

《智能信息处理》课程作业

## 多粒度形式概念分析

张贺

作业	分数[20]
得分	

2020 年 11 月 12 日

# 多粒度形式概念分析

张 贺

(大连海事大学 信息科学技术学院, 辽宁 大连 116026)

**摘 要** 形式概念分析是一种数据分析和规则提取的方法, 已在数据挖掘和机器学习等领域得到广泛应用。形式概念分析中概念格构造的指数时间复杂度在一定程度上导致数据处理效率不高。粒计算可以将复杂问题转化为简单问题。因而粒计算方法被引入传统形式概念分析方法中以更加有效的处理和分析数据。而多粒度计算的引入利于对数据进行多层次的知识发现, 进一步改善形式概念分析。

**关键词** 形式概念分析; 粒计算; 多粒度计算

## Multi-granularity formal concept analysis

Zhang He

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

**Abstract** Formal concept analysis is a method of data analysis and rule extraction, which has been widely used in data mining and machine learning. The exponential time complexity of the concept lattice construction in formal concept analysis leads to inefficient data processing to a certain extent. Granular computing can transform complex problems into simple ones. Therefore, granular computing methods are introduced into traditional formal conceptual analysis methods to process and analyze data more effectively. The introduction of multi-granularity computing facilitates multi-level knowledge discovery of data and further improves formal concept analysis.

**Keywords** formal concept analysis, granular computing, multi-granularity computing

### 1 引言

形式概念分析<sup>[1]</sup>是由 Wille 在 1982 年提出的一种建立在数学基础上的知识表示和知识发现的工具。形式概念分析的核心工具是概念格。形式概念分析的大致思路就是先构建形式背景, 然后构造概念格。相应的, 对于形式概念分析的改进, 也分别聚焦在两个层面, 形式背景的构建和概念格的构造, 但是概念格的构造算法的改进本质上并没有改变其指数时间复杂度。而对于形式背景构建层面的改进, 则引入了粒计算思想。

所谓形式概念分析的粒计算方法, 是指引入粒计算方法解决问题的基本思想, 分析形式背景中对象与属性之间的相互关系<sup>[2]</sup>。和改进概念格构造不同, 粒计算方法依据具体问题对形式背景进行粒化处理, 优化概念个数。和传统的形式概念分析相比, 形式概念分析的粒计算方法不需要计算所有概念, 只要根据具体问题找出部分概念。

多粒度形式概念分析是形式概念分析粒计算

方法的一个热门的研究方向。多粒度计算的引入, 通过对形式背景进行不同粒度的粒化, 进行概念分析, 能够获取数据的多层次概念知识表示和处理方法<sup>[2]</sup>。

### 2 形式概念分析的粒计算方法

本节简要介绍形式概念分析的粒计算方法的实现, 包括对形式背景的对象粒化、属性粒化、关系粒化等<sup>[2]</sup>。

对象粒化是通过粒化形式背景  $K=(U, A, I)$  的对象集  $U$  得到对象粒, 相当于对对象集进行子集划分, 使得研究目标从单一对象变成对象粒。通过对象粒化的处理, 形式背景得到简化, 依据形式背景构造的概念格比原概念格更加简洁, 利于更好的数据分析。

属性粒化和对象粒化在实现时相似。属性粒化是通过对属性集进行子集划分, 对属性进行合并得到属性粒, 实现知识的粗细转换。通过属性粒化可以进行不同知识层面的分析。

对于关系粒化，对象粒化和属性粒化都可看成是其特例。关系粒化相当于是从对象和属性两个角度进行粒化处理，但是又略有不同。一个形式背景就是一个关系，从关系角度出发，用粒化方法对原始形式背景进行处理，可以得到多个或多层次关联数据。

以具体问题为导向，用粒计算方法对原始形式背景进行粒化处理，预处理后的形式背景更简单，由此构造的概念格也相对于原始概念格更加简洁有效。本节只介绍这三种粒化处理，对于其他处理方法，见参考文献[2]。

3 形式概念分析的多粒度计算

通过不同程度粒化可以实现不同程度的知识粗细转换，达到在不同粒化程度上的形式背景上进行概念分析。本节主要重点关注三种引入多粒度计算的形式概念分析方法，而这四种方法的本质不同在于形式背景构建，因此下面每一小节的重点聚集在形式背景的构建。

3.1 基于属性粒化的形式概念分析方法

文献[3]中提出了一种基于属性粒化的形式概念分析方法。此方法对原形式背景中的属性进行不同粒度的分解，然后依据分解粒度构造粒度树，最后通过对粒度树的剪枝操作构造不同粒度的形式背景，以进行不同粒度的概念分析。

表 1 给出一个未进行属性粒化的传统形式背景示例，对象集为{a, b, c, d, e, f, g}，属性集为{X, Y, Z}。

表 1.传统形式背景示例  $K_1$

	X	Y	Z
a	1	0	1
b	1	0	1
c	1	0	1
d	0	0	1
e	0	0	1
f	1	1	0
g	0	1	0

对表 1 中的属性 Z 分解粒化为{Z1, Z2}，得到如表 2 中的基于属性粒化的形式背景。

{Z}和{Z1, Z2}相当于两种粒度，原始形式背景和表二中的形式背景即相当于两种粒度的形式背景。根据具体应用也可以进一步的粒化，方法

相同。在新的形式背景下可以构造新的概念格，进行不同层次的概念分析。

表 2. 基于属性粒化的形式背景示例  $K_1'$

	X	Y	Z1	Z2
a	1	0	1	0
b	1	0	1	0
c	1	0	0	1
d	0	0	1	0
e	0	0	0	1
f	1	1	0	0
g	0	1	0	0

3.2 基于属性聚类的形式概念分析方法

文献[4]种提出了一种基于属性聚类的形式概念分析方法。该方法是依据问题需求或先验知识将若干属性聚集在一起形成更高一级的新属性。在原始形式背景的基础上，对象集不变，属性集变为聚集后新的属性集，形成新的形式背景。这样实现不同粒度的概念分析。

表 3. 原始形式背景示例  $K_2$

	X	Y	Z	U	V
a	1	0	1	0	1
b	0	0	1	0	0
c	1	0	0	1	0
d	0	0	1	0	0
e	0	0	0	1	0
f	0	1	0	0	0
g	0	1	0	0	1

表 4.基于属性聚类的形式背景示例  $K_2'$

	{X,Y}	Z	{U,V}
a	1	1	1
b	0	1	0
c	1	0	1
d	0	1	0
e	0	0	1
f	1	0	0
g	1	0	1

表 3 和表 4 给出了一个基于属性聚类的形式背景示例, 原始的属性集  $\{X, Y, Z, U, V\}$  通过属性聚类形成新的属性集  $\{\{X, Y\}, Z, \{U, V\}\}$ , 构造的如表四的新形式背景。基于新的形式背景可以生成新的概念, 进行概念分析。

通过上述示例, 基于属性粒化的形式概念分析和基于属性聚类的形式概念分析对属性的粒化过程似乎是两个相反的过程, 但实质上是不同的。

### 3.3 基于经典形式背景的多粒度标记形式背景

文献[5]从经典形式背景的角度给出了多粒度标记形式背景的定义, 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。

该方法借助特定的属性间语义完成, 通过合并低层次粒度标记属性形成高层次粒度标记属性以构造多粒度标记形式背景, 合并过程中要考虑属性间的可融合性, 不能盲目合并。这似乎和属性聚类很相似, 但是属性聚类是基于数学中集合的等价关系实现的, 而这里的多粒度标记是依据属性间的语义关联, 由形式背景所对应的信息系统可以得到属性间的语义关系。

如下表 5 给出一个经典形式背景示例  $K_3$ , 对象集  $U=\{a,b,c,d,e,f\}$  表示六件衣服组成的集合, 属性集  $A=\{x_1, x_2, x_3, x_4, x_5, x_6\}$ , 其中  $x_1$  表示纯白衬衫,  $x_2$  表示白格子衬衫,  $x_3$  表示白条纹衬衫,  $x_4$  表示纯黑外套,  $x_5$  表示黑格子外套,  $x_6$  表示黑条纹外套。表 5 对应的信息系统如表 6, 信息系统属性  $m$  为白色衬衫的三个取值  $\{x_1, x_2, x_3\}$ , 属性  $n$  为黑色外套的三个取值  $\{x_4, x_5, x_6\}$ 。

属性合并后的属性集为  $\{y_1, y_2, y_3, y_4\}$ ,

$y_1$  为纯白衬衫,  $y_2$  为白格子或白条纹衬衫,

$y_3$  为纯黑外套,  $y_4$  为黑格子或黑条纹外套。多粒度标记后的形式背景如表 7。

表 5. 形式背景  $K_3$

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$
a	0	0	1	1	0	0
b	0	1	0	1	0	0
c	0	0	1	1	0	0
d	1	0	0	0	1	0
e	1	0	0	0	0	1
f	1	0	0	0	1	0

表 6.  $K_3$  对应信息系统

	m	n
a	白条纹衬衫	纯黑外套
b	白格子衬衫	纯黑外套
c	白条纹衬衫	纯黑外套
d	纯白衬衫	黑格子外套
e	纯白衬衫	黑条纹外套
f	纯白衬衫	黑格子外套

表 7. 多粒度标记形式背景  $K_3'$

	$y_1$	$y_2$	$y_3$	$y_4$
a	0	1	1	0
b	0	1	1	0
c	0	1	1	0
d	1	0	0	1
e	1	0	0	1
f	1	0	0	1

## 4 结语

本文重点关注了三种多粒度形式概念分析方法, 并举例简单介绍三种方法中粒化原理, 和形式背景的构建。多粒度计算的引入改变了经典形式背景的构建, 间接也改变了概念生成和表示。多粒度标记通过不同程度粒化形式背景中的属性可以实现不同程度的知识粗细转换, 达到在不同粒化程度的形式背景上进行概念分析。

## 参考文献

- [1]. Wille R. Restructuring lattice theory: an approach based on hierarchies of concepts[M] // Rival I, ed. Ordered Sets. Dordrecht: Reidel, 1982: 445—470
- [2]. 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(7): 1-12.
- [3]. Belohlavek R, De Baets B, Konecny J. Granularity of attributes in formal concept analysis[J]. Information Sciences, 2014, 260: 149-170.
- [4]. Liu Z, Li B, Pei Z, et al. Formal concept analysis via multigranulation attributes[C]//2017 12th International Conference on Intelligent Systems and Knowledge Engineering, 2017: 1-6.
- [5]. 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): 1-11.