

基于对象间相互作用的形式概念分析

牛昕悦

作业	分数[20]
得分	

2021年11月29日

基于对象间相互作用的形式概念分析

牛昕悦

(大连海事大学 信息科学与技术学院 116026)

摘要 论文对概念格构成所依赖的形式背景作了研究,针对形式背景中的对象集,提出了每个对象并不是孤立存在而是相互作用和相互影响的观点。定义了基于对象间相互作用的一系列相关概念,对于相互作用下不断变化的属性值引入属性加权法,结合概念定标原理,将动态多值背景通过两次转换为静态单值背景,进而对其进行形式概念分析。

关键词 形式概念分析 相互作用 动态多值背景 概念定标 概念格

Formal Concept Analysis Based on the Interaction among Objects

YANG Li SONG Zhen-ming

(Department of Applied Mathematics, Southwest Jiaotong University, Chengdu 610031)

Abstract: The formal context is the base of the constructing of concept lattice so the investigation of it is necessary. For the set of objects in the formal context, the author presents a viewpoint that objects are interactive, not exist alone. A series of relative notions based on the interaction among objects are defined. Then a weighting method of attributes is adopted to deal with the changeable attribute values. According to the conceptual scaling theory, the dynamic many-valued context will be transformed into the static single-valued context.

Keywords: formal concept analysis, interaction among objects, dynamic many-valued context, conceptual scaling theory, concept lattice

1 引言

形式概念分析理论又称为概念格理论。是由德国的 Wille 教授于 1982 年作为一种数学理论首先提出的^[1],它是一种基于概念和概念层次的数学化表达的应用数学的一个分支,在应用形式概念分析理论时,需要用数学的思维方式进行概念数据分析和知识的处理。近年来,形式概念分析被认为是数据分析的有力工具,然而,通常情况下的数据分析并非起源于单值属性的形式背景(简称为单值背景^[2]),而是多值属性的形式背景(简称为多值属性^[3]),并且将数据库看成对象的简单集合无疑会导致许多非常重要信息的丢失。每个对象都不是孤立存在的,它们之间存在着相互作用和相互影响,以往利用形式概念分析理论进行数据分析时,总是忽略这种关系的存在或是假设这种关系的不存在,那么势必会造成分析结果的不真实。

本文在提出对象间相互作用存在的必然性之后,给出了相互作用下的一系列相关概念,针对于不断变化的属性值,引入属性加权法,并结合概念定标原理通过两次转换将动态多值背景化为静态单值背景,从而对其进行形式概念分析。文章最后以具体实例来说明对象间相互作用下问题解决的过程。

2 基本概念

本章简要介绍所需要的形式概念分析的基本概念,相关的详细描述可参看文献[1]。

2.1 单值属性的形式背景

形式概念分析通常由形式背景这一基本概念开始。在形式

概念分析中,单值属性的形式背景简称为单值背景,可用一个三元组 (G, M, I) 来表示,其中, G 是对象集, M 是属性集,而 I 是 G 和 M 间的二元关系,即 $I \subseteq G \times M$; 用 gIm 或 $(g, m) \in I$ 来表示“对象 g 具有属性 m ”。

对于 $A \subseteq G$ 和 $B \subseteq M$, 定义: $A' = \{m \in M \mid \#g \in A, gIm\}$

$B' = \{g \in G \mid \#m \in B, gIm\}$

因此, A' 是 A 中所有对象所共有的属性, B' 是具有 B 中属性的对象之集,则形式背景 (G, M, I) 的概念定义为元素对 (A, B) , 这里 $A \subseteq G, B \subseteq M, A' = B$ 和 $B' = A$, 概念 (A, B) 的外延是 A , 而它的内涵是 B 。

2.2 多值属性的形式背景

对象—属性—值关系是对现实世界问题进行编码的一种常用的数据结构,在统计学中被表示为数据矩阵,在计算机科学中被表示为关系数据库,而在形式概念分析中,它被阐述为多值属性的形式背景即多值背景。例如对于属性项“颜色”、“重量”、“性别”、“级别”等,都具有多种属性值,对于这种具有多种值的属性就称为多值属性。

一个多值背景 (G, M, W, I) , 其中 G 是对象集, M 是多值属性集, W 是属性值的集合,而 I 是它们之间的一个三元关系, $I \subseteq G \times M \times W$, 使得对任意 $g \in G, m \in M$, 最多只有一个值 $w \in W$ 满足 $(g, m, w) \in I$, 即 $(g, m, w) \in I$ 和 $(g, m, v) \in I$ 总蕴含有 $w = v$, 表明相同的对象的同一个属性项的值应该相等; 用 $(g, m, w) \in I$ 表示“对于属性 m , 对象 g 具有属性值 w ”。

3 对象间相互作用下的相关理论

自然界的一切事物都存在着相互作用和相互影响,将数据库看成对象的简单集合无疑会导致许多非常重要的信息的丢失,对象间的相互作用和影响会引起实体属性的变化,通常体现在属性值大小的变化上。属性值的不断变化必会使所形成的多值背景发生变化,从而就有了动态多值背景,也必将会影响到形式概念分析的结果。因此我们有必要对对象间存在着的这种关系进行研究。

3.1 对象间相互作用存在的必然性

随着科学技术的进步,人们将研究的对象看成是一个系统,而系统是在外界因素不断变化的环境中演变发展的。对于某一个属性项而言,各对象所对应的属性值是不同的,这为对象间相互作用的发生创造了前提条件。对象间的相互作用可以理解为在时间的推移和环境的变化下同时发生在不同对象之间,由各对象在同一属性项上的差异所导致的知识溢出。知识溢出本是一个经济学概念,是指知识的扩散、传播、转移和获取^[7]。这一关系的发生必会使原有属性值发生正向或负向的变化。例如,学生的学科成绩在某一阶段会因各种因素的影响而上下波动,而这种波动必会对其评定结果有所影响。因此,对象间的互动关系是必然存在的,在很多情况下也是不容忽视的。

3.2 动态多值背景的形成

在对象间的相互作用下,属性值是不断变化着的,对于属性值所发生的每一次变化,我们可以采用数学方法作一次记录,并将记录次数用来表示。

定义 1 对于给定的属性集 $M=\{m_1, m_2, \dots, m_n\}$ 和在相互作用下用户所做的记录次数 k , 集合 $M^{(k)}=\{m_1^{(k)}, m_2^{(k)}, \dots, m_n^{(k)}\}$; $m_1^{(1)}, m_2^{(1)}, \dots, m_n^{(1)}; m_1^{(2)}, m_2^{(2)}, \dots, m_n^{(2)}; \dots, m_1^{(k)}, m_2^{(k)}, \dots, m_n^{(k)}$ 被称为是 M 的 k 次记录属性集, 它的每个元素都是一个 k 次记录属性值。

定义 2 对于给定的属性值集合 $W=\{w_{ij}|m_j(g)=w_{ij}\}$ 和在相互作用下用户所做的记录次数 k , 集合 $W^{(k)}=\{w_{ij}^{(k)}|m_j(g)=w_{ij}^{(k)}, k=0, 1, 2, \dots\}$ 被称为是 W 的 k 次记录属性值集, 它的每个元素都是一个 k 次记录属性值。

定义 3 在对象集 G 、 k 次记录属性集 $M^{(k)}$ 和 k 次记录属性值集 $W^{(k)}$ 之间, 可以定义一个三元关系 $I^{(k)}$, 其定义如下: $I^{(k)} \subseteq G \times M^{(k)} \times W^{(k)}$ 使得对任意 $g \in G, m_j^{(k)} \in M^{(k)}$, 最多只有一个 $w_{ij}^{(k)}$ 满足 $(g, m_j^{(k)}, w_{ij}^{(k)}) \in I^{(k)}$ 。

定义 4 在对象间的相互作用下, 对于给定对象集 G 、 k 次记录属性集 $M^{(k)}$ 以及 k 次记录属性值集 $W^{(k)}$, 根据上面定义的三元关系 $I^{(k)}$, 得到的多值背景 $(G, M^{(k)}, W^{(k)}, I^{(k)})$ 被称为 k 次记录多值背景。由于 k 的值是随着情况而变化的, 那么所形成的 k 次记录多值背景也就不同, 在这种动态数据情况下, 我们又称次记录多值背景为动态多值背景。

注: 对于 k 不同的取值, 就可以得到一系列次记录多值背景。特别地, 当 $k=0$ 时, 即对所有数据不做任何记录, 得到的 k 次记录多值背景 $(G, M^{(0)}, W^{(0)}, I^{(0)})$ 就等同于互动关系发生前所形成的多值背景 (G, M, W, I) 。

3.3 多值背景由动态向静态的转化

对于 k 次记录属性集 $M^{(k)}$ 和记录前的属性集 M 而言, 它们的元素实质上并没有发生变化, 也即对一组对象, 在相互作用

发生前后, 属性集中包含的属性项仍保持不变, 那么根据此特点, 就可以将动态中的属性值转化为静态中的属性值从而得到静态多值背景。为了使属性值更能反映实际情况, 那么根据各属性项的重要程度以及属性值在相互作用下的变化程度, 定义一个权重向量集 $U = \{u_1^{(k)}, u_2^{(k)}, \dots, u_n^{(k)}\}, u_j = \sum_{i=1}^n u_{ij}^{(k)} = 1, k=1, 2, \dots$, 权重集中的每组值可以根据专家调查法来决定。

定义 5 对于一个动态多值背景 $(G, M^{(k)}, W^{(k)}, I^{(k)})$, 称 (G, M, CW, I_c) 为一个静态多值背景, 其中 G 是对象集, M 是多值属性集, CW 是静态多属性值集, 且 $CW = \{w_j | w_j = \sum_{k=1}^n u_j^{(k)} w_{ij}^{(k)} | w \in W, w_j^{(k)} = w_j^{(k)} - w_j^{(k-1)}, k=1, 2, \dots\}$, 在对象间相互作用下, k 次记录所得到的动态多值背景, 通过属性加权法可以转换为静态多值背景, 经过多次记录和专家的评定, 我们不仅可以较客观和准确地得到分析结果, 而且还可以对对象在属性项上的变化程度做出预测。

对于一个静态多值背景, 若想采用形式概念分析理论分析数据、解决问题, 以目前现有的理论是无法直接解决的, 这就需要进行第二次转换——静态多值背景向静态单值背景的转换。以下所述的概念定标原理便实现了这种转换。

3.4 概念定标原理

为了对多值背景进行形式概念分析, 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^[4](conceptual scaling)。概念定标的基本思想是通过概念标尺^[6]从多值背景中导出单值背景。每个多值属性 $m \in M$ 可以被赋以一个概念标尺 S_m , 它也是一个形式背景 $S_m = (W_m, M_m, I_m)$, 其中: $W_m \subseteq M \cdot G = \{m(g) | g \in G\}$ 。即 W_m 是包含属性 m 的所有属性值的集合, 这样的形式背景就称为概念定标。多值背景 (G, M, W, I) 连同标尺的族集 $S_m (m \in M)$ 被称为简单定标的背景, 它确定的派生背景被定义为 $(G, \cup_{m \in M} \{m\} \times M_m, J)$, 其中 $gJ(m, n) = m(g) \cap n(g) \neq \emptyset, g \in G, m \in M, n \in M$, 可以看出, 概念标尺反映了专家对于属性值的内部结构的认识, 通过它的使用, 概念定标产生了存储于多值背景中的数据的概念模式的全局视图。

4 应用实例分析

例 1 在提拔干部时, 决策者为了把德才兼备的人才提拔到领导岗位上, 首先制定一些指标, 由群众推荐、评议, 对各项指标进行打分, 从中确定候选人。假设某单位拟从初选后的 5 位干部 $g(i=1, 2, \dots, 5)$ 中提拔 1 人担任领导职务。干部的优劣用 6 个属性去衡量, 它们分别是: m_1 —政策水平; m_2 —工作作风; m_3 —业务水平; m_4 —口才; m_5 —写作能力; m_6 —健康状况。群众为每个候选人打分, 每个候选人在各指标下的分数如表 1。

表 1 互动关系发生前的多值背景

	m_1	m_2	m_3	m_4	m_5	m_6
g_1	95	90	93	85	91	95
g_2	90	88	85	92	93	91
g_3	92	95	96	84	87	94
g_4	89	93	88	94	92	90
g_5	93	91	90	89	92	95

表 1 是一个不考虑对象间相互作用的多值背景 (G, M, W, I) , 其中 $G=\{g_1, g_2, \dots, g_5\}, M=\{m_1, m_2, \dots, m_6\}, W=\{w$ 表中所有的属

性值}, $I^{(1)} \bullet G \times M \times W$ 是它们之间的三元关系。

随着时间的变化,受各种因素的影响,对象间存在相互作用时,各候选干部的分值就会发生相应的变化,如果只做一次记录 $k=1$ 时,相应的可以得到表 2。

表 2 一次记录动态多值背景

	$m_1^{(0)}$	$m_2^{(0)}$	$m_3^{(0)}$	$m_4^{(0)}$	$m_5^{(0)}$	$m_6^{(0)}$	$m_1^{(1)}$	$m_2^{(1)}$	$m_3^{(1)}$	$m_4^{(1)}$	$m_5^{(1)}$	$m_6^{(1)}$
g1	95	90	93	85	91	95	93	88	93	88	91	91
g2	90	88	85	92	93	91	87	93	83	90	92	93
g3	92	95	96	84	87	94	93	92	90	94	90	94
g4	89	93	88	94	92	90	94	90	92	90	90	88
g5	93	91	90	89	92	95	95	85	88	92	92	93

表 2 为 1 次记录多值背景即动态多值背景($G, M^{(1)}, W^{(1)}$, $I^{(1)}$), 其中 $G=\{g_1, g_2, \dots, g_5\}, M^{(1)}=\{m_1^{(1)}, \dots, m_6^{(1)}; m_1^{(0)}, \dots, m_6^{(0)}\}$, $W^{(1)}=\{表中所有的属性值\}$ 。下面就采用属性加权法将上述动态多值背景转换为静态多值背景。对于动态多值背景($G, M^{(1)}, W^{(1)}$, $I^{(1)}$)由专家调查法确定了各属性项的权重分布 $u^{(1)}(j=1, 2, \dots, 6)$ 相应的权重向量 $u^{(1)}=(0.2, 0.2, 0.25, 0.1, 0.1, 0.15)$, 则根据公式计算可得到表 3。

表 3 静态多值背景

	m_1	m_2	m_3	m_4	m_5	m_6
g1	94.6	89.6	93.0	85.3	91.0	94.4
g2	89.4	89.0	85.5	91.8	92.9	91.3
g3	92.2	94.4	94.5	85.0	87.3	94.0
g4	90.0	92.4	89.0	93.6	91.8	89.7
g5	93.4	89.8	89.5	89.3	92.0	94.7

表 3 是由动态多值背景($G, M^{(1)}, W^{(1)}, I^{(1)}$)转换而得到的静态多值背景(G, M, CW, I)。表 3 中的每个属性值由 $w_{ij}^{(1)} = w_{ij} + u_{ij} \cdot w_{ij}^{(1)}$ 计算可得。例如, $w_{12}^{(1)} = w_{12} + u_{12} \cdot w_{12}^{(1)} = w_{12} + (w_{12}^{(1)} - w_{12}) = 90 + 0.2(88 - 90) = 89.6$ 。

在转换为单值背景之前,先来做以下过渡。对于表 3,每个属性项的属性值可以排成一个取值在 $[0, 100]$ 的链,而每个对象可通过这六条链所形成的直积来描述其属性情况,由此可以得到表 3- 1。

表 3- 1 静态多值背景的链状结构

	m_1	m_2	m_3	m_4	m_5	m_6
g1	≥ 90	≥ 85	≥ 90	≥ 85	≥ 90	≥ 90
g2	≥ 85	≥ 85	≥ 85	≥ 90	≥ 90	≥ 90
g3	≥ 90	≥ 90	≥ 90	≥ 85	≥ 85	≥ 90
g4	≥ 90	≥ 90	≥ 85	≥ 90	≥ 90	≥ 85
g5	≥ 90	≥ 85	≥ 85	≥ 85	≥ 90	≥ 90

此多值背景(G, M, CW, I_c)中 G, M 没发生变化,而 $CW=\{\geq 85, \geq 90\}, I^{(1)} \bullet G \times M \times CW$ 。

为了进行数据分析,并且需要寻找到所有的概念并在它们之间建立层次关系,那么就要采用概念定标原理将多值背景转换为单值背景。针对表 3- 1 中属性 m_1 ,可化为两个标准高、低(根据实际情况可以有不同的标准),分别以 sm_1, lm_1 来表示。例如($g_3, m_1, \geq 90$)就可转换为(“ \times ”代表“具有”)。

(sm_1, lm_1) $\geq 90 \rightarrow \times \times$, 那么由静态多值背景表 3- 1 所转换的静态单值背景如表 3- 1- 1。

表 3- 1- 1 静态单值背景

	m_1		m_2		m_3		m_4		m_5		m_6	
	1	2	3	4	5	6	7	8	9	10	11	12
g1	\times	\times	\times		\times	\times	\times		\times	\times	\times	\times
g2	\times		\times	\times	\times		\times	\times	\times	\times	\times	\times
g3	\times	\times	\times	\times	\times	\times	\times		\times		\times	\times
g4	\times	\times	\times		\times		\times	\times	\times	\times		\times
g5	\times	\times	\times		\times		\times		\times	\times	\times	\times

基于表 3- 1- 1 可以建立相应的概念格,由于考虑了对象间的相互作用,表 3- 1- 1 更能真实地反映出数据的变化。

5 结论

在形式概念分析理论已有的应用基础之上,重新对其所依赖的形式背景做了分析,提出了对象间相互作用存在的必然性的观点,并定义了相互作用下的一系列相关概念。进而引入属性加权法并结合概念定标原理,对对象间相互作用下的动态多值背景进行形式概念分析。这使得形式概念分析理论能够更真实和客观地在数据分析上得到应用。(收稿日期: 2006 年 5 月)

参考文献

1.B Ganter, R Wille.Formal Concept Analysis: Mathematical Foundations[M].Berlin, Heidelberg: Springer, 1999
2.B Ganter, R Wille.Conceptual Scaling[C].In: Roberts F ed.Applications of Combinatorics and Graph Theory to the Biological and Social Sciences, New York: Springer, 1989: 139~167
3.G Birkhoff.Lattice Theory[C].In: American Mathematical Society Colloquium Publications XXV, American Mathematical Society, Providence, RI, 1967; 25
4. Karl Erich Wolff.Concepts in Fuzzy Scaling Theory: order and granularity[J].Fuzzy Sets and Systems, 2002; 132: 29~48
5. 徐泽水.不确定多属性决策方法及应用[M].北京:清华大学出版社, 2004: 36~37
6.S Pollandt, R Wille.Functorial Scaling of Ordinal Data[J].Discrete Applied Mathematics, 2005; 147: 101~111
7. 孙兆刚, 徐雨森, 刘则渊.知识溢出效应及其经济学解释[J].科学学与科学技术管理, 2005; (1)

(接上)
cessing[J].Personal Communication, 1992
4.Hall D L, Llinas J.An introduction to multisensor data fusion[J].Proc of the IEEE, 1997; 85(1) : 6~23
5.L A Llein.Sensor and data fusion concepts and applications[M].SPIE Optical Engineering, 1999: 47~176

6.M Bedworth, J Obrien.The Omnibus Model: A new model of data fusion[J].IEEE Aerospace and Electronic Systems Magazine, 2000; 15(4) : 30~36
7.B Dasarathy.Sensor fusion potential exploitation- innovative architectures and illustrative applications[J].IEEE Proceedings, 1997; 85(1) : 24~38