

基于形式背景的属性转移与知识发现

郑书富 余高锋

(三明学院信息工程学院 福建 三明 365004)

摘要 概念格理论是一种有效的知识表示与知识发现工具,是知识表示、知识发现和知识获取的基础。利用形式背景信息熵与属性的重要性理论,讨论形式背景的属性知识转移的特点,得到了基于形式背景的属性转移原理,给出形式背景的知识发现与应用。

关键词 形式背景,概念格,信息熵,属性转移,知识发现

中图法分类号 TP181 **文献标识码** A

Attribute Transfer and Knowledge Discovery Based on Formal Context

ZHENG Shu-fu YU Gao-feng

(School of Information, Engineering San Ming University, Sanming, Fujian 365004, China)

Abstract The theory of concept lattice is an effective tool of knowledge representation and knowledge discovery, and is the basis of knowledge representation, knowledge discovery and knowledge acquisition. Based on the formal context information entropy and the importance of attribute theory, this paper discussed the characteristics of knowledge transfer of formal context attributes, obtained the attribute transfer principle based on formal context, and gave the knowledge discovery and application of formal context.

Keywords Formal context, Concept lattice, Information entropy, Attribute transfer, Knowledge discovery

知识转移是组织内或组织间跨越边界的知识共享,即知识以不同的方式在不同的组织或个体之间转移传播,强调知识转移不仅是知识的扩散,而是跨组织或个体边界的有目的、有计划的共享^[1]。国内外众多学者从知识转移的层次模型、知识转移的过程、知识转移的效果以及影响知识转移成败的因素等方面进行研究。从知识转移层次的角度来分析,无论知识在组织或个体间以何种形式实现知识转移,其核心就是个体间的知识转移,而知识往往由概念来表达。

概念格,又称为 Galois 格,是 Wille 于 1982 年首先提出的^[2]。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,其生动简洁地体现了概念之间的泛化和特化关系,已成为数据挖掘和知识处理的有力工具^[3-9]。知识是对象与属性的统一,对象集 $X \subseteq U$ 对应存在着属性集 $\alpha \subseteq R$ (这里 U 是非空有限集, R 是属性集),即若存在 $X \subseteq U$,则存在 $\alpha \subseteq R$ 与之对应;反之,若存在 $\alpha \subseteq R$,则必存在一个 $X \subseteq U$ 与之对应。显然,这是对知识的内涵与外延的认识,正如对同一知识概念的两种不同的定义形式。知识的对象描述易于知识的推理和计算,而知识的属性描述易于知识的理解和应用。本文利用概念格与粗集理论对知识转移中传递者与接受者的形式背景进行分析,引入知识概念的相似度、差异度等概念,结合形式背景的信息熵来研究属性知识的转移特点。

1 知识概念

概念格所分析的数据一般用形式背景 (Formal context) 来表示,形式背景的定义如下。

定义 1^[4] 称 $S = (U, R, V, f)$ 是一个形式背景。其中, U 是非空有限论域, R 是属性集, $V = \{0, 1\}$ 是属性值的集合, $f: U \times R \rightarrow V$ 是一个信息函数。对于任意 $a \in R$, 若 $f(x, a) = 1$, 表示对象 x 具有属性 a ; 反之, 表示对象 x 不具有属性 a 。

定义 2^[4] 设 U 是论域, R 是 U 上的属性集, 如果 $X \subseteq U$, 有 $F(X) = \{a \mid a \in R, \forall x \in X, f(x, a) = 1\}$, 则称 $F: U \rightarrow R$ 的属性映射。

定义 3^[4] 设 U 是论域, R 是 U 上的属性集, 如果 $\alpha \subseteq R$, 有 $F^*(\alpha) = \{x \mid x \in U, \forall a \in \alpha, f(x, a) = 1\}$, 则称 $F^*: U \rightarrow R$ 的对象映射。

定义 4^[4] 设 $S = (U, R, V, f)$ 是一个形式背景, $X \subseteq U$, 且 $\alpha \subseteq R$, 称 (X, α) 是形式背景的知识概念, 简称概念, 如果 $F(X) = \alpha$, 且 $F^*(\alpha) = X$ 。其中, X 称为概念的外延, α 称为概念的内涵。

形式背景 $S = (U, R, V, f)$ 的所有非空的知识概念的外延 (对象集), 记为:

$$J = \{X \mid X \subseteq U, F(X) = \alpha, F^*(\alpha) = X\}$$

命题 1 设 $S = (U, R, V, f)$ 是形式背景, 则形式背景中所有知识概念的外延是形式背景的一个覆盖。

本文受福建省中青年骨干教师教育科研项目(JA15461), 三明学院科技资助项目(B0702)资助。

郑书富(1971—), 男, 硕士, 副教授, 主要研究方向为粗系统理论与应用研究, E-mail: smxyzsf@126.com; 余高锋(1971—), 男, 博士, 讲师, 主要研究方向为决策分析和博弈论等。

例 1 给定形式背景 $S=(U,R,V,f)$, 如表 1 所列。

表 1 形式背景 $S=(U,R,V,f)$

U	a	b	c	d
x_1	1	0	0	1
x_2	1	0	1	0
x_3	1	0	0	1
x_4	1	1	1	0
x_5	1	1	1	0

显然, $(\{x_1, x_3\}, \{a, d\}), (\{x_2, x_4, x_5\}, \{a, c\}), (\{x_4, x_5\}, \{a, b, c\})$ 是形式背景中的可定义概念, 且所有可定义概念的外延是形式背景的一个覆盖。同时, $(\{x_4, x_5\}, \{a, b, c\})$ 是 $(\{x_2, x_4, x_5\}, \{a, c\})$ 的细化概念。在形式背景 $S=(U, R, V, f)$ 下的概念有 $[x_1]=\{x_1, x_3\}, [x_2]=\{x_2, x_4, x_5\}, [x_3]=\{x_1, x_3\}, [x_4]=[x_5]=\{x_4, x_5\}$ 。

若 (X, α) 是形式背景 $S=(U, R, V, f)$ 的知识概念, 则 X 与 α 是知识概念的两种不同的表达形式, 于是知识概念 (X, α) 又可称为知识 X 或知识 α 。

2 形式背景的知识转移

设 α 与 β 分别是传送者与接受者形式背景的知识概念, 而知识概念是由内涵来描述的。知识转移过程中传送者与接受者之间知识的相似度影响着知识转移的效率, 对知识的接受者来说, 评价和消化新知识的过程中, 与知识传送者有相似的知识是至关重要的, 缺乏这种知识往往会造成在知识转移过程中, 知识传送的效率降低及知识的缺损和丢失。

定义 5 知识 α 相对于知识 β 的知识相似度为:

$$Sim_{\beta}(\alpha, \beta) = \frac{|\alpha \cap \beta|}{|\beta|}$$

其中, $|\alpha|$ 表示知识概念 α 的属性(知识内涵)的个数, α 与 β 分别是对应传送者与接受者的知识。

知识的相似度是属性知识转移的基础。相似度越大, 说明知识越容易从传送者转移到接受者; 相似度越小, 说明知识越不容易从传送者转移到接受者。

定义 6 知识 α 相对于知识 β 的知识差异度为:

$$Div_{\beta}(\alpha, \beta) = \frac{|\alpha - \beta|}{|\beta|}$$

知识的差异度是属性知识转移量。差异度越大, 说明接受者有机会从传送者得到更多的知识; 差异度越小, 说明接受者从传送者得到的知识越少。同时, 传送者与接受者的知识相似度与差异度的大小还影响着知识转移的难度与效率。

命题 2 设 α 与 β 是相应的知识概念, 则知识 α 相对于知识 β 的知识相似度为 $0 \leq Sim_{\beta}(\alpha, \beta) \leq 1$ 。

命题 3 设 α 与 β 是相应的知识概念, 则知识 α 相对于知识 β 的知识差异度为 $Div_{\beta}(\alpha, \beta) \geq 0$ 。

命题 2 和命题 3 的结论是显然的。

3 形式背景的信息熵

为了研究属性知识的转移对丰富形式背景的概念的内涵的作用以及对概念细化的作用, 引进形式背景的信息熵与相对熵的概念。

定义 7^[6] 设 $S=(U, R, V, f)$ 是形式背景, 则称

$$IG(R) = \frac{1}{|U|} \sum_{x_i \in U} \frac{|[x_i]|}{|U|}$$

为 $S=(U, R, V, f)$ 形式背景概念的粒度。其中, $[x_i] \in J$ 。

定义 8^[6] 设 $S=(U, R, V, f)$ 是形式背景, 则称

$$IE(R) = \sum_{x_i \in U} \frac{1}{|U|} (1 - \frac{|[x_i]|}{|U|})$$

为 $S=(U, R, V, f)$ 形式背景的信息熵。

定义 9^[6] 设 $S=(U, R, V, f)$ 是形式背景, $P \subset R$, 且 $b \in R - P$, 则称

$$IE(b|P) = IE(P \cup \{b\}) - IE(P)$$

为形式背景的属性知识 b 相对于 P 的重要性。

命题 4 如果 $Div_{\beta}(\alpha, \beta) > 0$, 则 $IE(\beta|\alpha) \geq 0$ 。

证明: 由于 $Div_{\beta}(\alpha, \beta) > 0$, 因此 $|\alpha - \beta| \geq 1$, 即 $\beta \subset \beta \cup \alpha$ 。于是, 有 $IND(\beta \cup \alpha) \subseteq IND(\beta)$ 。由定义 7 和定义 8 可知, 形式背景的信息熵随着概念粒度的变小而单调增加, 因此有 $IE(\beta) \leq IE(\beta \cup \alpha)$, $IE(\beta|\alpha) \geq 0$ 。

4 属性知识转移与知识发现

知识转移就是知识由传送者向接受者的转移, 知识的相似度是指接受者与传送者之间某种程度上存在相同的知识体系, 是知识转移的基础。差异度决定了接受者所能得到的知识转移的量。知识转移的难度与效率与接受者的形式背景的信息熵有关, 对知识的接受者来说, 接受和消化新知识可能引起接受者的知识体系的重构, 即对接受者的形式背景的概念知识重新划分。因此, 对知识的接受者来说, 属性知识(知识内涵)的转移与知识接受者的信息熵相关。

命题 5 设 $r \in R_2$ 是传送者的属性知识, R_1 是接受者的知识, 若 $IE(r|R_1) = 0$, 则此知识最容易为接受者所接受, 由于接受者接受此属性知识, 因此不会引起知识的重构; 若 $IE(r|R_1) > 0$, 则此知识将引起接受者知识的重构。

证明: 若 $IE(r|R_1) = 0$, 由定义 8 可知, $IE(R_1 \cup \{r\}) = IE(R_1)$, 于是有 $IND(R_1 \cup \{r\}) = IND(R_1)$, 因此接受者接受此属性知识 r 不会引起知识的重构。

反之, 若 $IE(r|R_1) > 0$, 由定义 8 有 $IE(R_1 \cup \{r\}) > IE(R_1)$, 又由于 $R_1 \subset R_1 \cup \{r\}$, 因此有 $IND(R_1 \cup \{r\}) \subseteq IND(R_1)$, 故接受者接受此属性知识 r 将引起形式背景的概念知识重新划分, 即接受者的知识体系中的概念知识被重新刻画。

由命题 6 可知, 若 $IE(r|R_1) = 0$, 虽然属性知识 r 转移, 但不会引起知识系统的知识概念重新划分。从形式背景 $S=(U, R_1, V, f)$ 的知识概念角度来说, 属性 r 是拓广知识概念的内涵而出现的新知识。若 $IE(r|R_1) > 0$, 有 $IND(R_1 \cup \{r\}) \subseteq IND(R_1)$, 形式背景 $S=(U, R_1, V, f)$ 在属性集 $(R_1 \cup \{r\})$ 重新划分下, 发现新的知识概念, 进而发现新知识, 知识概念的内涵与外延同时发生变化。

基于形式背景信息熵的方法可以确定从传递者到接受者之间属性知识的转移原理: 遵循形式背景的属性重要性最小的属性知识最先转移。

属性的转移变化将引起对概念知识的重新刻画, 即拓展概念知识的内涵。对于知识的接受者来说, 先拓展知识的内涵, 再发生知识外延的拓广, 从而发现新知识。

5 应用

例1 给定知识接受者与传送者的形式背景分别是 $S_1 = (U_1, R_1, V, f)$ 与 $S_2 = (U_2, R_2, V, f)$, 如表2、表3所列。

表2 形式背景 $S_1 = (U_1, R_1, V, f)$

U_1	a	b	c
x_1	0	0	1
x_2	0	1	0
x_3	0	0	1
x_4	1	1	0
x_5	1	1	0
x_6	1	1	0

表3 形式背景 $S_2 = (U_2, R_2, V, f)$

U_2	a	b	d	e	f
x_1	0	0	1	1	0
x_2	0	1	1	1	0
x_3	0	0	1	1	1
x_4	1	1	1	0	1
x_5	1	1	1	0	1
x_6	1	1	1	1	1
y_1	1	0	1	0	1

知识接受者与传送者的知识相似度为:

$$Sim_{R_1}(R_2, R_1) = \frac{|R_2 \cap R_1|}{|R_1|} = \frac{2}{3}$$

知识差异度为:

$$Div_p(\alpha, \beta) = \frac{|R_2 - R_1|}{|R_1|} = \frac{2}{3}$$

传送者可传送的属性知识为 $R_2 - R_1 = \{d, e, f\}$, 接受者形式背景 $S_1 = (U_1, R_1, V, f)$ 的概念为: $[x_1] = \{x_1, x_3\}$, $[x_2] = \{x_2, x_4, x_5, x_6\}$, $[x_3] = \{x_1, x_3\}$, $[x_4] = [x_5] = [x_6] = \{x_4, x_5, x_6\}$ 。

于是,

$$IE(R_1) = \sum_{x_i \in U} \frac{1}{|U|} (1 - \frac{|[x_i]|}{|U|}) = \frac{1}{6} (\frac{2}{3} + \frac{1}{3} + \frac{2}{3} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2}) = \frac{19}{36}$$

形式背景 $S_1 = (U_1, R_1, V, f)$ 的属性集 R_1 更新为 $R_1 \cup \{d\}$ 时, 形式背景 $S_1 = (U_1, R_1 \cup \{d\}, V, f)$ 的概念为: $[x_1] = \{x_1, x_3\}$, $[x_2] = \{x_2, x_4, x_5, x_6\}$, $[x_3] = \{x_1, x_3\}$, $[x_4] = [x_5] = [x_6] = \{x_4, x_5, x_6\}$ 。

当属性集 R_1 更新为 $R_1 \cup \{e\}$ 时, 形式背景 $S_1 = (U_1, R_1 \cup \{e\}, V, f)$ 的概念为: $[x_1] = \{x_1, x_3\}$, $[x_2] = \{x_2, x_6\}$, $[x_3] = \{x_1, x_3\}$, $[x_4] = [x_5] = \{x_4, x_5, x_6\}$, $[x_6] = \{x_6\}$ 。

当属性集 R_1 更新为 $R_1 \cup \{f\}$ 时, 形式背景 $S_1 = (U_1, R_1 \cup \{f\}, V, f)$ 的概念为: $[x_1] = \{x_1, x_3\}$, $[x_2] = \{x_2, x_4, x_5, x_6\}$, $[x_3] = \{x_3\}$, $[x_4] = [x_5] = [x_6] = \{x_4, x_5, x_6\}$ 。

于是, $IE(R_1 \cup \{d\}) = \frac{19}{36}$, $IE(R_1 \cup \{e\}) = \frac{23}{36}$, $IE(R_1 \cup \{f\}) = \frac{20}{36}$, $IE(d|R_1) = IE(R_1 \cup \{d\}) - IE(R_1) = 0$, $IE(e|R_1) = IE(R_1 \cup \{e\}) - IE(R_1) = \frac{4}{36}$, $IE(f|R_1) = IE(R_1 \cup \{f\}) - IE(R_1) = \frac{1}{36}$ 。

由于 $IE(d|R_1) < IE(f|R_1) < IE(e|R_1)$, 因此属性知识

传递的顺序为: $\{d\} \rightarrow \{f\} \rightarrow \{e\}$ 。

知识接受者转移得到属性知识 $\{d\}$ 后, 并没有改变形式背景的划分, 但是丰富了知识概念的内涵, 从而得到了新知识, 如 $(\{x_1, x_3\}, \{c, d\})$ 。如果转移得到属性知识 $\{f\}$, 那么接受者不仅丰富了知识概念的内涵, 同时将知识概念 $(\{x_1, x_3\}, \{c\})$ 细化得到了一个新的知识概念 $(\{x_3\}, \{c, f\})$ 。

结束语 知识从传送者向接受者的转移主要是以概念知识为载体的, 知识转移的目的是使接受者在消化和吸收传送者的知识后能够发现新知识。形式背景的概念知识刻画是一个内涵与外延的统一, 外延的增加与减少往往引起知识内涵减少与增加。同时, 知识内涵的变化也必然会引起知识外延的变化。因此, 在知识转移过程中通过属性的转移, 可以使接受者对自身的概念知识重新刻画, 从而发现新知识。知识概念的研究是知识表示、知识发现和知识获取的基础, 知识概念将成为研究知识转移^[7]的新工具。若将本文的研究结果和 S-粗集的元素转移^[8-11]的特性纳入到知识转移的研究中, 将得到动态情况下的知识转移的新结果。

参 考 文 献

- [1] TEECE D. Technology transfer by multinational firms: the resource cost of transferring technological know-how [J]. The Economic Journal, 1997(87): 242-261.
- [2] RIVAL I. Ordered Sets [M]. Berlin: Reidel, 1982.
- [3] GANTER B, WILLE R. Formal Concept Analysis: Mathematical Foundations [M]. Berlin: Springer, 1999.
- [4] 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: 信息科学, 2005, 36(5): 628-639.
- [5] 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, 24(6): 4-6.
- [6] HUANG C C, LI J H, DIAS S M. Attribute significance, consistency measure and attribute reduction in formal concept analysis [J]. Neural Network World, 2016, 26(6): 607-623.
- [7] LI J H, MEI C L, LV Y J. Knowledge reduction in decision formal contexts [J]. Knowledge-Based Systems, 2001(24): 709-715.
- [8] PEI D, MI J S. Attribute reduction in decision formal context based on homomorphism [J]. International Journal of Machine Learning & Cybernetics, 2011, 2(4): 289-293.
- [9] LI L J, MI J S, XIE B. Attribute reduction based on maximal rules in decision formal context [J]. International Journal of Computational Intelligence Systems, 2014, 7(6): 1044-1053.
- [10] 汪应洛, 李勖. 知识的转移特性研究 [J]. 系统工程理论与实践, 2002, 22(10): 8-11.
- [11] 史开泉, 崔玉泉. S-粗集和它的一般结构 [J]. 山东大学学报 (理学版), 2002, 37(6): 471-474.
- [12] SHI K Q, CHANG T C. One direction S-rough sets [J]. International Journal of Fuzzy Mathematics, 2003, 11(2): 525-542.
- [13] SHI K Q. Two direction S-rough sets [J]. International Journal of Fuzzy Mathematics, 2004, 12(1): 178-181.
- [14] 史开泉. S-粗集和它的两类基本形式 [J]. 计算机科学, 2004, 31(10): 21-24.