
《智能信息处理》课程作业

多粒度形式概念分析的应用进展

刘程

作业	分数[20]
得分	

2020 年 11 月 12 日

多粒度形式概念分析的应用进展

刘程

(大连海事大学 信息科学与技术学院 大连 116026)

摘要 众所周知,形式概念分析的发展曾经一度遇到瓶颈,主要原因是该理论用于数据分析的核心工具概念格的构建代价是指数级的。令人欣慰的是,粒计算恰好是一个通过信息粒化思想简化复杂不确定性问题求解和进行大数据分析的数学工具,被认为是提高数据分析效率的有效方法,因此一些学者将粒计算引入形式概念分析中以应对概念数据分析出现的低效性,比如,姚一豫从认知信息学和三支决策的角度阐明了粒计算与概念形成相结合的必要性。本文就多粒度形式概念分析的应用进展进行了论述,有关讨论结果将为形式概念分析的粒计算方法的研究与发展提供借鉴。

关键词 形式概念分析;粒计算;信息粒化

Research status and Prospect of three branch concept analysis

Cheng Liu

(Department of Information Science Technology, Dalian Maritime University, Dalian 116026)

Abstract As we all know, the development of formal concept analysis once encountered a bottleneck. The main reason is that the construction cost of concept lattice, the core tool of formal concept analysis, is exponential. It is gratifying that granular computing is just a mathematical tool to simplify the solution of complex uncertain problems and big data analysis through the idea of information granulation. It is considered to be an effective method to improve the efficiency of data analysis. Therefore, some scholars introduce granular computing into formal conceptual analysis to deal with the inefficiency of conceptual data analysis. For example, Yao Yiyu explained the necessity of combining granular computing with concept formation from the perspective of Cognitive Informatics and three branch decision making. This paper discusses the application progress of multi-granular formal concept analysis, and the relevant discussion results will provide reference for the research and development of granular computing methods for formal concept analysis.

Keywords formal concept analysis; granular computing; information granulation.

1 引言

近年来,一些学者又将多粒度引入信息系统中以实现多源异构分布式数据分析。另外,属性粒化也被引入形式概念分析中以支持属性信息的合成与分解。这些研究结果促使学者们从各个角度讨论形式概念分析存在的多粒度问题。比如,文献[1]通过多个形式背景的并置很直观地得到了多粒度形式背景的定义,使得多粒度形式概念分析不再依赖于属性粒化这一抽象工具。在此基础上,文献[2]给出了如何在多粒度形

式概念分析中实现知识跨粒度组合的方法。文献[3]考虑了多粒度形式背景的面向对象和面向属性概念格的相互转化问题。文献[4]则是从各个角度总结分析了传统形式概念分析中能够体现多粒度思想的若干侧面。

对于现有的多粒度形式概念分析研究,仍然存在一些不足。比如,当多粒度形式概念分析进行知识跨粒度层发现时,只能借助于单粒度类属性块之间的粒度组合完成。换言之,单粒度类属性块内部信息均来源于同一个粒度空间且被视作基本的数据元而不能再进行分割,这与

实际应用中很多情况不符。以大学生“高等数学”考试成绩使用问题为例,改卷的时候通常是百分制,而补考取决于考试成绩是否及格,因此被认为是两级制;当评奖学金时,不及格的同学没有参评资格(即不及格的同学被视作一类不再进一步区分),但是及格的同学则需继续比较分数的高低,这意味着评奖学金时不及格的同学的分数应一律标记为“不及格”(两级制),而及格的同学的分数则应标记为百分制。因此,对于“高等数学”这一类属性块,限制属性信息来源于二级制或百分制当中的一种,无法满足评奖学金的要求。为了解决此类问题,本文对类属性块内部信息做进一步分类以引出多粒度类属性块的概念,分析了多粒度类属性块的内部结构,揭示了决策蕴涵随多粒度类属性块粒度粗细变化进行更新的机制。创新之处是推广后的多粒度形式概念分析的跨粒度知识发现方法,能够同时进行类属性块内部信息的粒度组合和类属性块之间的粒度组合。

形式概念分析的粒计算方法,除了庞大的概念个数 $|L \times K|$ 能够得到缓解之外,还可以带来一些新的研究挑战、机遇与应用前景。比如,通过对象粒化或属性粒化,获取数据的多层次概念知识表示与处理方法。此外,概念粒化、加工、转移等研究思路可以将注意力从所有概念成功转移到部分概念、局部认知与关联分析等得到强化,从而有望把所得结果应用于大数据分析、认知学习等热门领域。

2 属性粒化形式概念分析

文献[5]提出了一种基于属性粒化的形式概念分析方法。在该方法中,原形式背景 (G, M, I) 中的每一个属性在不同的粒度下可分解为不同的子属性,这些不同粒度的属性形成了一粒度树,文献[5]借助粒度树以及粒度树中的剪枝方法,从原背景出发导出了不同粒度的形式背景,为在不同粒度下进行概念分析提供了一个可能的框架。

定义 1 设 $K = (G, M, I)$ 为一形式背景, $m \in M$ 。属性 m 的粒度树 T_m 是满足如下条件的一个含有根节点树:

(I) 树中的每一个节点都用一个属性名称来标记,根节点记作 m ;

(2) 对于每一个节点 z , 赋予一对象集合 $\{z\}^\downarrow \subseteq G$, 该对象集 $\{z\}^\downarrow$ 由具有属性 z 的对象所构成;

(3) 若 z_1, z_2, \dots, z_n 是节点 z 的子节点, 则 $\{\{z_1\}^\downarrow, \{z_2\}^\downarrow, \dots, \{z_n\}^\downarrow\}$ 是 $\{z\}^\downarrow$ 的一个划分。

为方便起见, 往往用 z^\downarrow , 而不用 $\{z\}^\downarrow$, 其含义是不变的。

定义 2 粒度树 T_m 中的一剪枝(原文称之为 cut) C 是满足如下条件的一节点集: 对于每一个叶节点 u 而言, 在从 u 到根节点 m 的路径上, 存在唯一的节点 v 属于 C [5]。

例 1 表 1 给出了一基于属性粒化的形式背景, 在该表中, 属性 G 具有两层粒度, 图 1 给出了粒度树 T_G 中的所有剪枝, 即 $\{G\}$ 与 $\{IG, dG\}$ 。注意到 $C = \{G, dG\}$ 并不是粒度树 T_G 的一剪枝, 其原因在于从叶节点 dG 到根节点 G 的路径上, 存在两个节点属于 C 。同理, $\{G, IG\}$ 也不是一剪枝。

表 1 基于属性粒化的形式背景

	<i>L</i>	<i>R</i>	<i>G</i>
<i>a</i>	×		×
<i>b</i>	×		×
<i>c</i>	×		×
<i>d</i>			×
<i>e</i>			×
<i>f</i>	×	×	
<i>g</i>		×	

	<i>L</i>	<i>R</i>	<i>IG</i>	<i>dG</i>
<i>a</i>	×		×	
<i>b</i>	×		×	
<i>c</i>	×			×
<i>d</i>			×	
<i>e</i>				×
<i>f</i>	×	×		
<i>g</i>		×		


```
graph TD
    G1((G)) --> IG1[IG]
    G1 --> dG1[dG]
```

```
graph TD
    G2((G)) --> IG2[IG]
    G2 --> dG2[dG]
```

图 1 表 1 中的粒度树以及剪枝 $\forall z \in M, (g, m) \in I \Leftrightarrow g \in m^\downarrow$

由定义 1、定义 2 以及例 1 可见, 每一个属性都有唯一的粒度树, 但每一个粒度树上有不同的剪枝。在同一属性粒度 T_m 的剪枝集上定义如下偏序关系: $C_1 = \{y_1, y_2, \dots, y_r\}, C_2 = \{z_1, z_2, \dots, z_s\} \in T_m, C_1 \leq C_2$ 当且仅当 $C_1^\downarrow = \{y_1^\downarrow, y_2^\downarrow, \dots, y_r^\downarrow\}$ 是 $C_2^\downarrow = \{z_1^\downarrow, z_2^\downarrow, \dots, z_s^\downarrow\}$ 的细化, 即 $C_1^\downarrow = \{y_1^\downarrow, y_2^\downarrow, \dots, y_r^\downarrow\}$ 是通过对 $C_2^\downarrow = \{z_1^\downarrow, z_2^\downarrow, \dots, z_s^\downarrow\}$ 中每个集合进一步划分后得到的。

设 (G, M, I) 为一形式背景, 对于每一个属性 $m \in M$ 而言, T_m 为其粒度树, C_m 是 T_m 上的一剪枝, 不同属性粒度树上的剪枝可导出如下形式背景 (G, M_c, I_c) , 其中 $M_c = \bigcup_{m \in M} C_m$ 。

在每个属性的粒度树上任意选取一剪枝, 所有这些剪枝之集 C 代表了特定的属性粒化层

次 (Granularity Level)。对于任意两属性粒化层次 C_1 与 C_2 , 用 $C_1 \leq C_2$ 表示 $\forall m \in M, C_{1m} \leq C_{2m}$, 这里 C_{1m} 与 C_{2m} 分别表示属性 m 在两粒化层次中的剪枝。换言之, C_1 是 C_2 的细化。由此, 可在不同的粒化层次上进行概念分析。

在新的形式背景 (G, M_c, I_c) 中, 可定义如下形式的概念生成算子:

$$\forall X \subseteq G, Z \subseteq M_c$$

$$X^{\uparrow c} = \{m \in M_c | \forall x \in X, (x, m) \in I_c\} \quad (13)$$

$$X^{\square c} = \{m \in M_c | \forall g \in G, (g, m) \in I_c \Rightarrow g \in X\} \quad (14)$$

$$X^{\circ c} = \{m \in M_c | \exists x \in X, (x, m) \in I_c\} \quad (15)$$

$$Z^{\uparrow c} = \{g \in G | \forall m \in Z, (g, m) \in I_c\} \quad (16)$$

$$Z^{\square c} = \{g \in G | \forall m \in M_c, (g, m) \in I_c \Rightarrow m \in Z\} \quad (17)$$

$$Z^{\circ c} = \{g \in G | \exists m \in M_c, (g, m) \in I_c\} \quad (18)$$

为以下讨论方便, 引入如下记号: 设 C_1 、 C_2 为满足 $C_1 \leq C_2$ 的两剪枝, 对于 $m_1 \in M_{C_1}, m_2 \in M_{C_2}$, 用 $F_{C_2}(m_1)$ 表示 m_1 在 M_{C_2} 中的父节点, 用 $S_{C_1}(m_2)$ 表示 M_{C_1} 中的子节点集。

3 多粒度形式概念分析研究前景

3.1 多粒度形式概念分析研究现状

李金海针对现有的多粒度形式概念分析的介粒度标记方法无法实现类属性块内部信息的跨粒度层组合, 基于实际应用需要将类属性块内部信息进一步划分为子类, 通过跨粒度层重新组合各子类以提出多粒度类属性块, 在此基础上分析多粒度类属性块的内部结构, 揭示决策蕴涵随多粒度类属性块粒度粗细变化进行更新的规律, 完善了基于多粒度形式概念分析的多层次知识发现理论与方法[6]。

陈希邦首先通过定义一元、二元对象(属性)描述子来构建对象粒和属性粒, 并分别对其描述, 然后通过三种近似描述的精度对比, 验证了该描述方法的可行性、有效性; 在决策形式背景中, 结合经典形式背景以及粒计算理论基础, 讨论了基于属性粒协调决策形式背景的粒描述。粒计算作为目前数据挖掘方向最有效的解决问题的方法之一, 粒计算与形式概念分析理论的结合使得知识发现的过程更细致化, 能够更好的促使读者学习以及理解知识发现过程[7]。

刘之茗在多粒度形式背景存在两种表示方法: 一种是由多粒度标记属性诱导出的多粒度

形式背景, 另一种是由属性粒化诱导出的多粒度形式背景基础上定义可类属性块分割的形式背景, 表明了当原始形式背景为可类属性块分割时多粒度形式背景的两种表示方法能够相互转化。可类属性块分割的研究方法为进一步讨论多粒度形式背景奠定了基础[8]。

3.2 多粒度形式概念分析研究展望

从形式概念分析的粒计算方法的主要研究内容来看, 这一领域已取得一些初步的研究成果[9]。但是, 要将形式概念分析的粒计算方法发展成一个有前景的研究方向, 还要继续突破领域内挑战性问题。从粒计算层面统一粗糙集与概念格的研究非常巧合, 粗糙集与概念格是同一年被提出的, 而且有学者已阐明这两个理论实际上存在诸多的互补性。姚一豫是较早思考粗糙集与概念格相结合的学者之一, 这方面他的主要贡献是将粗糙算子引入概念的形式化描述中, 增强经典概念的应用分析能力; 此外, 他把概念外延与内涵运用到粗糙可定义集的刻画上。

众所周知, 粗糙集与概念格均是粒计算的一种有效方法。不管从构造性分析, 还是从公理化描述, 将粗糙集与概念格在粒计算层面上统一起来研究是一个值得深入思考的问题。基于“众包”思想的粒概念认知学习随着互联网的快速发展, 人们发现越来越多的概念(如, 人物, 事件, 事物)通常由成千上万的大众网络共同完成认识。这与互联网概念“众包”的思想不谋而合, 因此尝试研究基于“众包”思想的粒概念认知学习是非常实用的。这也是粒概念学习领域今后可能走向实际应用的一个契机。当然, 这个问题存在一些挑战。比如, 如何判断每个信息源提供的信息是否真实? 还有, 当信息源提供的信息不一致时, 如何处理? 是弃之不用, 还是对它们进行清洗, 需要根据实际情况作出恰当的选择。如果清洗它们, 应该选取何种数据清洗方法比较合适? 这些在技术上都要突破。类似于粗糙集知识更新研究, 现实中形式背景的对象个数、属性个数、属性取值等通常都处于动态环境下, 所以概念格中的粒约简问题也需考虑知识更新研究。概念格中的粒约简动态更新研究, 包括形式背景中数据更新引

起的基本粒改变,粒约简的变化规律,以及决策形式背景下的粒约简更新机制等。尽管粗糙集中这方面已取得诸多研究经验,但在概念格中有关约简更新的研究结果却寥寥无几。究其原因这是由于概念格中的粒约简动态更新机制相对较为复杂。因此,这个研究方向近期迫切需要取得突破性进展。

传统形式概念分析对形式背景几乎无约束,或最多要求其满足正则性。但是,多粒度形式概念分析对形式背景的数据结构要求较为严格,需要遵守类属性块的可分割性,这也是多粒度标记形式概念分析区分于属性粒化多粒度形式概念分析的关键所在。总之,如何对类属性块内部信息进行跨粒度组合,如何对类属性块之间实施粒度组合,以及如何从全局角度出发去寻找合适的粒度组合,这些都是多粒度形式概念分析当前的热点问题。

4 总 结

形式概念分析的粒计算方法的主要研究内容包括: Galois 连接的粒计算模型,对象粒化,属性粒化,关系粒化,关系诱导的概念粒化,粒规则,粒约简,粒概念,粒概念学习,概念粒(计算)系统等方向,该领域存在的挑战性问题较多。需要指出的是,不管是研究方向,还是挑战性问题,本文仅仅只是选择其中具有代表性的一部分给予介绍或归纳,并不涵盖所有的研究方向与问题。

本文从经典形式背景的角度给出了属性粒化形式概念分析定义,为多粒度标记思想引入到形式概念分析带来了更直观的认识,有助于摆脱长期从多粒度标记信息系统的正向尺度化间接理解多粒度标记思想的障碍,有益于今后搭建多粒度粗糙集与多粒度概念格之间的研究渠道。此外,分析了多粒度形式概念分析的研究现状和未来的发展方向,为将来进一步讨论多粒度概念形式分析奠定了理论基础。

参 考 文 献

- [1] 李金海,吴伟志,邓硕.形式概念分析的多粒度标记理论[J].山东大学学报(理学版),2019,54(2):30-40.LI Jinhai,WU Weizhi,DENG Shuo.Multi-scale theory in formal concept analysis[J].Journal of Shandong University (Natural Science),2019,54(2):30-40.
- [2] 李金海,李玉斐,米允龙,等.多粒度形式概念分析的介粒度标记方法[J].计算机研究与发展,2020,57(2):447-458.LI Jinhai,LI Yufei,MI Yunlong,et al.Meso-granularity labeled method for multi-granularity formal concept analysis[J].Journal of Computer Research and Development,2020,57(2):447-458.
- [3] SHE Yanhong,HE Xiaoli,QIAN Ting,et al.A theoretical study on object-oriented and property-oriented multi-scale formal concept analysis[J].International Journal of Machine Learning and Cybernetics,2019,10(11):3263-3271.
- [4] QI Jianjun,WEI Ling,WAN Qing.Multi-level granularity in formal concept analysis[J].Granular Computing,2019,4(3):351-362.
- [5] Belohlavek R, De Baets B, Konecny J.Granularity of attributes in formal concept analysis[J].Information Sciences, 2014, 260: 149-170.
- [6] 李金海,贺建君,吴伟志.多粒度形式概念分析的类属性块优化[J].山东大学学报(理学版),2020,55(05):1-12.
- [7] 陈希邦.基于粒计算与形式概念分析的面向对象(属性)概念的粒描述[D].江西理工大学,2020.
- [8] 刘之茗,李金海.多粒度形式背景的表示[J].海南热带海洋学院学报,2019,26(05):59-64.
- [9] 李金海,吴伟志.形式概念分析的粒计算方法及其研究展望[J].山东大学学报(理学版),2017,52(07):1-12.