

文章编号: 1007-791X (2014) 05-0388-06

大数据偏序结构生成原理

洪文学^{1,2,*}, 李少雄¹, 张涛³, 栾景民¹, 刘文远³

(1. 燕山大学 电气工程学院, 河北 秦皇岛 066004; 2. 东北大学秦皇岛分校 大数据可视化分析技术中心, 河北 秦皇岛 066004; 3. 燕山大学 信息科学与工程学院, 河北 秦皇岛 066004)

摘要: 形式概念分析是一种强有力的数据分析和可视化工具, 自提出至今已被应用到数据挖掘、知识发现等领域。但由于概念间复杂的关系使得概念格中的连线错综复杂, 尤其是在处理大的形式背景时显得尤为混乱。形式背景中所包含的属性间关系、对象间关系以及对象-属性关系是形式背景的本质关系, 本文在人类认知事物哲学原理的指导下, 构建了以描述属性间关系和区分对象为基本目的的偏序结构图, 并描述了其构建方法。该种图形层次分明、结构清晰, 不存在线的交叉连接, 具有比较好的可视化效果, 尤其计算方法简单, 存在应用到大数据的潜能, 为大数据数据分析和知识发现提供了一种潜在工具。

关键词: 形式背景; 偏序结构; 大数据; 属性偏序结构图; 对象偏序结构图

中图分类号: TP391.4 **文献标识码:** A **DOI:** 10.3969/j.issn.1007-791X.2014.05.003

0 引言

形式概念分析理论最初由德国数学家 Wille R 教授在 1982 年提出^[1], Ganter B 于 1999 年的学术著作中概括并总结了早期的形式概念知识框架及理论成果^[2]。形式概念分析用一个由一些对象组成的集合和由这些对象所共有的属性组成的集合组成的二元组来表示哲学意义上的概念。该理论建立在集合论及格应用理论基础上^[3-4]。近些年, 形式概念分析理论得到了比较广泛的应用, 包括: 知识发现^[5-7]、数据挖掘^[8]、软件工程^[9]、可视化^[10-11]、文献检索^[12]、在线分析^[13]等领域。但是形式概念分析理论高度抽象、严密的数学逻辑, 导致其数学门槛相对较高, 不利于其被更为广泛的应用。

大数据研究成为当今科学和社会发展的重大战略方向和必然趋势^[14]: 2008 年《Nature》上发表专刊讨论海量数据带来的挑战^[15]。2011 年《Science》上发表专刊, 联合其姊妹刊《Science Signaling》、《Science Translational Medicine》、《Science Careers》推出了“Dealing with data”专题,

讨论了与科研数据迅速增长有关的各种问题, 专题中的文章既强调了数据洪流所带来的挑战, 也强调了如果能够更好地组织和访问数据, 那么所能抓住和实现的机遇, 同时专题预见大数据合理利用对科学、经济和社会带来巨大推动作用^[16]。2012 年欧洲信息学与数学研究协会会刊《ERCIM News》发表 Big Data 专刊, 讨论了大数据时代的数据管理、数据密集型研究的创新技术等问题, 并介绍了欧洲科研机构开展的研究活动和取得的创新性进展^[17]。形式概念分析国际会议 (ICFCA 2006) 上公布的公开性问题中同样提到处理超大形式背景的问题。当形式背景数据量增大时, 形式概念的数量及概念间的关系量都可能是指数式增长, 必然会导致概念格无法清晰表达概念的层次结构及关系, 甚至无法在可接受时间内完成概念格的生成。

本文在研究了形式概念分析理论的优势和不足后, 从人类认知事物的哲学原理出发, 构建一种以发掘属性间关系和区分对象为基本目的新的数学结构, 以尝试解决上述形式概念分析面临的问题。

收稿日期: 2014-06-23 **基金项目:** 国家自然科学基金资助项目 (61273019, 61201111, 81273740, 81373767); 河北省自然科学基金资助项目 (F2013203368)

作者简介: *洪文学 (1953-), 男, 黑龙江依安人, 教授, 博士生导师, 主要研究方向为大数据偏序结构理论、复杂概念网络、混合数据信息融合与模式识别和中医工程学, Email: hongwx@ysu.edu.cn。

1 人类认知事物哲学的原理

众所周知,人类认知事物的过程由对象、属性和关系(包括对象-对象、属性-属性和对象-属性关系)3个基本要素构成。属性是各类事物特征的表达,属性间的关系表达了所研究问题的概念间关系。

图1表示了人类认知事物的哲学原理。图中包含对象和属性不同的观察尺度,对象尺度反映的是事物的相似性;属性尺度表示的是事物不同层次的特征。靠近原点表示含共性相对较多的对象和相对宏观的属性;远离原点的是含个性相对较多的对象和相对微观的属性。图1清晰地表明了人类认知事物的过程:从宏观共性中发现存在的普遍模式,从微观和个性中发现独有的模式。共有属性表现了事物之间共同拥有的属性,具有普遍性意义;个性对象凸显了有别于其他的独特对象,具有明显区分性。从普遍的共有属性中,发现事物的特异性;从共性对象中延伸出个性对象,是认知和区分事物本质的思维活动。

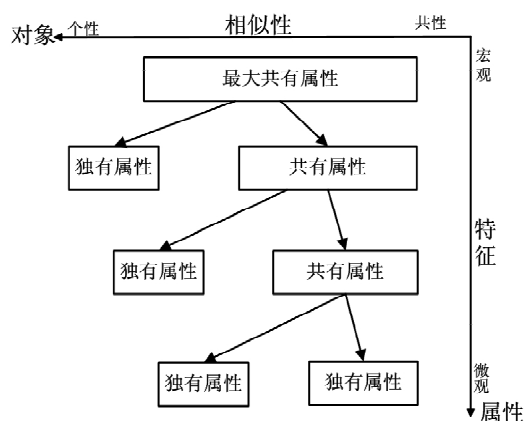


图1 人类认知事物的哲学原理示意图

Fig. 1 Philosophical principle of human cognition

从属性尺度上而言,最大共有属性是论域中所有对象都具有的属性,具有最为普遍的意义。独有属性是单独一个对象具有的属性,具有最为特异的性质。同理,从对象尺度上而言,“共有属性”对应对象具有较多的共同性质和较少的特异性质,而“独有属性”对应的对象具有明显的特异性。

2 偏序结构图

在人类认知事物的哲学思维的指导下可以构

建从两种不同角度上观察的偏序结构图形:属性偏序结构图(Attribute Partial Order Structure Diagram, APOSD)和对象偏序结构图(Object Partial Order Structure Diagram, OPOSD)。属性偏序结构图最大的优势在于发现事物的普遍性,而对象偏序结构图最大的优势在于发现事物的特异性。

其中,对于属性尺度而言,“最大共有属性”、“共有属性”和“独有属性”及其间关系是建立属性偏序结构图的3种基本元素和基本关系类型^[18];对象尺度与属性尺度为正交的尺度,对象偏序结构图可以由原形式背景转置后,以属性偏序结构图相同的构建方法获得,因此下文只介绍属性偏序结构图的构建方法。

“最大共有属性”、“共有属性”和“独有属性”及其间关系的计算成为构图的关键所在,这3种基本元素和基本关系的计算方法,以及基于此的属性偏序结构图构建方法如下描述。

2.1 特征属性的计算

特征属性的计算是构建偏序结构图的最基本运算,下面给出最大共有属性、共有属性和独有属性的计算方法。

给定一个形式背景 $K=(U, M, I)$,其中对象集合为 $U=\{u_1, u_2, \dots, u_n\}$, $n \geq 2$, 属性集合为 $M=\{m_1, m_2, \dots, m_k\}$, $k \geq 0$ 。

1) 计算最大共有属性。

根据最大共有属性的定义,假设存在属性 m_{o_i} ($i=1, 2, \dots, n$; $n \geq 1$)使得 $g(m_{o_i})=U$,那么, m_{o_i} 可称为最大共有属性,也就是说,形式背景中,可能存在多个最大共有属性。

2) 计算共有属性。

在属性偏序结构图中,每一个节点可以用一个由属性集合和对应对象集合组成的二元组表示。属性偏序结构图中的属性节点具有一定的相对性,共有属性和独有属性节点是从上一层父节点中衍生出来的,因此共有属性和独有属性的计算也是依图而计算。

假设上层的共有属性为 M_j ,且 $g(M_j)=U_j$,此时偏序图中的已有二元组组成的集合为 Ac_j 。用 M_j 与 $M=\{m_1, m_2, \dots, m_k\}$ 中的每个属性元素求并集得到

$M_j \cup \{m_i\}$ 作为新的二元组的属性集合, 并计算对应的对象集合 $U_j = g(M_j \cup \{m_i\})$ 组成二元组 $C_{ji} = (M_j \cup \{m_i\}, U_j)$ 。

这里约定, 属性集合求并时, 新元素追加到 M_j 到的末尾。

用这样的方法可以计算 M_j 下的二元组集合 $\{C_{j1}, C_{j2}, \dots, C_{jn}\}$ 。

共有属性: 假设 $C_{jp} = (M_{jp}, U_{jp})$ 是 M_j 下的一个二元组, 即 $C_{jp} \in \{C_{j1}, C_{j2}, \dots, C_{jn}\}$ 。如果 C_{jp} 满足下述 3 个条件:

a) $C_{jp} \notin Ac_j$;

b) $|U_{jp}| \geq 2$;

c) 所有与 C_{jp} 同层的二元组组成的集合为 $\{C_{j-m}, C_{j-m+1}, \dots, C_{j-1}, C_{j+1}, \dots, C_{j+n-1}, C_{j+n}\}$, $(m, n \geq 0)$, 其中任意一个同层二元组 C_{j+k} 可表示为 $C_{j+k} = (M_{j+k}, U_{j+k})$ 。这时, 如果 C_{jp} 的对象集合 U_{jp} 满足: $U_{jp} \not\subseteq \{U_{j-m}, U_{j-m+1}, \dots, U_{j-1}, U_{j+1}, \dots, U_{j+n-1}, U_{j+n}\}$, $(m, n \geq 0)$ 。

那么, 二元组 C_{jp} 的属性集合 M_{jp} 就可作 M_j 下的一个共有属性。

3) 计算独有属性。

假设 $C_{jp} = (M_{jp}, U_{jp})$ 是 M_j 下的一个二元组, 即 $C_{jp} \in \{C_{j1}, C_{j2}, \dots, C_{jn}\}$ 。如果 C_{jp} 满足下述 3 个条件:

a) $C_{jp} \notin Ac_j$;

b) $|U_{jp}| = 1$;

c) 所有与 C_{jp} 同层的二元组组成的集合为 $\{C_{j-m}, C_{j-m+1}, \dots, C_{j-1}, C_{j+1}, \dots, C_{j+n-1}, C_{j+n}\}$, $(m, n \geq 0)$, 其中任意一个同层二元组 C_{j+k} 可表示为 $C_{j+k} = (M_{j+k}, U_{j+k})$ 。这时, C_{jp} 的对象集合 U_{jp} 满足: $U_{jp} \not\subseteq \{U_{j-m}, U_{j-m+1}, \dots, U_{j-1}, U_{j+1}, \dots, U_{j+n-1}, U_{j+n}\}$, $(m, n \geq 0)$ 。

那么 C_{jp} 的属性集合 M_{jp} 就可作为 M_j 下的一个独有属性。

2.2 属性偏序结构图构建方法

根据上述的基本原理, 构建属性偏序结构图的方法步骤可描述如下:

Step1: 计算最大共有属性 M_0 及其节点 C_0 。共

有属性间直接建立纵向连线。

Step2: 在最大共有属性 M_0 下计算所有共有属性集合 $\{m_{11}, m_{12}, \dots, m_{1p}\}$ 和独有属性集合 $\{s_{11}, s_{12}, \dots, s_{1q}\}$ 及其对应二元组 $\{C_{m1}, C_{m2}, \dots, C_{mp}\}$ 和 $\{C_{s1}, C_{s2}, \dots, C_{sq}\}$, 并与最大共有属性的节点 C_0 间绘制有向边。

如果该层有序对节点与最大共有属性的节点间满足关系:

$$\bigcup_{i=1}^p g(m_{1i}) \cup \bigcup_{i=1}^q g(s_{1i}) = U, \quad (1)$$

则进入下一步; 不满足上式时, 在最底层节点与最大共有属性的节点间存在 k 条有向边。其中,

$$k = |U - \bigcup_{i=1}^p g(m_{1i}) \cup \bigcup_{i=1}^q g(s_{1i})|. \quad (2)$$

Step 3: 在上层每个共有属性节点 $\{C_{m1}, C_{m2}, \dots, C_{mp}\}$ 下, 计算上层共有属性 m_{t+1} , $t \in (1: p)$ (t 表示上一层的层数) 的共有属性 $\{m_{t+1,1}, m_{t+1,2}, \dots, m_{t+1,l}\}$ 和独有属性 $\{s_{t+1,1}, s_{t+1,2}, \dots, s_{t+1,k}\}$ 及其对应二元组节点 $\{C_{mt+1,1}, C_{mt+1,2}, \dots, C_{mt+1,l}\}$ 和 $\{C_{st+1,1}, C_{st+1,2}, \dots, C_{st+1,k}\}$, 每个独有属性的二元组节点下 $\{C_{s1}, C_{s2}, \dots, C_{sq}\}$ 计算独有属性 $\{sc_{t+1,1}, sc_{t+1,2}, \dots, sc_{t+1,q}\}$ 及其二元组节点 $\{Cs_{t+1}, Cs_{t+2}, \dots, Cs_{t+q}\}$ 。

如果上层节点 $\{C_{m1}, C_{m2}, \dots, C_{mp}\}$ 与其共有属性节点 $\{m_{t+1,1}, m_{t+1,2}, \dots, m_{t+1,l}\}$ 和独有属性节点 $\{C_{s1}, C_{s2}, \dots, C_{sk}\}$ 间满足关系:

$$\bigcup_{i=1}^p g(m_{t+1,i}) \cup \bigcup_{i=1}^q g(s_{t+1,i}) = g(m_j), \quad j \in (1: p), \quad (3)$$

则进入下一步; 不满足式 (3) 时, 在最底层节点与相应共有属性 m_j 的节点间绘制 l 条有向边,

$$l = |g(m_j) - \bigcup_{i=1}^p g(m_{t+1,i}) \cup \bigcup_{i=1}^q g(s_{t+1,i})|, \quad j \in (1: p). \quad (4)$$

Step 4: 当在共有属性或独有属性下不能生成新的二元组节点时则停止, 否则调至 Step 3。

Step 5: 算法结束。

上述算法的 Pseudo code 形式可表示为:

Algorithm: 构建属性偏序结构图

Input: 形式背景;

Output: 属性偏序结构图;

- 1 计算最大共有属性 M_0 , 及对应的二元组 C_0 ;
- 2 在 M_0 下计算 M_i, S_i , 及对应 C_{mi}, C_{si} ;
- 3 $C_i = C_{mi} + C_{si}$, 建立与 C_0 间的边;
- 4 If $\bigcup_{i=1}^p g(m_i) \cup \bigcup_{i=1}^q g(s_i) \neq U$;
- 5 在 C_0 与最底层点间建立 $|U - \bigcup_{i=1}^p g(m_i) \cup \bigcup_{i=1}^q g(s_i)|$ 条边;
- 6 Else
- 7 当前层索引 $t=t+1$;
- 8 While $C_t \neq \emptyset$
- 9 共有属性节点 $C_{mx} \in C_t$, 在 C_{mx} 下计算 M_{t+1}, S_{t+1} , 及对应 $C_{m,t+1}, C_{s,t+1}$;
- 10 独有属性节点 $C_{sx} \in C_t$, 在 C_{sx} 下计算 S_{t+1} 及对应 $C_{s,t+1}$;
- 11 $C_t = C_{m,t+1} + C_{s,t+1}$, 建立与 C_0 间的边;
- 12 If $\bigcup_{i=1}^p g(m_{t+1,i}) \cup \bigcup_{i=1}^q g(s_{t+1,i}) \neq g(m_j), j \in (1: p)$
- 13 在 C_t 最底层点间建立 $|g(m_j) - \bigcup_{i=1}^p g(m_{t+1,i}) \cup \bigcup_{i=1}^q g(s_{t+1,i})|, j \in (1: p)$ 条边;
- 14 End if
- 15 $t++$;
- 16 End while
- 17 End if

2.3 偏序结构图生成过程

利用属性特征定义, 从形式背景上建立属性偏序结构图的算法。下面应用文献 [19] 中的一个形式背景 (如表 1 所示), 简要说明从形式背景构建属性偏序结构图的具体过程。

表 1 一个形式背景

Tab. 1 A formal context

	a	b	c	d	e
1	x		x		
2	x	x	x		x
3		x			x
4			x	x	x

Step1: 计算最大共有属性 $m_0 = \emptyset$, 及其对应节点 $C_0 = (\emptyset, 1234)$, 位于第 0 层;

Step2: 计算 m_0 下的共有属性得到 $\{c, e\}$, 对应节点 $\{(c, 124), (e, 234)\}$ 位于第 1 层; 没有独有属性;

Step3: 在 $(c, 124)$ 下计算共有属性得到 $\{ce, ac\}$, 对应节点 $\{(ce, 24), (ac, 12)\}$ 位于第 2 层, 没有独有属性; 在 $(e, 234)$ 下计算共有属性得到 $\{ce, be\}$, 但 ce 已经存在, 此时只保留 be , 对应节点 $\{be, 23\}$ 位于第 2 层, 没有独有属性;

Step4: 在 $(ce, 24)$ 计算共有属性为 $\{\emptyset\}$, 独有属性为 $\{ace, cde\}$, 对应节点 $\{(ace, 2), (cde, 4)\}$, 位于第 3 层;

在 $(ac, 12)$ 计算共有属性为 $\{\emptyset\}$, 独有属性为 $\{ace, abc\}$, 但 ace 已存在, 且 $g(abc) = g(ace)$, 有重复对象存在, 此时该节点下无独有属性节点;

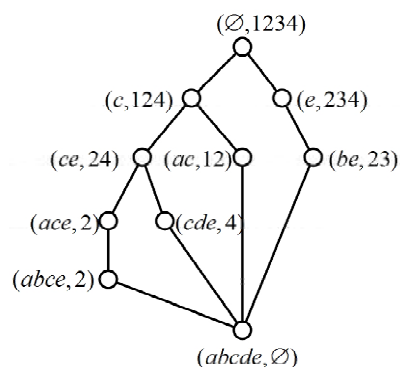
同理 $(be, 23)$ 下也无共有属性和独有属性节点;

但由于 $|g(ace) + g(cde)| < |g(ce) + g(ac) + g(be)|$, 在 $(ac, 12)$ 节点下建立与最底层节点间 $|g(ac) - (g(ace) + g(cde))| = |\{1, 2\} - \{2, 4\}| = |\{1\}| = 1$ 条边; 在 $(be, 23)$ 下建立与最底层节点 $|g(be) - (g(ace) + g(cde))| = |\{2, 3\} - \{2, 4\}| = |\{3\}| = 1$ 条边;

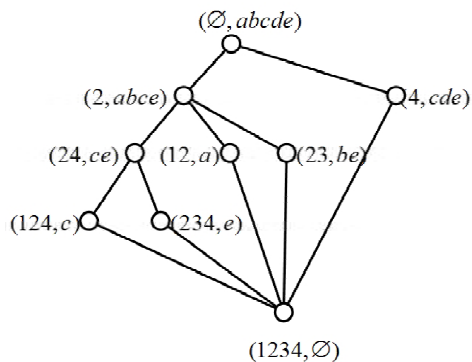
Step5: 在 $(ace, 2)$ 下计算独有属性得到 $abce$, 对应节点 $(abce, 2)$; 在 $(cde, 4)$ 下计算独有属性得到 $\{\emptyset\}$ 。但由于 $|g(abce)| < |g(ace) + g(cde)|$, 在 $(cde, 4)$ 与最底层节点间建立 1 条边;

Step6: 计算 $(abce, 2)$ 下独有属性为 $\{\emptyset\}$, 与最底层节点间建立 1 条边。

至此构建结束, 对应的属性偏序结构图如图 2 (a) 所示。将原形式背景转置后以同样的方法计算, 可以构建对象偏序结构图, 如图 2 (b)。



(a) 属性偏序结构图



(b) 对象偏序结构图

图 2 表 1 形式背景的属性偏序结构图 and 对象偏序结构图
Fig. 2 The attribute partial order structure diagram and object partial order diagram of context of Tab. 1

3 偏序结构图应用实例

一个例子来自于 Formal Concept Analysis 主页的经典形式背景: Live in water, 如表 2 所示。其中各对象符号代表的内容为: 1: fish leech, 2: bream, 3: frog, 4: water weeds, 5: reed, 6: bean, 7: corn; 各属性符号代表的内容为: *a*: needs water to live, *b*: lives in water, *c*: lives on land, *d*: needs chlorophyll, *e*: dicotyledon, *f*: monocotyledon, *g*: can move, *h*: has limbs, *i*: breast feeds。构建此形式背景的概念格、属性偏序结构图和对象偏序结构图, 如图 3 所示。

表 2 Live in water 的形式背景

Tab. 2 Formal context of Live in water

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	<i>i</i>
1	×	×					×		
2	×	×					×	×	
3	×	×	×				×	×	
4	×		×				×	×	×
5	×	×		×		×			
6	×	×	×	×		×			
7	×		×	×	×				
8	×		×	×		×			

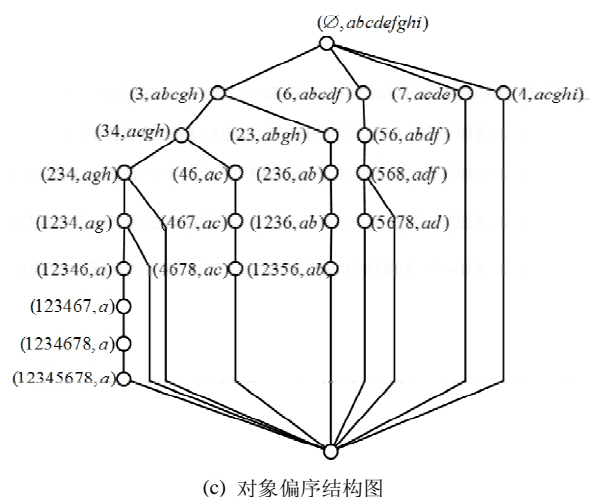
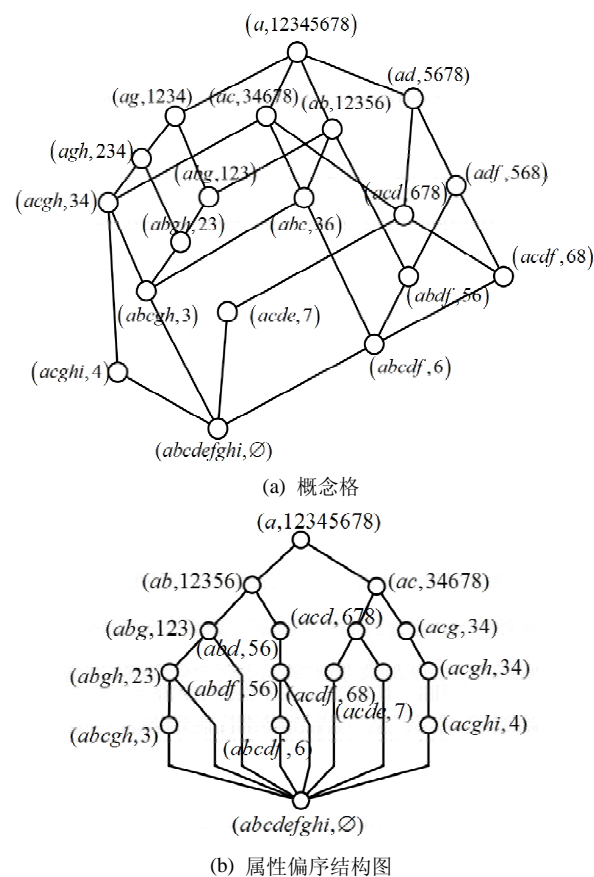


图 3 Live in water 的概念格、属性偏序结构图和对象偏序结构图

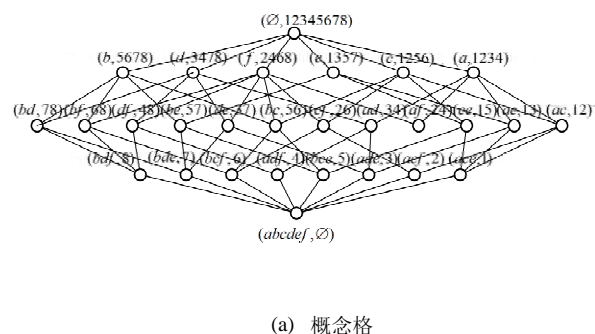
Fig. 3 The concept lattice, attribute partial order structure diagram and object partial order structure of Live in water

另一个典型的形式背景: Bodies of water 如表 3 所示, 其中各对象符号代表的内容为: 1: Fluss, 2: Bach, 3: Kanal, 4: Graben, 5: See, 6: Tumpel, 7: Teich, 8: Becken。各属性符号代表的内容为: *a*: fließend, *b*: stehend, *c*: natürlich, *d*: künstlich, *e*: groß, *f*: klein。使用同样的方法可以得到对应的概念格、属性偏序结构图和对象偏序结构图, 如图 4 所示。

表 3 Bodies of water 的形式背景

Tab. 3 Bodies of water formal context

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>
1	×		×		×	
2	×		×			×
3	×			×	×	
4	×			×		×
5		×	×		×	
6		×	×			×
7		×		×	×	
8		×		×		×



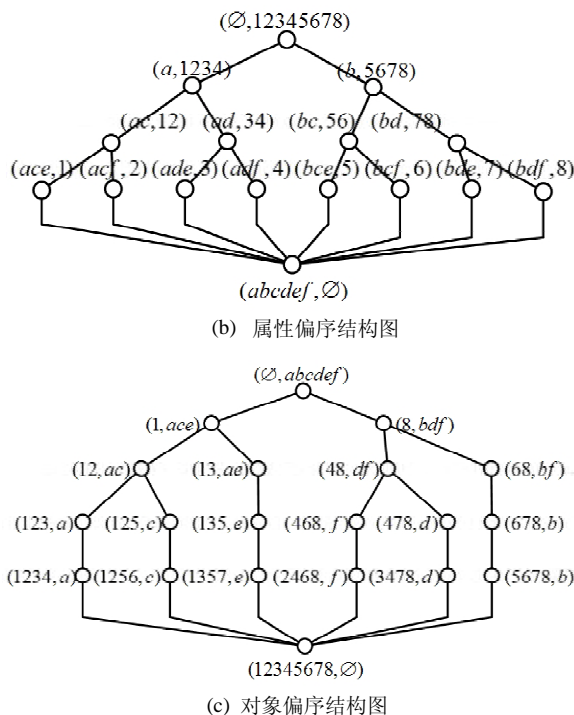


图4 Bodies of water 的概念格、属性偏序结构图和对象偏序结构图

Fig. 4 The Concept lattice, attribute partial order structure diagram and object partial order structure diagram of bodies of water

4 结束语

本文从分析形式概念分析的优势和不足入手,针对大数据应用的必然趋势,以人类认知事物的哲学原理为指导,提出了一种以表达属性间关系和区分对象为基本目的的属性偏序结构图和对象偏序结构图。以“最大共有属性”、“共有属性”和“独有属性”的计算为基础,详细描述了偏序结构图的构建方法。同时以两个国际公认的经典形式背景为实例,展示了其对应的属性偏序结构图和对象偏序结构图,并与相应的概念格做比较,通过对比可以发现偏序结构图所具有的层次分明、结构图清晰、节点间关系简要明了以及连线无交叉的优点。同时,也显示出这种偏序结构图应用到大数据数据分析的潜能,为大数据数据分析、知识发现提供了一种新的工具。

参考文献

[1] Wille Rudolf. Restructuring lattice theory: an approach based on hierarchies of concepts [C] // Proceedings of the NATO Advanced

Study Institute, Banff, Canada, 1982: 445-470.

- [2] Ganter B, Stumme G, Wille R. Formal concept analysis: foundations and applications [M]. Berlin Heidelberg: Springer-Verlag, 2005.
- [3] Wolff K E. A first course in formal concept analysis [C] // the 7th Conference on the Scientific Use of Statistical Software, Heidelberg, 1993: 429-438.
- [4] Ganter B, Wille R. Formal concept analysis: mathematical foundations [M]. Berlin: Springer-Verlag, 1999.
- [5] 翟岩慧. 形式概念分析中的知识表示与知识获取 [D]. 太原: 山西大学, 2006.
- [6] 杨淑群, 蔡声镇, 丁树良, 等. 基于 FCA 具有认知诊断功能 CAT 的设计与实现 [J]. 南京航空航天大学学报, 2008,40 (5): 696-701.
- [7] 刘旭龙, 洪文学, 刘杰民, 等. 基于红外热成像与形式概念分析的面瘫病情客观评估方法 [J]. 光谱学与光谱分析, 2014,34 (4): 932-936.
- [8] 罗忠诚, 张志强, 王忠浩, 等. 语义背景下基于形式概念分析的产品族组件规划与类型识别研究 [J]. 中国机械工程, 2007,18 (22): 2729-2733.
- [9] 蒋平, 任胜兵, 林鹏, 等. 形式概念分析在软件工程中的应用 [J]. 计算机技术与发展, 2008,18 (4): 127-129,213.
- [10] Tom Huysegoms, Monique Snoeck, Guido Dedene, et al.. Visualizing variability management in requirements engineering through formal concept analysis [J]. Procedia Technology, 2013,9: 189-199.
- [11] 刘旭龙, 洪文学, 张涛, 等. 基于形式概念分析的中医辨证可视化方法 [J]. 燕山大学学报, 2010,34 (2): 162-164,168.
- [12] 刘兆庆, 伏玉琛, 凌兴宏, 等. 基于形式概念分析的博客社区发现 [J]. 计算机应用, 2013,33 (1): 189-191,198.
- [13] Korobko A V, Penkova T G. On-line analytical processing based on formal concept analysis [J]. Procedia Computer Science, 2012,1 (1): 2311-2317.
- [14] 李国杰, 程学旗. 大数据研究: 未来科技及经济社会发展的重大战略领域 [J]. 中国科学院院刊, 2012,27 (6): 647-657.
- [15] Doctorow C. Big data: Welcome to the petacenter [J]. Nature, 2008,455 (7209): 16-21.
- [16] Reichman O J, Jones G A, Bony S, et al.. Challenges and opportunities of open data in ecology [J]. Science, 2011,331 (6018): 703-705.
- [17] Stefan Manegold, Martin Kersten, Costantino Thanos, et al.. Special theme: Big Data [J]. ERCIM News, 2012, (89): 1-55.
- [18] 洪文学, 栾景民, 张涛, 等. 形式背景中属性特征与对象特征的完备定义 [J]. 燕山大学学报, 2014,38 (5): 381-387.
- [19] 马垣, 曾子维, 迟呈英, 等. 形式概念分析及其新进展 [M]. 北京: 科学出版社, 2011.

(下转第 402 页)

[13] Hong Wenxue, Li Shaoxiong, Yu Jianping, et al.. A new approach of generation of structural partial-ordered attribute diagram [J].

ICIC Express Letters, Part B: Applications, 2012,3 (4): 823-830.

A new method for knowledge discovery based on partial ordered structure theory

HONG Wen-xue^{1,2}, LUAN Jing-min¹, ZHANG Tao³, LI Shao-xiong¹, YAN En-liang¹

(1. College of Electrical Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China; 2. Big Data Visualization Technology Center, Northeastern University at Qinhuangdao, Qinhuangdao, Hebei 066004, China; 3. College of Information Science and Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China)

Abstract: Knowledge discovery in big data is one of key problems for big data application. From the view of highly abstract of cognition, a generalized method of knowledge discovery based on the theory of attribute partial ordered structure diagram (APOSOD) and object partial ordered structure diagram (OPOSOD) is proposed in this paper. Then, the analysis and compare has been discussed for structural relation of data features in aspects of structure of groups, structure of subgroups, branches and nodes of partial ordered structure diagram. In APOSOD, objects with certain characteristics have congregated into several clusters, which express the common attributes of the objects. Using the unique attribute of objects, it can distinguish specific object from the others quickly and effectively. Finally, taking the compatibility of traditional Chinese medicine problem as an example, compatibility rules for prescription of apoplexy are analyzed, and application of knowledge discovery in study of compatibility of traditional Chinese medicine is discussed in this paper, and the method is proved to be effective and effective and practical by the pre-experiment results.

Key words: knowledge discovery; partial ordered structure theory; attribute partial ordered; object partial ordered

(上接第 393 页)

Generation principle of partial ordered structure towards big data

HONG Wen-xue^{1,2}, LI Shao-xiong¹, ZHANG Tao³, LUAN Jing-min¹, LIU Wen-yuan³

(1. College of Electrical Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China; 2. Big Data Visualization Technology Center, Northeastern University at Qinhuangdao, Qinhuangdao, Hebei 066004, China; 3. College of Information Science and Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China)

Abstract: Formal concept analysis is a powerful tool in data analysis and visualization, and has been applied to data mining, knowledge discovery and many other fields since proposed. However, in the concept lattice, the complex relations between concepts make the lines rather complicated and crossed, especially when dealing with a large-scale formal context. The relation among attributes, objects and attribute-object are the essential relations in a formal context. Therefore, under the guidance in the philosophical principle of human being's cognition, the partial ordered structure diagram aiming to delineate the relations among attributes and distinguish distinctive objects is proposed, and construction method is described. Its distinct hierarchy, clear structure, uncrossed lines provide a better visualization. Apart from that, simple computational method of it makes a large potential in allusion to big data. Hence, a novel and efficient tool towards data mining and knowledge discovery of big data is provided by this diagram.

Key words: formal context; partial ordered structure; big data; attribute partial ordered structure diagram; object partial ordered structure diagram