文章编号：1007-791X (2014) 05-0394-09

基于偏序结构理论的知识发现方法

洪文学[1,2,*]，栾景民[1]，张 涛[3]，李少雄[1]，闫恩亮[1]

（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：大数据中的知识发现是大数据应用中的核心热点。本文从高度抽象认知事物视角出发，以表征事物普遍性为特征的概念驱动与表征事物特异性为特征的数据驱动两种方法学为哲学原理，提出了基于属性偏序结构图和对象偏序结构图的知识发现方法。分别从群结构、子群结构、支路、节点等角度对数据特征之间的结构关系进行讨论分析。属性偏序结构图将数据中具有某些共同特征的对象聚类到一起，是数据共性的表达；对象偏序结构图中，通过数据的独有属性可以快速有效的将特异性对象区分于其他对象。最后，以中医药方剂配伍研究问题为例，对张锡纯治疗中风的 32 个处方进行数据挖掘和知识发现，证明了该方法的有效性和实用性，为大数据知识发现研究提供了新的思路和方法。

关键词：知识发现；偏序结构理论；属性偏序；对象偏序

中图分类号：TP391.4 **文献标识码**：A DOI：10.3969/j.issn.1007-791X.2014.05.004

0 引言

随着信息科学技术的高速发展，无处不在的信息感知和采集终端为我们带来了大数据时代，其中如何从大数据中分析挖掘对决策有价值的对象和规则，就成为了目前科技与商业中急需解决的一个重要课题[1]。

大数据本身只是复杂而几乎无序的数据，而其背后的关联性，是大数据巨大的价值所在。大数据研究的重中之重，就是从海量的数据中提取相关内容，形成有意义、有价值的信息。大数据提出的科学挑战问题之一就是重点研究数据背后的关系网络。观察各种复杂系统得到的大数据，直接反映的往往是个体和个别链接的特性。反映相互关系的网络的整体特征隐藏在大数据中，数据科学的主要任务就是搞清楚数据背后的“关系网络”。如何表述和分析关系网络也成为了数据背后的共性问题。大数据关系网络研究的本质是大数据知识发现研究的基础，大数据应用最终目标是发现大数据的价值。本文以偏序结构理论为核心，开展了大数据结构与关系发现研究，为偏序结构理论在大数据知识发现研究中的应用提供了理论基础[2]。

1 偏序结构理论

德国的数学家 Wille 教授于 20 世纪 80 年代初提出了形式概念分析，主要研究“概念”和“概念分层”的数学化描述，其主要思想是：从被表示为形势背景的数据中获取形式概念以及形式概念之间的联系，形成一种以形式概念为元素的格结构—概念格[3-4]。近十几年来，随着形式概念分析理论的不断发展和完善，该方法已在知识发现、信息检索和软件工程等领域得到了广泛的应用[5]。从数据挖掘和知识发现的角度，形式概念可看作特定知识背景中基本知识单元的形式化。形式概念分析能够有效地实现对数据中所隐含的基本知识单元的提取。从理论上讲，运用形式概念分析理论，结合知识库、数据库等相关背景，从大数据中抽取出有用体系知识，如概念等，是有效可行的，主要优点在于可以将数据中的内在逻辑和组织结构完整地图示化，从而为分析概念数据之间的关联提供系统的

收稿日期：2014-06-13 **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、混合数据信息融合与模式识别、复杂概念网络和中医工程学，Email: hongwx@ysu.edu.cn。

可视化工具。

然而，在实际的大数据研究应用中，随着形式背景中对象数量和属性数量的增加，其概念生成的计算复杂度呈现指数级的增长[6]，生成的概念格结构也会随着数据量的增大而变得特别复杂，概念节点之间的关联关系不容易发现，概念层次关系不清楚，其可读性和可理解性大大降低[7]。此外，由于形式概念分析理论的数学门槛较高，很难被非数学专业的研究者理解和应用，影响了形式概念分析在大数据背景下的应用。

燕山大学的洪文学教授及其团队，在多项国家自然科学基金的支持下，开展了属性偏序结构理论的研究。在形式概念分析的基础上，研究了形式背景中属性的属性特征和性质，提出了以属性覆盖对象程度为偏序关系的属性偏序结构图理论，并以此为基础，研究了基于偏序结构图的数据挖掘和知识发现方法，并在中医诊断模式分类、中药方剂配伍、英语语义排歧等领域得到了很好的应用[8-13]。

1.1　属性偏序结构图原理

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之间的关系。共有属性表达的一定是事物普遍存在的现象，是共性的表达，具有较大的外延和较浅的内涵；独有属性是区别与其他事物的表达，是个性的表达，具有较深的内涵和较小的外延。从人类认识模式知识的角度来看，可以构造出以属性特征和对象相似性为指标的层次结构图。图 1~3 清楚地表明了以描述自顶向下概念驱动原理的模式分类哲学原理：从宏观共性中发现存在的普遍模式，从微观和个性中发现最独有的模式。

图 1 清楚地表明了基于属性偏序知识发现的哲学原理：从宏观共性中发现存在的普遍模式知识，从微观和个性中发现最独有的模式知识。图 1 中有两个坐标一个是属性，一个是对象。属性坐标靠近原点是宏观，远离原点是微观。对象坐标靠近原点是共性，远离原点是个性。这种构图的方式是将最普遍存在的事物积聚在原点附近，而将个性化的事物远离原点。

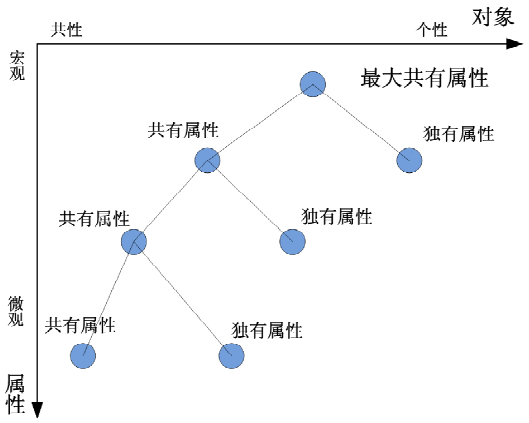

图 1　基于属性偏序的知识发现方法原理示意图

Fig. 1　Philosophical principle of knowledge discovery based on attribute partial ordered

属性偏序结构图是一个封闭非循环的树图结构。图中最顶层和最底层分别有且只有一个节点，树图中边的方向是由上向下的有向边。偏序图中从某一节点沿边到另一节点的走向称为路径。对于偏序图中的任意节点一定能够找到一条路径，其起始于最顶层节点，结束于最底层节点。

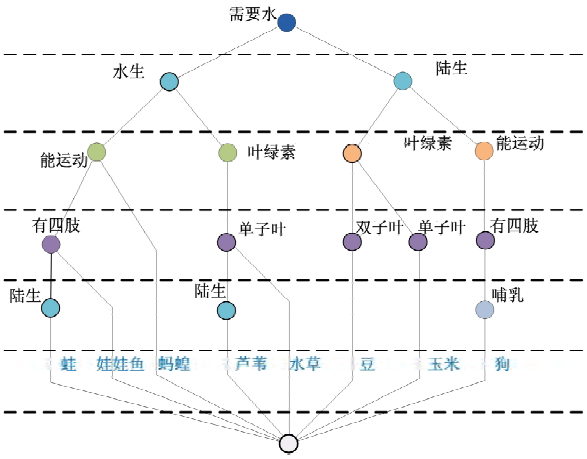

图 2　“生物和水”属性偏序结构图

Fig. 2　Attribute partial ordered structure diagram of lives in water

图 2 给出的是以属性包含对象程度为偏序关系的属性偏序结构图。从认知事物的角度来看，属性与属性之间是结构性关系，而对象与对象之间是相似性关系。从宏观角度考察对象之间的相似性，可以达到认知和区分事物的目的。形式背景中，对象和属性为对偶关系。如果将形式背景中的数据进行转置，即以原来的对象集为新形式背景的属性集，原来的属性集作为新形式背景的对象集，按照属性偏序图生成方法，则可以生成对象偏序结构

图。图 3 给出的是“生物和水”形式背景的对象偏序结构图。

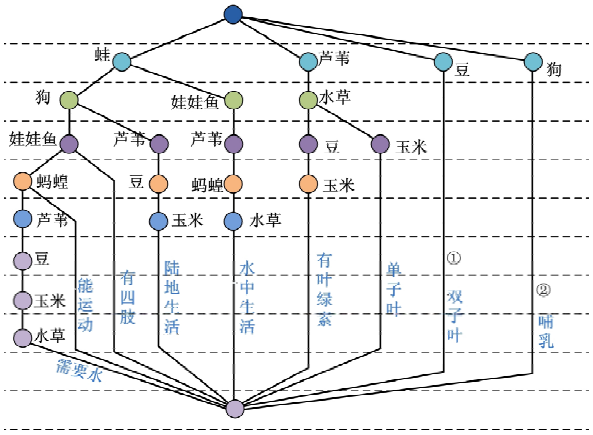

图 3 “生物和水”对象偏序结构图

Fig. 3 Attribute partial ordered structure diagram of lives in water

2 基于属性（对象）偏序结构图表示的知识发现方法

人们认知事物基本原理就是从事物的普遍性和特异性角度，从事物相关的全体事件序结构来区分和认知事物。人类处理信息最常用的模式有自顶向下和自底向上模式。自顶向下模式又称为概念模型驱动模式，这种模式将一个问题（论域）逐步细化为更小的问题（子域）。按照粒计算理论，一个较大的粒（等价类）被分解成若干较小的粒（低层次等价类），较小的粒还可以分成更小的粒（更低层次等价类），全部的粒按照层次和关联建立连接，从而构成偏序结构关系。自底向上模式又称为数据驱动。从原理上讲，自顶向下思维表达的是：普遍性越大，层次越高，特异性越大，层次越低。

属性偏序结构图，从认知事物的基本原理出发，根据数据本体之间的特征关系，构建以属性覆盖对象的程度为偏序关系的层次结构图，这种结构图可以反映出数据属性之间的普遍性和特异性。将形式背景转置后，可以生成对象偏序结构图。对象偏序结构图可以反映出数据对象之间的相似程度，以及数据对象的特异性表达。属性偏序结构图和对象偏序结构图，从认知事物原理的不同角度出发，为发现数据之间的属性结构性关系和对象相似性关系提供了理论基础。

基于属性偏序结构图和对象偏序结构图，可以从集群结构、集子群结构、支路和节点等不同角度对原始数据进行知识发现。表 1 给出了基于属性偏序结构和对象偏序结构的大数据知识发现方法归纳。

表 1 大数据偏序结构图知识发现方法归纳表

Tab. 1 Knowledge discovery methods towards big data based on partial ordered structure diagram

比较项目	属性偏序结构图	对象偏序结构图
原理	自顶向下；普遍性强在上层	自底向上；特异性强在上层
群结构	数据中属性构成的可视化整体偏序结构	数据中对象构成的可视化整体偏序结构
子群结构	数据中属性构成的可视化局部偏序结构	数据中对象构成的可视化局部偏序结构
节点	单个属性	单个对象
支路	属性集合表达的一个对象，一个对象具有什么属性	对象集合表达的一个属性，一个属性具有多少对象
主要用途	发现同一子群结构表达的对象间相似性；不同子群结构表达对象差异性	发现对象独有属性；相邻对象表达相似性；不同子群结构表达属性构成模式不同
规则提取	一条支路自顶向下构成对象表达规则	一条支路自顶向下构成属性表达规则
知识发现	组合应用上面原理，实现知识发现	组合应用上面原理，实现知识发现

2.1 偏序结构图中的集群（Cluster）结构关系

集群结构，可简称为群结构，在属性偏序结构图中是对象所具有的特征构成的结构关系；在对象偏序结构图中是指属性所包含的对象构成的结构关系。

在属性偏序结构图中，每条支路代表一个对象，也表示为一种模式。一些模式聚集在一个（或多个）共有属性节点下，组成一个集群结构（Cluster Structure），简称为集群 (Cluster)，或者簇，英文缩写为 Cr。一个集群表示了一些含有一个（或多个）共有属性的对象的集合，具有聚类作用，能够用于归纳和总结。集群可以由一个属性的集合从属性偏序结构图中提取获得。例如，生物和水的属性偏序结构图中，可以由“在水中生活”、“在陆地生活”和“两栖”这 3 个共有属性分别得到：水生生物集群 {娃娃鱼，蚂蟥，水草}，陆生生物集群 {豆，玉米，狗}，以及两栖生物集群 {芦苇，蛙}。

同理，在对象偏序结构图中，每条支路代表一

种属性。一些属性聚集在一个（或多个）共有的对象节点下，组成一个集群结构。 这样的一个集群结构表示包含了一个（或多个）共同对象的属性的集合，同样具有聚类的作用。此集群可由一个对象的集合从对象偏序结构图中提取获得。例如，在生物和水的对象偏序结构图中，陆生生物特征集群{需要水，能运动，有四肢，在陆地生活} 可由 {蛙，狗} 属性集合获得。

2.2 偏序结构图中的子集群(sub-Cluster)结构关系

子集群结构，可简称为子群结构，在属性偏序结构图中是部分对象所具有的特征构成的局部结构关系；在对象偏序结构图中是指部分属性所包含的对象构成的局部结构关系。

在属性偏序结构图的一个集群 Cr 中，由不同的共有属性节点可以再分离出不同的包含对象数更少的集群，这些集群包含于大集群 Cr，称为大集群 Cr 的子集群，或可称为子簇，英文简写为 sub-Cr。一个集群内的子集群反映了集群内部的大致结构关系，是更进一层次的结构表示。例如，在生物和水形式背景的属性偏序结构图中，在集群 {水生} 中，可以再分割为 {有叶绿素} 集群和 {能运动集群}，也就是植物集群和动物集群。

同理，在对象偏序结构图的一个集群 Cr 中，由不同的公共对象可分离出不同的具有更少属性的集群，同样，这些集群称为大集群 Cr 的子集群，或子簇。

2.3 偏序结构图中的支路模式

在偏序结构图中，每个分支是一个模式，每个模式可以被视为一个规则，模式之间的异同，可以构成一个完全不同的知识。

属性偏序结构图中，支路由多个属性节点以及节点之间的有向边组成，属性节点和有向边组成了属性偏序结构图中的属性模式。每一条支路为一个对象的表达，支路上的节点，为该对象区别于其他对象的规则模式。

对象偏序结构图中，每一条支路为一个属性的表达。如果某一条支路上的节点越多，则该支路表达的属性越具有普遍性；如果某一条支路上的节点越少，则支路表达的属性越具有特异性。应注意的是，当某一支路上的节点只有一个时，则该支路表达的属性的这一节点的独有属性，独有属性表达了事物的独有性质，可以用来区分当前对象区与其他对象。

2.4 偏序结构图中节点之间的关系

属性偏序结构图中，每个节点代表了一个属性。通过图中节点之间的关联，可以看出形式背景中属性之间的结构关系。顶层节点表达了形式背景中所有对象的共有属性，底层节点表达了形式背景中对象的独有属性。

对象偏序结构图中，每个节点表示一个对象。同一支路上的节点，距离越近，其对象的相似度越高；距离越远，对象的相似度越低。

在图 2 生物和水的属性偏序结构图中，顶层的属性节点 {需要水} 是形式背景中的最大共有属性，是形式背景中所有的对象都具有的属性，是最普遍的属性；在最大共有属性下面，分出了两个较大的子群结构，分别是以属性节点 {水中生活} 和 {陆地生活} 为顶点的簇集。在属性节点 {水中生活} 和 {陆地生活} 下面，又分为 {能运动} 和 {有叶绿素} 两个子群结构。从属性偏序结构图中，可以清楚的看到，当前形式背景中的对象，分别聚集于 {水中生活} 和 {陆地生活} 两种属性，说明当前形式背景中的对象，主要分为 {水中生活} 和 {陆地生活} 两大类。

在图 3 生物和水的对象偏序结构图中，左边的第 1 条支路中，包含对象节点为 {蛙，狗，娃娃鱼，蚂蝗，芦苇，豆，玉米，水草}，其中，对象节点“蛙”和“狗”距离最近，其相似性也越高，在当前形式背景中，“蛙”和“狗”均具有属性：{需要水，陆地生活，能运动，有四肢}。对象节点“水草”和“狗”距离最远，相似性也最小，仅仅具有共同的属性：{需要水}。对象偏序图中，每条支路是一个属性的表达。右边支路①和支路②中，分别只含有 1 个对象，此时支路①表达的属性为 {双子叶} 为该支路上的节点对象 {豆} 的独有属性，在当前形式背景中，可以以此作为区分 {豆} 与其他对象的分类依据；支路②表达的属性为 {哺乳}，当前形式背景中，如果某一未知对象具有 {哺乳} 的属性，那么此未知对象一定是 {狗}。

3 基于偏序结构图表示的中药方剂配伍知识发现

中医是中国特有的传统医学，是中华民族在长期与疾病斗争过程中积累的宝贵经验和财富。中药方剂学是中医理论中的一个重要研究领域，主要研究各中药方剂配伍之间的复杂关系结构。本文依据清代著名医学家张锡纯在治疗中风时用到的处方及组成构建中药方剂配伍知识发现形式背景。其中，形式背景的对象集合为张锡纯治疗中风常用的32个经典处方，属性集合是67味中药。表2给出了形式背景中的部分数据，如果某一个处方中含有某一味中药，则用用“×”表示，若没有用到此中药，则用空白表示。

表2 张锡纯中风32方形式背景

Tab. 2 The formal context of ZHANG Xi-chun 32 prescriptions for apoplexy

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	×	×	×			×			×		×		×							×				
2	×	×	×						×	×			×	×									×	
3	×	×	×		×		×	×				×		×		×			×					×
4	×	×	×				×	×				×		×		×					×			×
5	×		×						×		×			×							×			
6				×	×		×	×																
7				×	×		×	×			×					×								×
8				×	×		×	×			×				×									
9		×		×	×		×	×													×	×		
10				×			×	×													×	×		×
11				×	×		×	×									×						×	
12					×																			
13				×	×																			
14																					×			
15				×	×																			
16			×	×	×																	×		
17				×	×																	×		
18	×	×	×			×			×	×	×			×		×								
19	×	×	×			×			×	×	×	×		×					×					
20	×	×	×			×				×	×	×						×	×					
21	×	×	×			×			×	×		×	×					×					×	
22	×	×	×			×				×	×	×			×					×				
23	×	×	×			×			×	×		×			×			×	×				×	
24	×	×	×						×						×			×	×				×	
25	×	×	×	×		×					×	×			×			×						
26	×	×	×	×		×			×									×						
27	×	×	×			×									×			×						
28	×	×	×			×			×	×				×			×							
29	×	×	×			×			×	×	×			×		×				×				
30	×	×	×			×	×	×		×		×			×				×	×		×		
31	×	×		×	×		×	×							×		×							
31	×	×		×	×		×	×							×		×							×

32个处方分别是：1 建瓴汤、2 镇肝熄风汤、3 起痿汤、4 养脑利肢汤、5 熄风汤、6 加味补血汤、7 干颓汤、8 补脑振痿汤、9 振颓汤、10 振颓丸、11 补偏汤、12 升陷汤、13 回阳升陷汤、14 搜风汤、15 逐风汤、16 加味黄芪五物汤、17 加味玉屏风散、18 医案处方1、19 医案处方2、20 医案处方3、21 医案处方4、22 医案处方5、23 医案处方6、24 医案处方7、25 医案处方8、26 医案处方9、27 医案处方10、28 医案处方11、29 医案处方12、30 医案处方13、31 医案处方14、32 医案处方15。所涉及的中药分别有：代赭石，怀牛膝，生杭芍，当归等共67味常见中草药。

图4给出的是形式背景对应的属性偏序结构图。从图中可以清楚的看到，从整体群结构来看，图中主要分为了两个较大的簇集，分别是属性节点{生赭石}和{当归}，也就是说张锡纯在治疗中风的32处方中，绝大部分的处方中，添加了生赭石和当归这两味中药 (32个处方中，有30个处方含有生赭石或当归)。在属性偏序结构图的第3层，属性节点{生赭石}的下面，含有一个较大的子群结构，即属性节点{怀牛膝}，即在使用生赭石的处方中，往往伴随着怀牛膝（从o1到o20的20个使用生赭石的处方中，有19个使用了怀牛膝）。而在属性节点{当归}下，同样含有一个较大子群结构，即属性节点{生黄芪}，说明在使用了当归的处方中，大部分同时使用了生黄芪（从o21到o30的10个使用当归的处方中，有9个使用了生黄芪）。

从属性偏序图中可以清楚的看到，张锡纯在治疗中风的处方中，主要应用到{生赭石}和{当归}。其中，{生赭石}主要与{怀牛膝}配伍使用，{当归}往往与{生黄芪}配伍使用。在属性偏序图的第3层，在使用{生赭石}和{怀牛膝}的情况下，往往与{生杭芍}配伍使用；在使用{当归}和{生黄芪}的情况下，往往与{没药}或{桂枝}配伍使用。

将表2中医肾病形式背景进行对象属性转置后，依据属性偏序结构图生成算法，生成相应的对象偏序结构图，如图5所示。由对象偏序结构图的性质可知：如果某一条支路上的节点数较多，则该支路对应的属性，是该支路上对象所共有的表达，对应的属性普遍性较强。图5中，支路 (1) 和支

路 (2) 的节点数量是最多的，对应的属性为 {生赭石} 和 {怀牛膝}，说明在张锡纯治疗中风的 32 个经典处方中，用药最频繁的是生赭石和怀牛膝；支路 (3) 至支路 (10) 的节点个数也比较多，对应的属性分别是 {生杭绍}、{生怀地黄}、{甘草}、{当归}、{没药}、{乳香}、{生黄芪} 和 {生龙骨}，说明在张锡纯治疗中风的处方中也经常用到。值得说明的是，支路 (7) 和支路 (8)，在同一个节点下分出的两个分支（或多个分支），且各分支上再无其他节点，那么此时这两个分支（多个分支）对应的属性，为属性伴生对。对应于当前形式背景，属性伴生对即中药配伍研究中的药对。支路 (7) 和支路 (8) 对应的属性分别是 {没药} 和 {乳香}，即在张锡纯治疗中风的处方中，{没药} 和 {乳香} 总是药对的方式使用。对象偏序结构图中，同样出现的属性伴生支路有：支路 (11) 和支路 (12)，对应的属性分别是：生龟板和川楝子；支路 (13) 至支路 (16)，对应的属性分别是：半夏、石膏、僵蚕和麝香；支路 (17) 和支路 (18)，对应的属性分别是：熟地和附子；支路 (19) 和支路 (20)，对应的属性分别是：黄蜡和白矾；支路 (21) 和支路 (22)，对应的属性分别是：秦艽和陈皮；支路 (23) 至支路 (25)，对应的属性分别是：羌活、独活和全蝎；支路 (26) 至支路 (28)，对应的属性分别是：柴胡、桔梗和升麻。

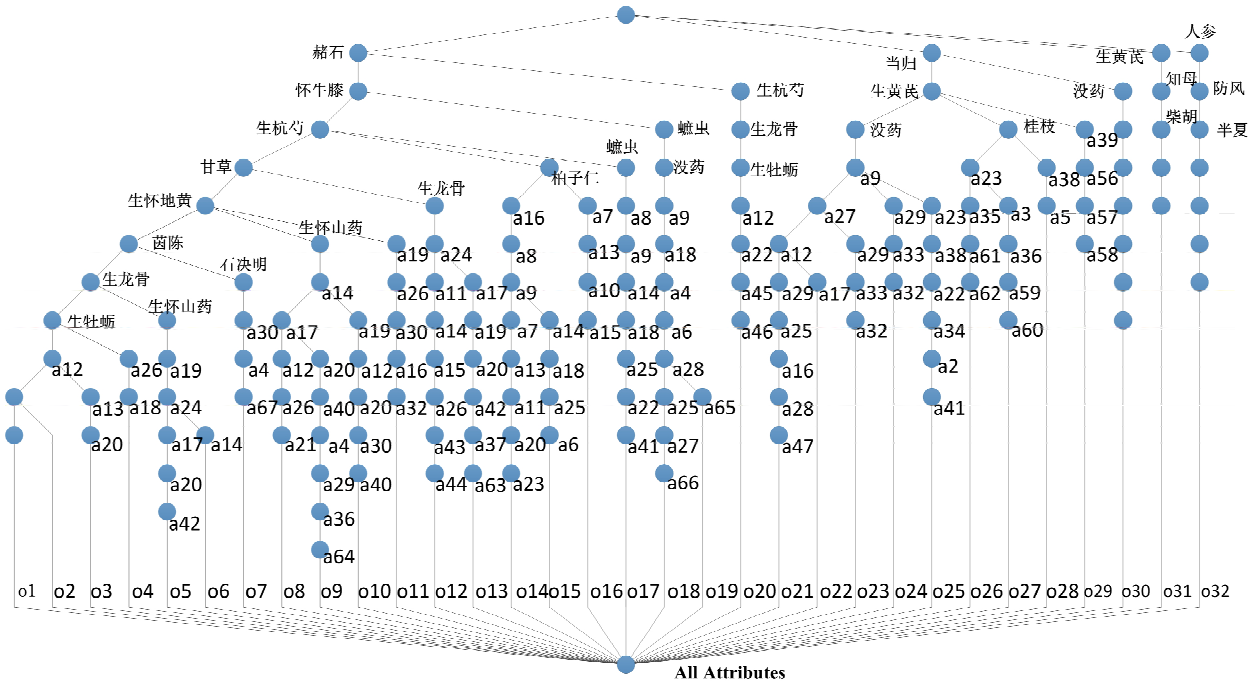

图 4　张锡纯治疗中风 32 方的属性偏序结构图

Fig. 4　Attribute partial ordered structure diagram of ZHANG Xi-chun 32 prescriptions for apoplexy

此外，若某一支路上，只有一个对象节点，则该支路对应的属性，一定是对象节点的独有属性。从图 5 中可以清楚地看到，支路 (29) 中，只有一个节点 {医案处方 9}，该支路对应的属性为 {龙胆草}，则属性 {龙胆草} 是对象 {医案处方 9} 的独有属性。此时，即可以得到中药配伍规则：在张锡纯治疗中风的处方中，如果出现中药龙胆草，则可以直接判断处方为：{医案处方 9}。

支路 (13) 至支路 (16) 只有一个对象节点：{搜风汤}，对应的独有属性为：{半夏}，{石膏}，{僵蚕}，{麝香}，则可以得到张锡纯治疗中风的处方使用规则：{半夏} → {搜风汤}、{石膏} → {搜风汤}、{僵蚕} → {搜风汤}、{麝香} → {搜风汤}。

同样的，支路 (17) 和支路 (18) 有一个对象节点：{熄风汤}，对应的独有属性为：{熟地黄}、{附子}，则可以得到处方使用规则：{熟地黄} → {熄风汤}、{附子} → {熄风汤}。

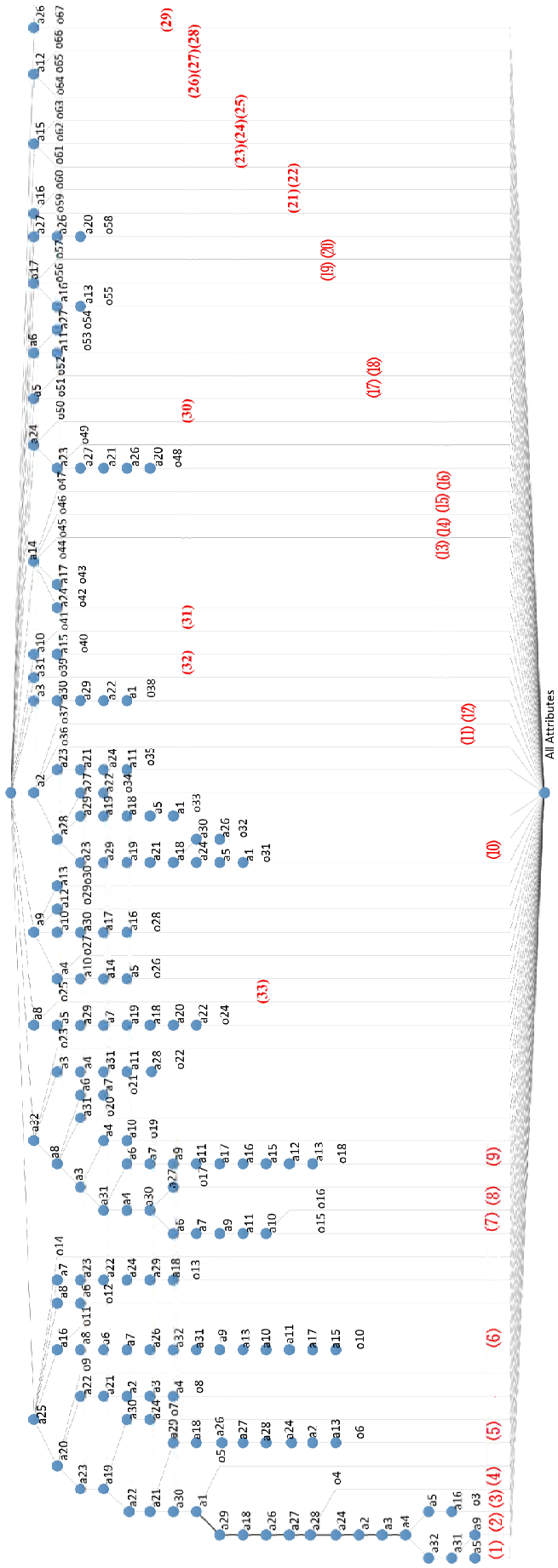

图 5 张锡纯治疗中风 32 方的对象偏序结构图

Fig. 5 Object partial ordered structure diagram of ZHANG Xi-chun 32 prescriptions for apoplexy

支路 (19) 和支路 (20) 对应的节点为：{加味玉屏风散}，对应的独有属性为：{黄蜡}、{白矾}，则可以得到处方规则：{黄蜡} → {加味玉屏风散}、{白矾} → {加味玉屏风散}。

支路 (21) 和支路 (22) 对应的节点为：{加味黄芪五物汤}，对应的属性为 {秦艽}、{陈皮}，则可以直接得到处方规则 {秦艽} → {加味黄芪五物汤}、{陈皮} → {加味黄芪五物汤}。

支路 (23)、支路 (24)、支路 (25) 对应的节点为：{逐风汤}，对应的属性为 {羌活}、{独活}、{全蝎}，则可以直接得到处方规则 {羌活} → {逐风汤}、{独活} → {逐风汤}、{全蝎} → {逐风汤}。

支路 (26)、支路 (27)、支路 (28) 对应节点为：{升陷汤}，对应的属性为 {柴胡}、{桔梗}、{升麻}，则可以直接得到处方规则 {柴胡} → {升陷汤}、{桔梗} → {升陷汤}、{升麻} → {升陷汤}。

支路 (30) 有：{医案处方 7}，对应的独有属性为：{黑脂麻}，则可以得到处方规则：{黑脂麻} → {医案处方 7}。

支路 (31) 有：{振颓丸}，对应的独有属性为：{穿山甲}，则可以得到处方规则：{穿山甲} → {振颓丸}。

支路 (32) 有：{医案处方 14}，对应的独有属性为：{丝瓜络}，则可以得到处方规则：{丝瓜络} → {医案处方 14}。

支路 (33) 有：{补脑振萎汤}，对应的独有属性为：{胡桃肉}，则可以得到处方规则：{胡桃肉} → {补脑振萎汤}。

4　结论

本文介绍了以大数据知识发现为应用背景，在偏序结构理论的基础上，提出了一种基于属性偏序结构图和对象偏序结构图的知识发现方法。从认识事物的基本原理出发，属性偏序结构图表达了数据属性之间的结构性关系；对象偏序结构图表达了数据对象之间的相似性关系。属性偏序结构图将数据中具有某些共同特征的对象聚类到一起，图中的每条支路是一个模式的表达，可以从这些模式里抽取对事物认知和分类的规则提取。对象偏序结构图中，每一条支路为一个属性的表达，节点是对象的表达。对象偏序图中的某一支路中，节点之间的距离，表达了节点对象之间的相似程度，若某一支路上仅存在一个节点，则该支路对应的属性为此对象节点的独有属性，可以依据对象的独有属性将对象与其他数据对象进行分类。最后，以中医方剂配伍，验证了基于属性偏序结构图和对象偏序结构图在中医药配伍知识发现方法的可行性和有效性。

参考文献

[1] Silva J P D, Zarate L E, Vieira N J. Formal concept analysis for data mining: Theoretical and practical approaches [C] //2006 IEEE International Conference on Engineering of Intelligent Systems, 2006: 1-6.

[2] 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——大数据的研究现状与科学思考 [J]. 中国科学院院刊, 2012,27 (6): 647-657.

[3] Wille Rudolf. Restructuring lattice theory: an approach based on hierarchies of concepts [C] // Rivall. Ordered Sets. Dordrecht-Boston: Reidel, 1982: 445-470.

[4] Ganter B, Stumme G, Wille R. Formal concept analysis: foundations and applications [M]. Berlin Heidelberg: Springer-Verlag, 2005.

[5] Priss U. Formal concept analysis in information science [J]. Annual Review of Information Science and Technology, 2006,40 (1): 521-543.

[6] 马垣, 曾子维, 迟呈英, 等. 形式概念及其新进展 [M]. 北京: 科学出版社, 2011.

[7] Zhang T, Ren H L, Wang X M, et al.. A calculation of formal concept by attribute topology [J]. ICIC Express Letters, Part B: Applications, 2013,4 (3): 793-800.

[8] Xu Y H, Zhang T, Wang X Y, et al.. Data mining in Traditional Chinese ophthalmologic formulae based on theory of Structural Partial-Ordered Attribute Diagram[J].ICIC Express Letters,2013,7 (3B): 953-958.

[9] Fan F J, Hong W X, Song J L, et al.. A visualization method of Chinese Medicine knowledge discovery base on Formal Concept Analysis [J]. ICIC Express Letters, Part B: Applications, 2013,4 (3): 801-808.

[10] Song J L, Yu J P, Yan E L, et al.. Syndrome differentiation of Six Meridians for warm disease based on Structural Partial-Ordered Attribute Diagram [J]. ICIC Express Letters, 2013,7 (3B): 947-952.

[11] 樊凤杰, 洪文学. 基于属性偏序结构图表示原理的中药方剂配伍规律研究 [J]. 生物医学工程学杂志, 2013,30 (4): 719-723.

[12] Yu jianping, Chen Nan, Sun Rui, et al.. Word sense disambiguation and knowledge discovery of English modal verb can [J]. ICIC Express Letters, 2013,7 (2): 577-582.

[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-min[1], ZHANG Tao[3], LI Shao-xiong[1], YAN En-liang[1]

(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 (APOSD) and object partial ordered structure diagram (OPOSD) 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 APOSD, 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-xiong[1], ZHANG Tao[3], LUAN Jing-min[1], LIU Wen-yuan[3]

(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