

靖鲲鹏, 宋之杰. 基于属性偏序结构图的文本型灾情多元信息可视化[J]. 灾害学, 2014, 29(3): 57-63. [Jing Kunpeng, Song Zhijie. Textual disaster multivariate information visualization based on attribute partial order structure diagram [J]. Journal of Catastrophology, 2014, 29(3): 57-63.]

# 基于属性偏序结构图的文本型灾情多元信息可视化\*

靖鲲鹏, 宋之杰

(燕山大学 经济管理学院, 河北 秦皇岛 066004)

**摘要:** 在处理和析文本型灾情多元信息时, 由于传统概念格层次结构不够清晰, 不利于对其进行数据挖掘和知识发现以支持决策。应用形式概念分析理论, 在分层概念格建格算法的基础上, 通过形式背景行列交换原理, 对形式背景进行优化, 提出了属性偏序结构图表示方法。该方法可以实现层次化的属性聚类, 便于分析概念构成, 达到分层递阶可视化的效果, 实现知识发现和多元海量数据的信息融合。应用该方法, 绘制了中国1995-1999年农作物受灾和成灾情况的属性偏序结构图。结果表明, 该方法清晰、直观, 可用于支持文本型灾情信息的分析和趋势判断。

**关键词:** 可视化; 灾情信息; 文本型数据; 属性偏序结构图

**中图分类号:** C934; X915.5; X43 **文献标志码:** A **文章编号:** 1000-811X(2014)03-0057-07

**doi:** 10.3969/j.issn.1000-811X.2014.03.012

无论是应急中的实时动态信息, 还是事后的静态统计分析, 文本型数据都是应急管理中灾情信息的主要表现形式。如: 在应急管理工作各阶段, 不同应急部门报送的统计数据、报表、人员伤亡、财产损失等。2002-2008年每年发表于《安全与环境学报》的系列论文“我国事故与灾害状况综述”, 以及从2009年开始《中国减灾》杂志每月发布的“全国灾情月报”等, 都以文本数据(或多属性数据集)的形式对我国灾情做了较为全面、准确的统计和分析。

文本型数据主要包含多维数据(multi-dimension data)和多元数据(multi-variate data)。对于多属性数据集, 相互完全独立的属性被称作“维度”(dimension), 相关的属性被称为“变元”(variate)。由于属性间的相关性往往难以明确判断, 因此多维数据和多元数据经常被称为多维多元(multi-dimensional multi-variate, mdmv)数据。本文将统称为“多元”数据。

灾情信息通常为多属性数据集, 可以表示成矩阵形式。若所要研究的多元数据样本数为 $n$ , 每个样本的变量数为 $m$ , 整个多元数据可以表示为 $n$ 行 $m$ 列的数据矩阵 $X_{(n \times m)}$ 。矩阵中的元素为 $x_{ij}$ , 其中 $i = 1, 2, \dots, n; j = 1, 2, \dots, m$ 。

$$X_{(n \times m)} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} = [X_1, X_2, \dots, X_m] \quad (1)$$

这种以表格或矩阵形式表示的文本型数据具有简单、清晰的优点, 可以进行基本的统计分析。但由于灾情信息具有海量、多源、异构、时变等特征, 使得数据处理比较棘手。特别是在多个时间段、多个空间上比较分析灾情状况, 进行应急决策时, 这种文本型的数据不便于对灾情进行整体把握。

可视化方法是帮助人们“立刻理解”大量数据和信息的有效手段。Gaynor认为危机中, 人们在知觉时间感知压力下所做出的决策必须对动态的不确定状态做出反应<sup>[1]</sup>。也就是说, 管理者必须获取实时的环境数据, 能够“立刻理解”这些数据并采取适当的行动。灾情信息可视化在本质上是一种数据挖掘和知识发现过程, 高效地挖掘出对决策有用的信息, 避免“大数据时代知识贫乏”的现象。在诸多知识发现方法中, 形式概念分析(Formal Concept Analysis, FCA)表现出较大的潜力, 并被认为是有一种有力的知识分析和知识发现工具。

Wille教授基于“概念是由外延和内涵组成的思想单元”这一哲学理解, 在Birkhoff对格理论(lattice theory)贡献的基础上, 于1982年首先引入了概念格(concept lattice)并将其作为一种数学理论, 从数学的角度描述了哲学范畴的“概念”, 奠定了形式概念分析的理论基础。该理论对“概念”进行了形式化描述: 外延是由概念所覆盖的对象构成的集合, 内涵是由概念的所有对象的共有属性构成的集合。这种描述符合人们对世界从感性到抽象的认

\* 收稿日期: 2013-09-24 修回日期: 2013-11-11

基金项目: 国家自然科学基金项目(70871101); 燕山大学博士基金项目(B804)

作者简介: 靖鲲鹏(1977-), 男, 陕西西安人, 副教授, 博士, 主要从事应急管理、信息可视化研究. E-mail: jkp@ysu.edu.cn

知规律。

形式概念分析主要研究“概念”和“概念分层”的数学化描述,其主要实现是:基于对象和属性间的关系,构建形式背景(formal context);从被表示为形式背景的数据中,定义对象与属性的统一体,获取形式概念(formal concept);通过形式概念之间的对象包含关系(或者属性间的包含关系),定义偏序关系,建立以形式概念为元素的层次结构——概念格。全部概念与概念间的例化(特化)和泛化关系组合形成概念格结构<sup>[2]</sup>。因此,形式概念分析又叫概念格理论,是一种用数学的形式化语言来反映人形成概念的过程的集合理论模型,用来研究特定领域可能存在的概念的几何结构、概念格形式<sup>[3]</sup>。

作为形式概念分析的核心数据结构,概念格可以显示对象与特征之间的联系,表明概念之间的泛化与例化关系。使用 hasse 图实现数据的可视化,为提取规则知识提供了一个很好的平台,适合用来发现数据中潜在的概念和规则型知识。

自 1982 年提出形式概念分析后,有较多的论文和著作详细地介绍了形式概念分析<sup>[4-7]</sup>。概念格可以用来作为知识表示和可视化表示的独特而强大的手段<sup>[8]</sup>。形式概念分析提供了一个表示概念定义的语义基础,已经被应用到知识表示领域,比如:本体构建<sup>[9-10]</sup>,本体映射和合并<sup>[11-12]</sup>。与许多其他知识表示形式化比较,本体(Ontology)和形式概念分析方法都是旨在建模“概念”。文献[13]中讨论了如何将这两个形式化方法互补推动建模概念研究。形式概念分析可以用来支持本体工程,以及利用本体可以在形式概念分析得到应用。形式概念分析作为一个学习技术可以支持构建、分析本体,本体可以被利用改善形式概念分析的应用。

形式概念分析是一种无监督的学习概念聚类技术,可用于词汇数据库和分类建模<sup>[14-15]</sup>。文献[16]中引入了冰山概念格的概念,并应用到数据库知识发现。冰山格设计的目的是分析非常大的数据库,从一个已知的关联规则挖掘中得到常见的模式。

形式概念分析也可用来表示和处理领域背景知识,比如:病人病案的说明,解释治疗的决定与治疗规则的表示<sup>[17]</sup>。

文献[18]中提出了基于形式概念分析的信息系统模型验证的方法学。研究证明形式概念分析对于理解概念模型拓扑是有用的,并且它能够用来改善概念模型的结构。

通常,形式背景可以用二维表格来表示。在许多应用场合,对象和属性的关系不是二进制关系,而是多值关系,多值的形式背景通过概念标度(conceptual scaling)的方法转换到单值形式背景<sup>[19]</sup>。

针对二维表形式的文本型形式背景(Textual Formal Concept, TFC),文献[20]中提出了一种文本型形式背景的约简方法 TFC-Reducing,采用信息

损失熵和语义覆盖度评价背景约简。

使用形式概念分析从数据中获取概念与其他基于统计的传统数据分析方法不同,形式概念分析用概念表示数据分析结果,用概念格显示知识视图。概念格已成为近年来获得飞速发展的数据分析的有力工具。目前,形式概念分析已被广泛研究并应用到模式识别、机器学习、软件工程、信息检索、专家系统、决策分析等领域<sup>[21]</sup>。

但是,在处理和文本型灾情多元信息时,传统概念格层次结构不够清晰,不利于对其进行数据挖掘和知识发现以支持决策。本文借助形式概念分析这一描述概念和概念层次的数学模型,通过对形式背景进行优化,提出概念格改进算法——属性偏序图。以灾情多元数据为形式背景,通过将多值形式背景转化为单值形式背景,绘制相应的属性偏序结构图,实现灾情多元数据的可视化表达,以支持灾情分析和应急管理。

## 1 基于属性偏序结构图的多元数据分层递阶表示

### 1.1 形式背景的分层递阶概念格表示

根据形式概念分析的基本原理,可以通过形式背景的概念格,画出相应的 hasse 图,实现形式背景的分层递阶表示<sup>[19,21-22]</sup>。

利用文献[22]中提出的分层建格算法,对形式概念分析中著名的“生物和水”形式背景(表1)进行分析,可以得到相应的概念格(图1)。但是,这种概念格存在线条交叉、层次结构不够清晰的不足之处,不利于海量多元数据的数据挖掘和知识发现。

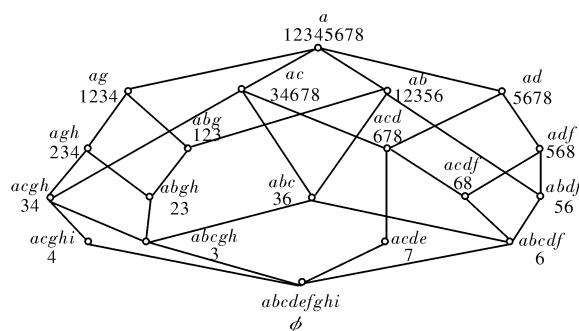


图1 生物和水形式背景的分层概念格

为了更好地构建能够处理以多元数据为形式背景的分层递阶概念格,本文通过对初始形式背景进行优化,提出概念格的改进算法——基于形式背景行列交换原理的属性偏序结构图。该方法可对形式背景进行分层递阶可视化表示,具有较显著层次关系结构。

### 1.2 概念格的改进算法——属性偏序结构图

基于形式背景行列交换原理的属性偏序结构图表示方法,是通过形式背景行列交换原理将多元

表 1

生物和水形式背景

对象		属性								
		<i>a</i> 需要水	<i>b</i> 在水里生活	<i>c</i> 在陆地生活	<i>d</i> 有叶绿素	<i>e</i> 双子叶	<i>f</i> 单子叶	<i>g</i> 能运动	<i>h</i> 有四肢	<i>i</i> 哺乳
1	蚂蝗	×	×					×		
2	娃娃鱼	×	×					×	×	
3	蛙	×	×	×				×	×	
4	狗	×		×				×	×	×
5	水草	×	×		×		×			
6	芦苇	×	×	×	×		×			
7	豆	×		×	×	×				
8	玉米	×		×	×		×			

海量数据描述成分层递阶结构,对形式背景进行分层优化,借助属性偏序结构图这一特殊的格生成方法,实现分层优化形式背景的可视化。

一个形式背景  $K = (O, A, R)$ , 其中  $O$  是对象的集合,  $A$  是属性的集合,  $R$  是  $O$  和  $A$  之间的一个二元关系, 并且具有  $(O_1, O_2, O_3, \dots, O_i, \dots, O_m)$  对象排列次序,  $(A_1, A_2, A_3, \dots, A_j, \dots, A_n)$  属性排列次序。如果  $a_{ij}$  为属性值, 且  $a_{ij} \in A$ 。则当对象  $O_i$  具有属性  $A_j$  时,  $a_{ij} = 1$ ; 否则  $a_{ij} = 0$ 。形式背景  $K = (O, A, R)$  转换为分层形式背景  $K_{y_0} = (O, A, R)$  的充分必要条件: 具有新的对象排列次序  $(O'_1, O'_2, O'_3, \dots, O'_m)$  和新的属性排列次序  $(A'_1, A'_2, A'_3, \dots, A'_n)$ 。

新的对象排列次序和属性排列次序可以由下面的方法确定。

(1) 求出属性值求和最大值所对应的列。即:

$$a_{mj} = \max(\sum_{i=1}^m a_{i1}, \sum_{i=1}^m a_{i2}, \dots, \sum_{i=1}^m a_{ij}, \dots, \sum_{i=1}^m a_{in}), \quad (2)$$

(2) 将属性列的第一列与  $a_{mj}$  对应的第  $j$  列进行交换, 得到一个新的属性排列次序:  $(A'_1, A'_2, A'_3, \dots, A'_j, \dots, A'_n)$ 。

(3) 再做行交换。使得属性值  $a_{i1} = 1$  从  $a_{i1}$  开始连续排列, 得到一个新的对象排列次序  $(O'_1, O'_2, O'_3, \dots, O'_i, \dots, O'_m)$ 。

为了定义形式背景的层次结构, 在这里介绍子背景和不相交(互斥)子背景的概念。

**定义 1**<sup>[19]</sup> 如果  $K = (O, A, R)$  是一个形式背景, 而且  $H \in O, N \in A$ 。则

$$(H, N, R \cap H \times N), \quad (3)$$

就是  $K = (O, A, R)$  的子背景。

**定义 2**<sup>[19]</sup> 设形式背景  $K = (O, A, R)$  有两个子背景:  $K_1 = (O_1, A_1, R_1)$  和  $K_2 = (O_2, A_2, R_2)$ 。

如果  $K = (O_1 \cup O_2, A_1 \cup A_2, R_1 \cup R_2)$ , 那么  $K_1 = (O_1, A_1, R_1)$  和  $K_2 = (O_2, A_2, R_2)$  是不相交(互斥)的背景。

由上面两个定义, 我们可以将分层形式背景  $K_{y_0} = (O, A, R)$  拆分成两个不相交(互斥)子背景  $K_1 = (O_1, A_1, R_1)$  和  $K_2 = (O_2, A_2, R_2)$ 。

其中  $K_1 = (O_1, A_1, R_1)$  为属性值  $a_{i1} = 1$  对应的背景部分,  $K_2 = (O_2, A_2, R_2)$  为属性值  $a_{i1} = 0$  对应的背景部分。

对这两个形式背景做如下变换:

(1) 对子背景  $K_2 = (O_2, A_2, R_2)$  (不含第一列)的列重新排序, 确保  $a_{12} = 1$ , 并且那些  $a_{1j} = 1$  的列连续地排在新序列的前面。同时, 子背景  $K_1 = (O_1, A_1, R_1)$  的对应列的排列顺序也做同样的变换。

(2) 子背景  $K_1 = (O_1, A_1, R_1)$  的行重新排序, 使得同一属性  $a_{ij} = 1$ 。

(3) 变换后得到新的形式背景  $K_{y_1} = (O, A, R)$ , 并且具有新的对象排列次序  $O''_1, O''_2, O''_3, \dots, O''_i, \dots, O''_m$  和新的属性排列次序  $A''_1, A''_2, A''_3, \dots, A''_j, \dots, A''_n$ 。注意  $K_{y_0} = (O, A, R)$  和  $K_{y_1} = (O, A, R)$  的第一个属性是相同的。

进行第二次分层变换后, 再重复上面的变换, 直到属性值求和最小。整个分层优化过程完成。

从数学意义上看, 该分层过程实质是根据属性集合的普遍性对集合  $A$  进行子集划分。其目的是使普遍性高的子族外延的并包含普遍性低的子集成员的并。

该过程的数学描述为: 设形式背景中的属性集合  $A = \{A_1, A_2, \dots, A_m\}$ ,  $A_i$  表示形式背景中的第  $i$  个属性。定义属性  $A_i$  的度为:

$$\text{Degree}(A_i) = \|A'_i\|_0. \quad (4)$$

式中:  $\text{Degree}(A_i)$  的值表示属性  $A_i$  的普遍性大小。 $\text{Degree}(A_i)$  的值越大, 表示在当前形式背景下属性  $A_i$  越具有普遍性; 该值越小, 表示属性  $A_i$  越具有特异性。但直接的  $\text{Degree}(A_i) = \|A'_i\|_0$  计算并未考虑集合间的包含关系, 因此需要进行修正。

设  $\text{Degree}(A_i)$  的集合为  $D, D = \{\|A'_i\|_0 \mid i = 1, 2, \dots, m\} = \{0, 1, 2, \dots, d \mid d \in N\}$ 。根据集合论, 必有  $d \leq \text{Degree}(A)$  且  $d \leq \text{Degree}(O)$ 。

可得: 属性数为  $j$  的对象集合是  $D_j = \{A'_i \mid \|A'_i\|_0 = j, i = 1, 2, \dots, m\}$ , 属性数为  $j$  的属性集合是  $MD_j = \{A_i \mid \|A'_i\|_0 = j, i = 1, 2, \dots, m\}$ 。

比较相邻两个对象集合  $D_j$  和  $D_{j-1} (j > 0)$ , 遍历  $A_i \subseteq MD_{j-1}$ , 如果

$$A'_i \mid_{A_i \subseteq MD_{j-1}} \subseteq (U_{A'_i \subseteq D_j} A'_i), \quad (5)$$

说明对于当前  $A_i \subseteq MD_{j-1}$ ,  $D_j$  中所有元素的并包含  $D_{j-1}$  中所有元素的并, 不需要进行修正。否则, 令  $\|A'_i\|_0 = j$ , 更新  $D_j$  和  $MD_j$ , 再次执行修正操作, 直至满足(式 3) 或  $\|A'_i\|_0 = d$ 。

表 2 分层优化后的生物和水形式背景

对象		属性							
		<i>a</i> 需要水	<i>b</i> 在水里生活	<i>c</i> 在陆地生活	<i>d</i> 有叶绿素	<i>e</i> 双子叶	<i>f</i> 单子叶	<i>g</i> 能运动	<i>h</i> 有四肢
1	蚂蝗	×	×					×	
2	娃娃鱼	×	×					×	×
3	蛙	×	×	×				×	×
6	芦苇	×	×	×	×		×		
5	水草	×	×		×		×		
4	狗	×		×				×	×
7	豆	×		×	×	×			
8	玉米	×		×	×		×		
属性的度		8	5	5	4	3	3	4	3

表 1 所示为未经分层优化的形式背景。利用上面提出的形式背景优化的分层方法,处理表 1 中生物和水的形式背景。因为表 1 中形式背景较为简单,只需将对象 4 和对象 6 做行交换即可,得到分层优化后的生物和水形式背景(表 2)。依据表 2 形式背景,可以做出生物和水关系的属性偏序结构图(图 2)。

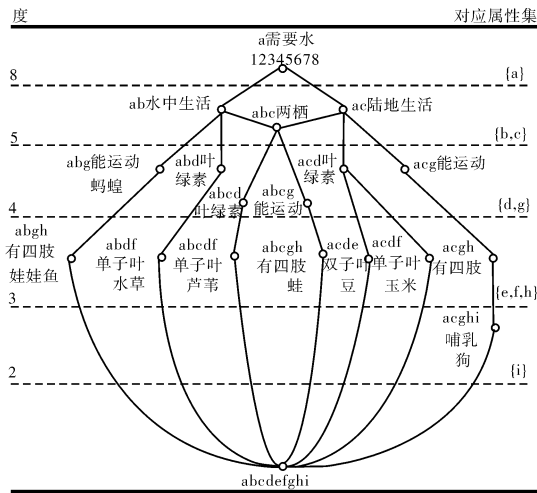


图 2 形式背景分层优化后的生物和水属性偏序结构图

比较表 1 和表 2, 可以看到: 优化后的形式背景是严格分层表示的。比较图 1 和图 2, 可以得到以下结论: 基于形式背景分层优化方法, 生成的属性偏序结构图可视化表达, 其层次关系图表示简练, 并且无交叉连线, 类别意义清楚, 层次关系鲜明。具有某一属性对象的聚类表示, 有助于挖掘知识体系。适合应用于对复杂系统多元海量数据进行分层递阶化的可视化信息融合研究。

但是, 需要指出的是: 尽管基于形式背景分层优化原理构造的属性偏序结构图与概念格相似, 但属性偏序结构图不是形式概念分析理论中的概念格。依据形式概念分析理论将属性偏序结构图与概念格做一个简要比较, 可知: 从层次描述角度看, 概念格是概念层次描述; 而属性偏序结构图是由形式背景特定属性不为零的数目决定层次, 由形式背景特定对象不为零的属性数目决定层次层级, 是概念

关系描述。形式背景的概念格唯一, 形式背景的属性偏序结构图不唯一。

## 2 中国 1995 – 1999 年农作物受灾和成灾情况分析

### 2.1 基本概念

(1) 农作物受灾面积: 是指年内因遭受旱灾、水灾、风雹灾、霜冻、病虫害及其他自然灾害, 使农作物较正常年景产量减产 10% 以上的农作物播种面积。受灾面积不得重复计算, 在同一块土地上如先后遭受几种或几次灾害, 只按其受灾最大最重的一次计算受灾面积。

(2) 成灾面积: 是指在遭受上述自然灾害的受灾面积中, 农作物实际收获量较常年产量减少 30% 以上的播种面积。

(3) 绝收面积: 是指在遭受上述自然灾害的受灾面积中, 农作物实际收获量较常年产量减少 70% 以上的播种面积。

### 2.2 分级标准的基础

根据《自然灾害风险分级办法》<sup>[23]</sup>, 自然灾害风险

$$R = P \times C. \quad (6)$$

式中:  $R$  为自然灾害风险;  $P$  为自然灾害风险事件发生的可能性;  $C$  为自然灾害风险事件产生的后果。

可能性  $P$  和后果  $C$  都划分为 4 个等级: 1 为“极高”; 2 为“高”; 3 为“中”; 4 为“低”。

在《自然灾害风险分级办法》中提出分级原则具有“可扩展性”, 即该办法的分级指标可以依据实际需要进行调整。

在《自然灾害风险分级办法》的附录“洪水灾害风险事件风险等级划分示例”中, 根据后果指标中“需政府救助人数占农牧业人口的比率或人数( % 或万)”, 将后果划分为 4 个等级: 1 级为  $> 30\%$ ; 2 级为  $26\% \sim 30\%$ ; 3 级为  $21\% \sim 25\%$ ; 4 级为  $15\% \sim 20\%$ 。

### 2.3 属性划分标准

根据“农作物受灾和成灾面积”原始数据表( 国

家统计局网站,作者整理),计算相关指标。

SZ = 受灾面积 / 播种总面积:表示受灾面积占总播种面积的百分比。

SH = 旱灾受灾面积 / 受灾面积:表示旱灾受灾面积占受灾面积的百分比。

SS = 水灾受灾面积 / 受灾面积:表示水灾受灾面积占受灾面积的百分比。

SF = 风雹灾受灾面积 / 受灾面积:表示风雹灾受灾面积占受灾面积的百分比。

SD = 霜冻灾受灾面积 / 受灾面积:表示霜冻灾受灾面积占受灾面积的百分比。

CZ = 成灾面积 / 受灾面积:表示成灾面积占受灾面积的百分比。

CH = 旱灾成灾面积 / 成灾面积:表示旱灾成灾面积占成灾面积的百分比。

CS = 水灾成灾面积 / 成灾面积:表示水灾成灾面积占成灾面积的百分比。

CF = 风雹灾成灾面积 / 成灾面积:表示风雹灾成灾面积占成灾面积的百分比。

CD = 霜冻灾成灾面积 / 成灾面积:表示霜冻灾成灾面积占成灾面积的百分比。

可以得到:1995 - 1999 年全国和 30 个省农作物受灾和成灾面积指标计算结果。

结合指标计算结果,根据“可扩展性”原则,对“洪水灾害风险事件等级划分”中灾害后果等级划分指标进行调整,制定“农作物受灾和成灾情况”形式背景属性的划分标准如下:1 级为  $> 50\%$ ;2 级为  $31\% \sim 50\%$ ;3 级为  $26\% \sim 30\%$ ;4 级为  $21\% \sim 25\%$ ;5 级为  $15\% \sim 20\%$ ;6 级为  $< 15\%$ 。

将“农作物受灾和成灾”中的每个指标(共 10 个),都划分为 6 个等级,用于衡量农作物受灾或成灾的总体严重性,或某一种灾害(旱灾、水灾、风雹、霜冻)对受灾或成灾影响的严重性。如:SZ1 为“受灾面积 / 播种总面积”  $> 50\%$ ;CZ2 为“成灾面积 / 受灾面积”在  $31\% \sim 50\%$ ;SF4 为“风雹灾受灾面积 / 受灾面积”在  $21\% \sim 25\%$ 。

## 2.4 建立形式背景

以“1995 - 1999 年全国农作物受灾和成灾面积”为例,形式背景(部分)如表 3 所示。

## 2.5 生成属性偏序结构图

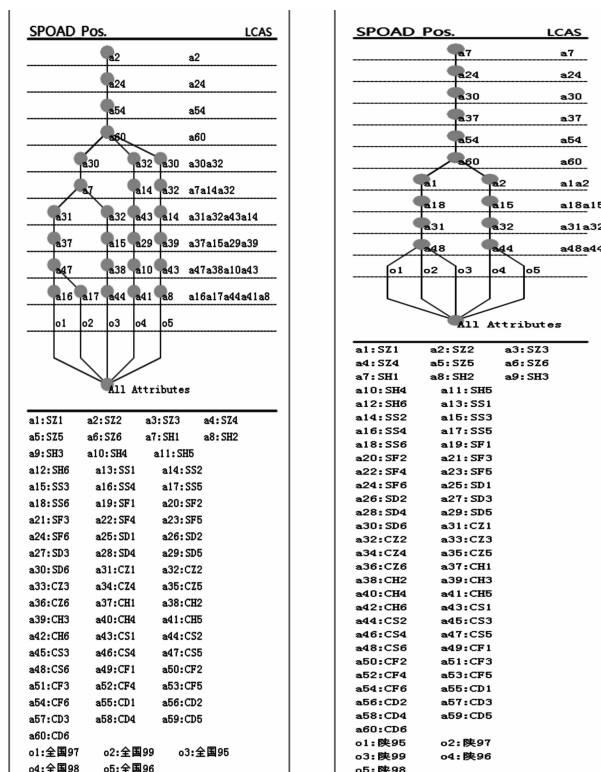
根据全国 1995 - 1999 年农作物受灾和成灾面积的形式背景,可得到该形式背景的属性偏序结构图;同理可得到陕西省 1995 - 1999 年农作物受灾和成灾情况的属性偏序结构图(图 3)。

## 2.6 分析属性偏序结构图

从图 3 所示的属性偏序图中,可以看到某一对象所具有的所有属性集合。如在图 3(a)中,对象 O1 的属性集合为:

{a2, a24, a54, a60, a30, a7, a31, a37, a47, a16}。即 1997 年全国农作物受灾和成灾情况对应的指标为:

$$\{SZ2, SF6, CF6, CD6, SD6, \{SH1, CZ1, CH1, CS5, SS4\}\}。$$



(a) 全国 (b) 陕西省  
图 3 全国和陕西省 1995 - 1999 年农作物受灾和成灾情况的属性偏序结构图

同理,对象 O3 的属性集合为:

{a2, a24, a54, a60, a30, a7, a31, a37, a47, a16}。即 1995 年全国农作物受灾和成灾情况对应的指标为:

$$\{SZ2, SF6, CF6, CD6, SD6, \{SH1, CZ2, SS3, CH2, CS2\}\}。$$

在这两年农作物受灾和成灾 10 个指标中,前 6 个指标和对应级别均相同;后 4 个指标相同,级别有所差别。可看出:1995 年和 1997 年灾情相近;1995 年旱灾成灾较弱,水灾受灾和成灾都较 1997 年严重;1995 年农作物总体成灾情况比 1997 年低。

根据图 3(a),可以看出:

(1) 全国 1995 - 1999 年间,每年农作物受灾总体情况均处于 2 级水平(SZ2),即受灾面积占农作物播种面积的百分比为  $30\% \sim 50\%$ 。

(2) 1997 年和 1999 年灾情最为严重(CZ、SH 和 CH 均处于 1 级),即总体成灾面积、旱灾受灾面积和旱灾成灾面积均高于  $50\%$ 。这两年的水灾受灾面积(SS)处于中等偏下水平,1999 年(SS5)相对于 1997 年(SS4)低了 1 个级别,即水灾受灾面积从 1997 年的  $21\% \sim 25\%$ 降低到  $15\% \sim 20\%$ 。

(3) 1996 年和 1998 年灾情相似,其中 CZ、SS、CS 均处于 2 级,即这两年中水灾是造成农作物受灾和成灾的主要因素,且灾害损失较为严重,处于  $31\% \sim 50\%$  之间。但 1996 年旱灾比 1998 年较为严重,旱灾造成的受灾面积和成灾面积都高。

表 3 1995 – 1999 年全国农作物受灾和成灾面积的形式背景(部分)

对象	S Z 1	S Z 2	S Z 3	S Z 4	S Z 5	S Z 6	S H 1	S H 2	S H 3	S H 4	S H 5	S H 6	S S 1	S S 2	S S 3	S S 4	S S 5	S S 6	S F 1	S F 2	S F 3	S F 4	S F 5	S F 6	S D 1	S D 2	S D 3	S D 4	S D 5	S D 6
全国 9 5		1					1							1										1						1
全国 9 6		1						1						1										1						1
全国 9 7		1					1								1									1						1
全国 9 8		1							1					1										1					1	
全国 9 9		1					1									1								1						1

(4)1995 年旱灾和水灾同时存在,水灾受灾面积(SS3)中等偏上,处于 26%~30%。旱灾和水灾造成的农作物成灾面积较大(CH2、CS2),均处于 31%~50%,总体成灾损失严重(CZ2)。

根据图 3(b),可以看出:

(1)陕西省 1995 – 1999 年间,农作物受灾和成灾主要是由旱灾引起(SH1、CH1),每年的旱灾受灾面积超过农作物播种面积的 50%,旱灾成灾面积超过受灾面积的 50%。

(2)5 年间,陕西省由风雹灾和霜冻灾造成的农作物损失很少,均低于 15%。

(3)在 1995 年、1997 年和 1999 年,受灾和成灾情况均处于最高级别(SZ1、CZ1);1996 年和 1998 年,由于雨量相对较多,水灾造成了一定的损失(SS3、CS2),但也缓解了一定的旱情,使得这两年的受灾和成灾情况有所降低(SZ2、CZ2),受灾和成灾面积处于 31%~50%。

以同样的方法,可以得到:全国 30 个省份 1995 – 1999 年农作物受灾和成灾情况的属性偏序结构图、1995 – 1999 年各年度 30 个省农作物受灾和成灾情况的属性偏序结构图。

以 1995 年 30 个省农作物受灾和成灾情况的属性偏序结构图(部分)为例(图 4)。(限于篇幅,其余的属性偏序图本处省略)。可以根据图 4 分析 1995 年各省农作物受灾和成灾情况的特征。

同理,可以根据以上的方法,分析全国各省在 1995 – 1999 年的农作物受灾和成灾情况。也可以按年度对全国 30 个省的农作物受灾和成灾情况进行分析。

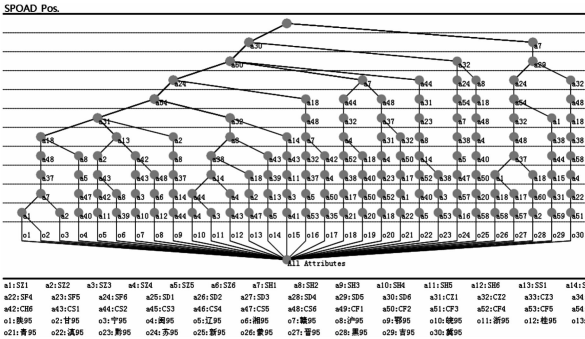


图 4 1995 年 30 个省农作物受灾和成灾情况的属性偏序结构图(部分)

3 结论

在文本型灾情多元信息分析和可视化中,将形式概念分析理论、知识库、数据库等相结合,从海量数据中抽取有用信息和知识,是切实可行并且是有意义的。该方法的主要优点在于:可以将灾情多元海量数据库中的表面或隐含数据,通过图示化,完整地表现出这些数据和知识之间的内在逻辑和组织结构,从而为分析数据之间的关联信息提供系统的可视化工具。

但是,在建立形式背景时,需要根据指标计算值确定属性划分标准,进而将多值背景转换为单值形式背景。这种属性划分的标准,还需要经过相关主管部门确认,以便该方法具有更好的指导性。

参考文献:

[1] Gaynor M, Seltzer M, Moulton S, et al. A Dynamic, data-driv-

- en, decision support system for emergency medical services [M]//Computational Science-ICCS 2005, Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2005, 3515: 703–711.
- [2] 康向平. 基于形式概念分析理论的知识获取模型研究[D]. 太原: 山西大学, 2012.
- [3] 钱杰. 基于形式概念分析的本体构建与映射方法研究[D]. 长沙: 国防科学技术大学, 2006.
- [4] Ganter B, Wille R, Franzke C. Formal concept analysis: Mathematical foundations[M]. New York: Springer-Verlag, 1997.
- [5] Kalfoglou Y, Dasmahapatra S, Chen-Burger Y H. FCA in Knowledge Technologies: Experiences and Opportunities [M]//Concept Lattices. Berlin, Heidelberg: Springer, 2004: 252–260.
- [6] Diaz-Agudo B, Gonzalez-Calero PA. Formal concept analysis as a support technique for CBR[J]. Knowledge-Based System, 2001, 14(3): 163–171.
- [7] Priss U. Formal concept analysis in information science[J]. Annual Review of Information Science and Technology, 2006, 40: 521–543.
- [8] Jiang G Q, Pathak J, Chute CG. Formalizing ICD coding rules using formal concept analysis [J]. Journal of Biomedical Informatics, 2009, 42(3): 504–517.
- [9] Schoening J. IEEE P1600. 1: Standard Upper Ontology Working Group (SUO WG) [EB/OL]. (2003–12–28) [2013–11–12]. <http://suo.ieee.org/>.
- [10] Jiang G, Ogasawara K, Endoh A, et al. Context-based Ontology Building Support in Clinical Domains Using Formal Concept Analysis[J]. International Journal of Medical Informatics, 2003, 71(1): 71–81.
- [11] Kalfoglou Y, Schorlemmer M. IF-Map: An ontology-mapping method based on information-flow theory [M]//Journal on Data Semantics I. Berlin, Heidelberg: Springer, 2003: 98–127.
- [12] Stumme G, Maedche A. FCA-Merge: Bottom-up merging of ontologies [C]//International Joint Conference on Artificial Intelligence. Lawrence Erlbaum Associates LTD., 2001, 17(1): 225–234.
- [13] Cimiano P, Hotho A, Stumme G, et al. Conceptual knowledge processing with formal concept analysis and ontologies [M]//Concept Lattices. Berlin Heidelberg: Springer, 2004: 189–207.
- [14] Priss U, Old LJ. Modelling lexical databases with formal concept analysis [J]. Journal of Universal Computer Science, 2004, 10(8): 967–984.
- [15] Priss U. Formalizing Botanical Taxonomies [M]//Conceptual Structures for Knowledge Creation and Communication. Berlin Heidelberg: Springer, 2003: 309–322.
- [16] Stumme G. Efficient data mining based on formal concept analysis [C]//Database and Expert System Applications. Berlin Heidelberg: Springer, 2002: 534–546.
- [17] Schnabel M. Representing and processing medical knowledge using formal concept analysis [J]. Methods of Information in Medicine, 2002, 41(2): 160–167.
- [18] Laukaitis A, Vasilecas O, Plikynas D. Formal concept analysis for business information systems [J]. Information Technology and Control, 2008, 37(1): 33–37.
- [19] 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, 2010.
- [20] 杨小平, 何伟, 孙亚琳, 等. TFC-Reducing: 一种基于属性语义距离和规则的文本型形式背景约简方法[J]. 小型微型计算机系统, 2012, 33(10): 2170–2176.
- [21] 许研. 基于 FCA 的信息检索模型研究及应用[D]. 开封: 河南大学, 2007.
- [22] 潘跃建. 基于 FCA 面向多数据源的领域本体创建方法研究[D]. 南京: 南京航空航天大学, 2010.
- [23] 中华人民共和国民政部. MZ/T031–2012 自然灾害风险分级方法[EB/OL]. [2013–11–05]. <http://files.mca.gov.cn/yunnan/201209/20120901201208905.pdf>.

## Textual Disaster Multivariate Information Visualization based on Attribute Partial Order Structure Diagram

Jing Kunpeng and Song Zhijie

(Economic and Management College, Yanshan University, Qinhuangdao 066004, China)

**Abstract:** An hierarchy of traditional concept lattice is not clear in processing and analyzing multivariate information about textual disaster. This is not conducive to carry on data mining and knowledge discovery to support decision making. It is proposed that the representation method of attribute partial order structure diagram by applying the ranks exchange principle of formal context to optimize formal context. The method has advantage in analyzing some concepts and hierarchical attribute clustering. It can realize knowledge discovery, hierarchical information visualization and information fusion of huge amounts of multivariate data. The method was applied to draw multiple attribute partial order structure diagrams of the national crop disaster and inundated area during 1995–1999. The results show that this approach is clear, intuitive, and support to information analysis and trend judgment of textual disaster information.

**Key words:** visualization; information on disastrous situation; text-data; attribute partial order structure diagram