy Sets and Systems, 2006, 157(8): 1114–1125.
- [16] ARZEM K J, WITTENMARK B. Computer controlled systems[M]. New Jersey, US: Prentice Hall, 1997.