

模糊概念格构建的 Bordat 方法

乌弘毅, 黄映辉

(大连海事大学 信息科学技术学院, 辽宁 大连 116026)

摘要:概念格是近年来兴起的知识表示模型。现实世界中的事物大多具有不精确性特征,如何将若干模糊对象构建成一个概念格具有重要的理论与应用价值。选择适于构建精确概念格的 Bordat 方法,应用模糊集理论对其进行改进,重新定义了概念格的顶节点确定过程和子节点生成算法,提出了构建模糊概念格的 Bordat 方法,结合实例说明了其应用,即生成相关的模糊概念,同时构建与之对应的模糊概念格。该方法不仅保留了作为数据源的模糊形式背景的所有信息,而且所需生成的模糊概念数量少,方法简单快捷,易于计算机实现。

关键词:模糊形式背景;模糊概念;模糊概念格;Bordat 方法

中图分类号:TP391.1

文献标识码:A

文章编号:1673-629X(2010)10-0127-04

Bordat Method of Constructing Fuzzy Concept Lattice

WU Hong-yi, HUANG Ying-hui

(Information Science and Technology College, Dalian Maritime University, Dalian 116026, China)

Abstract: Concept lattice is the knowledge representation model which is rising in recent years. Most things in the real world have imprecision features, how to construct a concept lattice based on some fuzzy objects has important values in both theory and practice. This paper uses fuzzy set theory to improve Bordat method which is suitable for constructing the precise concept lattice, redefines the process determining the top node and the algorithm obtaining the child nodes, proposes the Bordat method of constructing fuzzy concept lattice, and includes an example to illustrate the method that generates the relevant fuzzy concepts while constructing a corresponding fuzzy concept lattice. This method not only preserves all the information of a fuzzy formal context which is the data source, but also generates a small number of fuzzy concepts. It is simple and quick, and easy to be implemented by a computer.

Key words: fuzzy formal context; fuzzy concept; fuzzy concept lattice; Bordat method

0 引言

概念(concept)是人类认识客观世界及其规律的基本形式。考察概念与概念之间的关系,对语义网络、信息处理以及人工智能的研究有着非常重要的意义。表示概念以及概念之间的关系有多种方法,概念格(concept lattice)方法能够通过生成 Hasse 图生动简洁地呈现概念之间的层次关系,且具有结构的唯一性^[1]。对概念格的研究成为计算机科学领域的热点之一。

概念格最初只是针对精确概念的,然而客观世界中的绝大多数事物却是模糊的,模糊概念(fuzzy concept)是概念的主要形式。这样,模糊概念格(fuzzy concept lattice)^[2]的构建与应用就显得十分有意

义^[3,4]。文中通过对适于构建精确概念格的 Bordat 方法进行改进,使之成为构建模糊概念格的一种简单有效方法。

1 模糊概念格

概念格,由 R. Wille 于 1982 年首次提出,是一种基于形式背景(formal context)表示形式概念(formal concept)的模型,是形式概念分析(formal concept analysis, FCA)的核心数据结构,被认为是进行数据分析的有力工具^[5,6]。概念格的每个节点表示一个形式概念,由外延和内涵两部分组成。概念的外延是指此概念所覆盖的对象的集合;概念的内涵则是外延所具有的共同属性的集合^[7]。

对于由精确概念构成的概念格已被学者们熟知。为更好地理解由模糊概念构成的模糊概念格,现比较分析其与概念格的共性与差异,如表 1 所示。

在模糊形式背景中,对象与属性之间的模糊关系用属性隶属函数表示。例如“李明今年 30 岁”,根据模

收稿日期:2010-01-19;修回日期:2010-04-04

基金项目:国家自然科学基金资助项目(60972090);辽宁省自然科学基金资助项目(20072142)

作者简介:乌弘毅(1986-),男,内蒙古赤峰人,硕士研究生,从事智能信息处理的研究;黄映辉,教授,从事智能信息处理的研究。

模糊“年轻”的隶属函数 Y , 计算“30岁”隶属于“年轻”的程度为 $Y(30) = 0.5$, 即对象“李明”具有“年轻”这个模糊属性, 相应的模糊关系表示为 $I_{李明}(年轻) = 0.5$ 。显然, 客观世界中大多数是不确定的模糊关系, 而这样的模糊关系用(精确)概念格是无法表示的。

表1 概念格与模糊概念格之比较

	(精确)概念格	模糊概念格
形式背景	(O, D, R)	(O, D, I)
二元关系	确定性关系 R	模糊性关系 I
$f: F(O) \rightarrow F(D)$	精确属性映射	模糊属性映射
$g: F(D) \rightarrow F(O)$	精确对象映射	模糊对象映射
对象的属性值	$\{0, 1\}$	$[0, 1]$
概念(节点)	精确概念	模糊概念
概念的内涵	D 的子集	D 的模糊子集

可见, 与(精确)概念格不同, 模糊概念格是在概念外延与概念内涵之间具有模糊性关系的前提下, 研究和处理概念的外延与内涵之间的属性隶属关系。

2 Bordat 方法

构建概念格的过程实际上是概念聚类的方法。对于概念格, 已经有一些成熟的构建方法, 一般分为两类: 批处理算法(batch algorithm)和渐进式算法(incremental algorithm)^[8]。

批处理算法简单直观, 易于并行化, 而且有一些具体方法能够在建格的同时生成 Hasse 图。Bordat 方法就是其中的代表, 它不仅能生成 Hasse 图, 而且相比其他方法更便于概念格的剪枝以及分类规则的提取^[9]。

Bordat 方法支持自顶向下的建格过程。首先, 确定格的顶节点; 其次, 生成顶节点的所有子节点; 第三, 选择某个子节点作为父节点, 生成其全部子节点; 最后, 对于每个子节点, 递归重复上一步。

子节点的生成是 Bordat 方法的关键。设形式背景 $K = (O, D, R)$, 对于某精确概念 $C_i(O_i, D_i)$ (其中 $O_i \subseteq O, D_i \subseteq D$), 在其子节点生成过程中, 主要是对 D_i 的补集 D_i^c 集合中的属性进行操作。具体过程是: 对于 D_i^c 集合中的每一个属性 y , 求出相应的 $g(y) = \{x \mid xRy, \forall x \in O_i\}$, 并对除去 y 之外的 D_i^c 集合中的每一个属性 y' 进行 $g(y')$ 求值, 然后判断最大子集, 根据判断的结果, 决定是否生成子节点。

3 模糊概念格构建的 Bordat 方法

迄今针对模糊概念格构建的方法所做的研究还不多^[10, 11]。文中通过对概念格构建的 Bordat 方法的改进, 提出了构建模糊概念格的一种新尝试。

3.1 模糊子节点的生成

处理模糊概念 $C_i(O_i, D_i)$ ($i = 1, 2, \dots, m$) 之间的包含关系时, 可以设模糊形式背景的属性集 D 为模糊运算的论域, 则每个概念的内涵都为 D 的模糊子集。利用 Zadeh 表示法, 模糊概念 $C_i(O_i, D_i)$ 的内涵可表示为: $D_i = \sum_{j=1}^n \frac{\mu_i(d_j)}{d_j}$, 其中 n 为模糊形式背景中的属性个数, $\mu_i(d_j)$ 为属性 d_j 对概念 C_i 的内涵 D_i 的隶属度。

根据 Zadeh 模糊集合理论^[12], 两个模糊子集之间的运算实际上就是逐点对隶属度 μ 作相应的运算。模糊集 A 和模糊集 B 等价, 记为 $A = B$, 是指当且仅当对于任意 $x \in U$, 有 $\mu_A(x) = \mu_B(x)$ 成立。模糊集 A 是模糊集 B 的子集, 记为 $A \subseteq B$, 是指当且仅当对于任意 $x \in U$, 有 $\mu_A(x) \leq \mu_B(x)$ 成立。

设模糊形式背景 $K = (O, D, I)$, 若其某概念 C 的外延为 $O_c = \{o_1, o_2, \dots, o_i\}$, 则其内涵可以表示为:

$$\begin{aligned} D_c &= \sum_{j=1}^n \frac{\bigwedge_{t \in O_c} I_t(d_j)}{d_j} \\ &= \sum_{j=1}^n \frac{\inf_{t \in O_c} I_t(d_j)}{d_j} \\ &= \sum_{j=1}^n \frac{\min(I_1(d_j), I_2(d_j), \dots, I_i(d_j))}{d_j} \end{aligned} \quad (1)$$

设模糊形式背景 $K = (O, D, I)$ 的属性集 D 为模糊运算的论域, $A(O_1, D_1)$ 和 $B(O_2, D_2)$ 是 K 上的两个模糊概念。若对任意 $d_j \in D$, 有 $\mu_A(d_j) < \mu_B(d_j)$ 成立, 则 $D_1 \subset D_2$; 根据概念的定义, 若 $O_1 \supset O_2$, 则 A 是 B 的父节点(上层节点)。

模糊形式背景 $K = (O, D, I)$, I 为属性隶属函数, 属性集 D 的属性个数为 n 。生成模糊概念 $C(A, B)$ 的子节点的步骤如下:

(1) 判断单对象概念。判断 A 是否为单元素集合,

若是, 则子节点为 $(\emptyset, \sum_{i=1}^n \frac{\sup_{t \in A} I_t(d_i)}{d_i})$, 将其连接到节点 C 上, 方法结束; 若不是, 方法继续。

(2) 生成临时子节点。

① 标号节点。令 $i = 1$, 即对第一个属性 d_1 进行操作, 同时将 i 作为生成的临时子节点的标号。

② 计算外延。对每个临时子节点设置连接标记 $C_i.linkFather = true$, 求概念 C_i 的外延 $A_i = \{o \mid o \in A, I_o(d_i) > I_A(d_i)\}$ 。

③ 计算内涵。判断 A_i 与已生成临时子节点 C_j 的外延 A_j ($j < i$) 是否有包含关系, 若有, 即 $A_i \subseteq A_j$ 或 $A_i \supseteq A_j$, 则置 $C_i.linkFather = false$; 若无, 则求概念 C_i

的内涵 $B_i = \sum_{j=1}^n \frac{I_i(d_j)}{d_j}$ 。

④ 判断是否跳出循环。令 $i = i + 1$ 。若 $i \leq n$, 转②; 否则转(3)。

(3) 连接父节点。linkFather 值为 true 的临时子节点, 作为 $C(A, B)$ 的子节点并与之连接。

3.2 建格步骤

加入模糊关系运算后的 Bordat 方法由两种操作构成: 顶节点的确定和子节点的确定。

顶节点的确定直接取决于模糊形式背景。顶节点的外延就是形式背景的对象集 O , 顶节点的内涵由式(1)求得。

子节点的确定依赖于三个关键选择:

(1) 选择父节点。在当前生成的一系列节点中选择一个作为父节点, 以进行下一步的子节点生成过程。

(2) 判断子节点。判断该节点是否有子节点, 若有, 则选取其中一个进行操作。

(3) 判断兄弟节点。判断当前节点是否有兄弟节点, 若有, 则选取其中一个进行操作。

综上所述, 模糊概念格构建的 Bordat 方法的具体流程如图 1 所示。

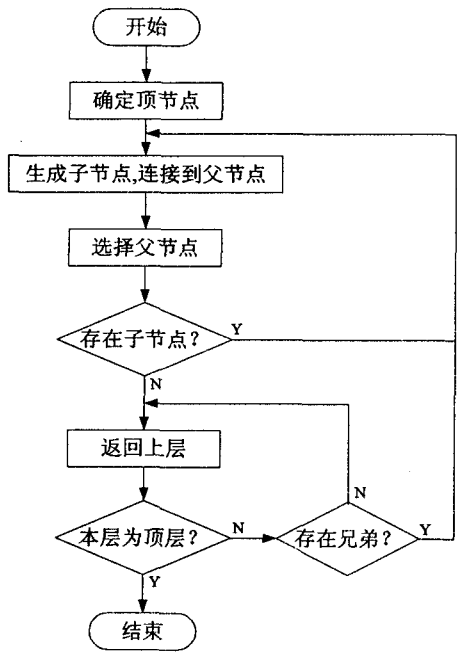


图 1 模糊概念格构建的 Bordat 方法的流程

4 应用实例

选择表 2 所示的模糊形式背景, 其对象集 $O = \{1, 2, 3, 4\}$, 属性集 $D = \{a, b, c, d, e\}$ 。

采用上述的改进 Bordat 方法构建相应的模糊概念格。

首先, 确定概念格的顶节点。由式(1), 求得其内涵为 $\frac{0.1}{a} + \frac{0.2}{b} + \frac{0.2}{c} + \frac{0.4}{d} + \frac{0.1}{e}$, 则可确定顶节点为

$(\{1, 2, 3, 4\}, \frac{0.1}{a} + \frac{0.2}{b} + \frac{0.2}{c} + \frac{0.4}{d} + \frac{0.1}{e})$ 。

表 2 模糊形式背景实例

	a	b	c	d	e
1	0.2	0.4	0.6	0.8	0.5
2	0.1	0.2	0.7	0.9	0.4
3	0.7	0.6	0.3	0.4	0.3
4	0.5	0.8	0.2	0.6	0.1

其次, 生成子节点。根据文中提出的模糊概念子节点的生成方法, 可求得顶节点的子节点分别为:

$(\{1, 3, 4\}, \frac{0.2}{a} + \frac{0.4}{b} + \frac{0.2}{c} + \frac{0.4}{d} + \frac{0.1}{e})$

$(\{1, 2, 3\}, \frac{0.1}{a} + \frac{0.2}{b} + \frac{0.3}{c} + \frac{0.4}{d} + \frac{0.3}{e})$

$(\{1, 2, 4\}, \frac{0.1}{a} + \frac{0.2}{b} + \frac{0.6}{c} + \frac{0.1}{d} + \frac{0.1}{e})$

最后, 分别对这三个子节点重复上述过程, 并对其每个子节点递归该过程。

基于表 2 所示模糊形式背景建立的模糊概念格的 13 个模糊概念如下(依照各概念的生成顺序排列):

#1($\{1, 2, 3, 4\}, \frac{0.1}{a} + \frac{0.2}{b} + \frac{0.2}{c} + \frac{0.4}{d} + \frac{0.1}{e}$)

#2($\{1, 3, 4\}, \frac{0.2}{a} + \frac{0.4}{b} + \frac{0.2}{c} + \frac{0.4}{d} + \frac{0.1}{e}$)

#3($\{1, 2, 3\}, \frac{0.1}{a} + \frac{0.2}{b} + \frac{0.3}{c} + \frac{0.4}{d} + \frac{0.3}{e}$)

#4($\{1, 2, 4\}, \frac{0.1}{a} + \frac{0.2}{b} + \frac{0.2}{c} + \frac{0.6}{d} + \frac{0.1}{e}$)

#5($\{3, 4\}, \frac{0.5}{a} + \frac{0.6}{b} + \frac{0.2}{c} + \frac{0.4}{d} + \frac{0.1}{e}$)

#6($\{1, 3\}, \frac{0.2}{a} + \frac{0.4}{b} + \frac{0.3}{c} + \frac{0.4}{d} + \frac{0.3}{e}$)

#7($\{1, 4\}, \frac{0.2}{a} + \frac{0.4}{b} + \frac{0.2}{c} + \frac{0.6}{d} + \frac{0.1}{e}$)

#8($\{3\}, \frac{0.7}{a} + \frac{0.6}{b} + \frac{0.4}{c} + \frac{0.4}{d} + \frac{0.3}{e}$)

#9($\{4\}, \frac{0.5}{a} + \frac{0.8}{b} + \frac{0.2}{c} + \frac{0.6}{d} + \frac{0.1}{e}$)

#10($\emptyset, \frac{0.7}{a} + \frac{0.8}{b} + \frac{0.7}{c} + \frac{0.9}{d} + \frac{0.5}{e}$)

#11($\{1\}, \frac{0.2}{a} + \frac{0.4}{b} + \frac{0.6}{c} + \frac{0.8}{d} + \frac{0.5}{e}$)

#12($\{1, 2\}, \frac{0.1}{a} + \frac{0.2}{b} + \frac{0.6}{c} + \frac{0.8}{d} + \frac{0.4}{e}$)

#13($\{2\}, \frac{0.1}{a} + \frac{0.2}{b} + \frac{0.7}{c} + \frac{0.9}{d} + \frac{0.4}{e}$)

同时生成的 Hasse 图如图 2 所示。

5 结束语

模糊概念格的构建过程比较复杂, 当采用人工方式构建时只能应用于对象和属性较少的情形; 当面对

大量的对象和属性时只有采用计算机构建才可以快速、准确、高效地实现模糊概念格的构建。

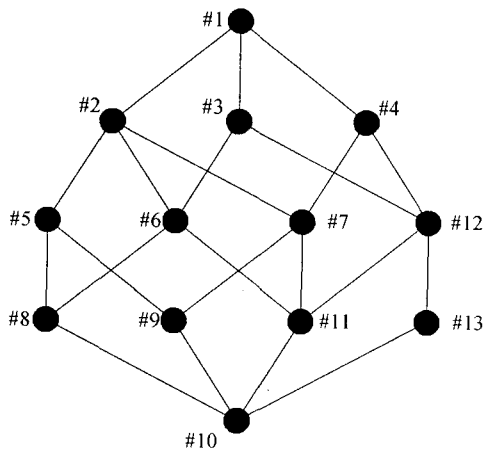


图 2 实例的模糊概念格 Hasse 图

文中基于模糊关系运算,改进了概念格构建的 Bordat 方法,提出了模糊概念格构建的一种新方法。由此可将属性值介于 0 和 1 之间的模糊形式背景结构层次化为模糊概念格,同时绘制出其 Hasse 图。该方法简单明了,易于计算机实现。进一步的研究将致力于建格方法效率的改进、模糊概念格约简和模糊概念格的合并等。

参考文献:

- [1] Quan T, Siu C. Automatic fuzzy ontology generation for semantic web[J]. IEEE Transactions on Knowledge and Data

Engineering, 2006, 18(6): 842-856.

- [2] 陈世权,程里春. 模糊概念格[J]. 模糊系统与数学, 2002, 16(4): 12-18.
- [3] 蒋平,任胜兵,林鹏. 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.
- [4] Belohlavek R, Dvork J, Outrata J. Fast factorization by similarity in formal concept analysis of data with fuzzy attributes[J]. Journal of Computer and System Sciences, 2007, 73(6): 1012-1022.
- [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, 40(9): 77-81.
- [6] 胡明涵,张俐,任飞亮. 模糊形式概念分析与模糊概念格[J]. 东北大学学报: 自然科学版, 2007, 28(9): 1274-1277.
- [7] Ganter B, Wille R. Formal concept analysis: mathematical foundations[M]. Berlin: Springer, 1999.
- [8] 杨强,赵明清. 概念格研究进展[J]. 计算机工程与设计, 2008, 29(20): 5293-5296.
- [9] 赵志坤,李义杰. 对 Bordat 建格算法的一个改进[J]. 中国科技信息, 2007(4): 275-278.
- [10] 刘瑞新,孙士保,秦克云. 基于不同蕴含算子的模糊概念格建格算法研究[J]. 计算机工程与应用, 2007, 43(17): 179-181.
- [11] 张涛,周爱武,谢荣传. 基于概念格和关联规则 Web 个人化系统[J]. 计算机技术与发展, 2008, 18(2): 139-142.
- [12] 陈世权. 模糊性的层次结构与模糊论[J]. 模糊系统与数学, 1999, 13(4): 1-6.

(上接第 126 页)

5 结束语

文中提出了车载激光雷达数据处理软件的设计方案,并对设计过程中的关键问题进行了阐述。该软件将多个测量对象的处理功能集成在一起,并实现了数据的实时处理与显示。降低了用户数据处理的工作量,给用户对测量物质分布的研究工作带来了极大的便利。同时,文中对其他激光雷达系统的数据处理软件的设计有较高的参考价值。

参考文献:

- [1] 吕勇辉. AML-2 车载激光雷达中的信号检测及控制技术[D]. 合肥:中国科学院安徽光机所, 2004.
- [2] 方欣,胡顺星,赵培涛,等. ATS-1 激光雷达实时处理显示系统设计[J]. 计算机工程, 2009, 35(11): 276-279.
- [3] Megie G. Simultaneous Nighttime Lidar Measurements of Atmospheric Sodium and Potassium[J]. Planetary and Space Science, 1978, 26(1): 27-35.

- [4] 杜小勇,张寅超,屈凯峰,等. 车载激光雷达探测低层大气中 NO₂ [J]. 大气与环境光学学报, 2006, 1(2): 97-100.
- [5] Kruglinski D J, Wingo S, Shepherd G. Programing VC++ 6.0[M]. Beijing: Hope Electronic Press, 1999.
- [6] 吴涵. 基于 VC++ 的研究生信息管理系统的设计与实现[J]. 计算机技术与发展, 2006, 16(12): 184-186.
- [7] 柳钰,王明辉. 基于 GDI+ 的快速图形图像绘制技术研究[J]. 现代显示, 2006(11): 66-69.
- [8] 武晓波,王世新,肖春生. Delaunay 三角网的生成算法研究[J]. 测绘学报, 1999, 28(1): 28-35.
- [9] 石宏斌,陈性义,赖祖龙. 基于离散数据的等值线绘制[J]. 工程地球物理学报, 2006, 3(6): 484-487.
- [10] 韦美雁,陈学工. 基于规则格网的等值线光滑算法研究[J]. 现代计算机, 2007(4): 30-32.
- [11] 武晓波,王世新,肖春生. Delaunay 三角网的生成算法研究[J]. 测绘学报, 1999, 28(1): 28-35.
- [12] 于扬,杨泽红,贾培发. 计算机安全监控系统的关键技术研究[J]. 计算机工程, 2007, 33(24): 146-148.