

# 三支决策思想下概念格的分析与构建

韩炜

作业	分数[20]
得分	

2020年11月25日

# 三支决策思想下概念格的分析与构建

韩炜

(大连海事大学信息科学与技术学院 大连 116026)

**摘要:** 概念格结构模型是形式概念分析中的核心数据结构,而三支决策是一种新兴的符合人类认知的概念模式,是处理不确定性决策的有效方法,利用不完备形势背景下两者的比较来体现出三支决策思想在概念格理论中的重要意义。将三支决策中三分而治的思想理论方法化而与经典概念格相结合进行研究,利用碰撞产生的新的问题与思想可能引申出经典概念认知学习触及不到的问题,不断挖掘之下开辟出一个新的研究前景与方向。

**关键字:** 三支决策;概念格;不完备形势背景;研究展望

中图法分类号 TP311.20 DOI号 10.3969/j.issn.1001-3695.2014.01.030

## Analysis and construction of concept lattice based on Three-way Decision

Han Wei

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

**Abstract:** The concept lattice structure model is the core data structure in the formal concept analysis, and the three decisions are an emerging conceptual model which conforms to the human cognition. It is an effective way to deal with the uncertainty decision. Using the conceptual lattice of the Incomplete Formal Context to reflect the importance of three-way decision in conceptual lattice theory. The combination of the trisecting-and-acting of thinking and management of the three-way decision and the combination of classical concept lattice research, the use of the collision of new problems and ideas may lead to the classic concept of cognitive learning can not reach the problem, continue to tap the Under the development of a new research prospects and direction.

**Key words:** Three-way Decision; Concept lattice; Incomplete Formal Context; research forecast

## 1 引言

形式概念分析(Formal Concept Analysis, FCA)主要研究内容是形式背景及以形式背景为基础形成的形式概念和概念格。FCA作为数据分析的一种重要方法,以形式背景中属性和对象的二元关系为基础建立形式概念,并分析其隐含信息,建立概念格。FCA在数据挖掘、知识管理、Web服务等众多领域有着广泛的应用<sup>[1]</sup>。再经多人研究与实践,形式概念的构建算法<sup>[2]</sup>已比较成熟。

FCA支持二支决策即考虑接受和拒绝两种选择,此时不接受就等同于拒绝,不拒绝就等同于接受。然而实际应用中却非如此,例如日常生活中的

投票可以选择赞成、反对或弃权,而在医疗诊断中则是治疗、不治疗或进一步诊断,与之对应的理论即三支决策。

三支决策理论是为具有三个选项的决策问题提出的一个与学科无关的统一框架<sup>[3]</sup>。作为二支决策理论的一个推广,三支决策理论根据某个评判准则将决策分为接受、拒绝、不承诺3大类,近些年已有许多与其相关的研究及应用。有丰富的语义信息。同时概念格能清晰地表达概念与概念之间的层次结构,即本体的阶层,本文基于形式概念的构造和形式背景的构成,从三支决策的视角分析已有的概念格模型,探究它们之间的内在联系,展现三支决策思想的优势及发展前景。

## 2 相关概念定义

### 2.1 概念格

概念格是以形式概念为元素并配备一个偏序关系得到的完备格。

**定义 1<sup>[4]</sup>**: 形式背景

称一个 3 元组  $(U, V, R)$  为形式背景, 则其中  $U = \{x_1, x_2, \dots, x_n\}$  为对象集合,  $V = \{v_1, v_2, \dots, v_n\}$  为属性集合,  $R$  是从  $U$  到  $V$  的一个二元关系, 用  $xRv$  或者  $(x, v) \in R$  来表示对象  $x$  具有属性  $v$ 。

**定义 2<sup>[5]</sup>**: 概念诱导算子\*

$$X^* = \{v \in V \mid \forall x \in X(xRv)\}$$

$$A^* = \{u \in U \mid \forall a \in A(uRa)\}$$

其中  $X \subseteq U$  是对象子集,  $A \subseteq V$  是属性子集,  $(xRv)$  表示对象  $x$  具有属性  $v$ 。  $X^*$  为  $X$  中所有对象共同具有的属性集合,  $B^*$  为共同具有  $B$  中所有属性的对象集合。

**定义 3<sup>[4]</sup>**: 形式概念

设  $(U, V, R)$  是一个形式背景。如果一个二元组  $(X, A)$  满足  $X^* = A$  且  $X = A^*$ , 称  $(X, A)$  是一个形式概念 (简称概念)。其中  $X$  称为概念的外延,  $A$  称为概念的内涵。

**定义 4<sup>[4]</sup>**: 概念格

我们将形式概念之间的偏序关系  $\leq$  约定为

$$(X_1, A_1) \leq (X_2, A_2) \Leftrightarrow X_1 \subseteq X_2 \Leftrightarrow A_1 \supseteq A_2$$

那么所有概念连同上述偏序关系构成一个完备格, 称为概念格。

截止目前, 概念格研究已取得众多的研究成果, 比如概念格构造, 概念格规则提取, 概念三元格, 以及概念格与模糊集、粗糙集、粒计算、认知计算等的各种结合。

### 2.2 三支决策

三支决策是姚一豫于 2009 年提出的以“三分而治(trisecting-and-acting) 为主要思想的一个有效决策理论<sup>[6]</sup>, 即将一个整体(论域)分为 3 个部分, 并采取有效的策略处理这 3 个部分。为了更好地论述

三支决策理论处理问题的核心思想, 假设问题空间 (即论域) 为  $U$ , 问题描述评价函数为  $f_{A_i}(\bullet)$  (即由属性子集  $A_i$  产生的函数), 三分结果为 Part 1, Part 2, Part 3, 治略分别为 Strategy 1, Strategy 2, Strategy 3, 并且解空间为  $\Pi$ , 那么三支决策理论解决问题的一般框架就如图 1 所示:

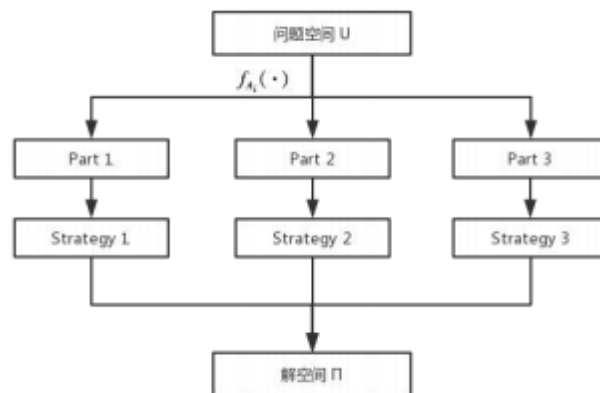


图 1 三支决策解决问题的一般框架

其中第 1 层(自上而下)为问题空间层, 第 2 层为三分结果层, 第 3 层为治略层, 第 4 层为解空间层。实际上, 整个框架充分体现了“分而治之”解决问题的基本思想。自顶而下相当于将整个问题子问题化, 通过分解模式解决问题; 自底而上相当于多策略局部优化, 最后逼近全局问题解。

实际中很多复杂问题涉及的“三分”与“治略”还需在一般框架的基础上继续延伸或扩展。比如, 三支决策理论中的序贯三支决策问题, “三分”与“治略”这两层就需要循环往复, 通常循环往复的原因是随着时间推移信息不断更新, 使得解决问题的约束条件在增加, 从而优化过程出现迭代致使目标结果更加具体。

三支决策就是在这种思想下根据一组标准对论域进行划分, 得到 3 个两两不交的区域, 分别称为正域、负域和边界域, 通过评估对象满足标准的程度, 即满足度, 可以进行如下三支决策<sup>[7]</sup>。

1) 接受决策。如果对象对于某标准的满足度大于或等于某一程度, 采取接受决策, 并认为该对象满足标准。

2) 拒绝决策。如果对象对于某标准的满足度小于或等于另一程度, 采取拒绝决策, 并认为该对象不满足标准。

3) 不承诺决策。如果对象的满足度既不适合采取接受决策也不适合采取拒绝决策, 采取不承诺决策或延迟决策, 并对其进行进一步观察。

### 3 三支决策思想下概念格的比较

本节从三支决策的视角分析和比较不完备形式背景和经典形式背景中的概念格模型来体现三支决策思想在概念格理论中的重要意义。

#### 3.1 不完备形式背景

经典形式背景中的二元关系确定，对象和属性之间要么具有关系，要么不具有关系。然而现实中，对象和属性之间的关系有时无法确定，因此提出不完备形式背景。

**定义 5<sup>[8]</sup>：**不完备形式背景

称四元组  $IC = (U, A, V, T)$  为一个不完备形式背景，其中  $U = \{x_1, x_2, \dots, x_n\}$  为对象集合， $A = \{a_1, a_2, \dots, a_m\}$  为属性集合， $V = \{1, ?, 0\}$  为值域， $T \subseteq U \times A \times V$  为一个三元关系， $(x, a, 1) \in T$  为对象  $x$  具有属性  $a$ ， $(x, a, 0) \in T$  为对象  $x$  不具有属性  $a$ ， $(x, a, ?) \in T$  为无法确定对象  $x$  是否具有属性  $a$ 。表 1 给出一个具体的不完备形式背景的例子。

表 1 不完备形式背景 IC

$U$	$a$	$b$	$c$	$d$	$e$
1	0	1	1	1	0
2	?	0	1	0	?
3	1	0	0	0	?
4	1	?	?	0	1

在不完备形式背景中，对象和属性之间的关系包括具有、不具有、无法确定 3 种情况，分别使用“1”、“0”和“?”表示。从这个角度来看，不完备形式背景可以看成是三支决策的一种具体表现，其可以看成是一种三支形式背景。而在经典形式背景中，对象  $x$  和属性  $a$  之间的关系仅包括具有和不具有 2 种情况，可以看成是一种二支形式背景。

基于粗糙集理论在不完备形式背景中给出一种近似概念的构造方法。

#### 3.2 粗糙近似概念及其形式背景

**定义 6<sup>[9]</sup>：**粗糙近似概念

设  $(U, A, V, T)$  为一个不完备形式背景， $P(U)$  和  $P(A)$  分别表示  $U$  和  $A$  的幂集，

$$\forall X \in P(U), (B, C) \in P(A) \times P(A)$$

及一对派生算子

$$\square : P(U) \rightarrow P(A) \times P(A),$$

$$\diamond : P(A) \times P(A) \rightarrow P(U),$$

定义如下：

$$X^\square = (\underline{R}(X), \overline{R}(X)),$$

$$(B, C)^\diamond = \{x \in U \mid (B, C) \subseteq (\underline{R}(X), \overline{R}(X))\}$$

$$= \{x \in U \mid B \subseteq \underline{R}(X), C \subseteq \overline{R}(X)\}$$

其中

$$\underline{R}(X) = \{a \in A \mid (x, a, 1) \in T, \forall x \in X\},$$

$$\overline{R}(X) = \{a \in A \mid (x, a, 1) \in T \text{ or } (x, a, ?) \in T, \forall x \in X\}$$

可以证明， $(\square, \diamond)$  是  $(P(U), \subseteq)$  和  $(P(A) \times P(A), \subseteq)$  之间的一对 Galois 连接。

如果  $X^\square = (B, C)$  且  $(B, C)^\diamond = X$ ，称  $(X, (B, C))$  为一个近似概念。

从粗糙集的角度上看， $\underline{R}(X)$  表示  $X$  中对象肯定共同具有的属性，可以看成是  $X$  内涵的下近似；

$\overline{R}(X)$  表示  $X$  中所有对象可能共同具有的属性，可以看成是  $X$  内涵的上近似。因此，本文称  $(\square, \diamond)$  诱导的概念为粗糙近似概念。

从三支决策的视角来看，

$$\forall X \in P(U), X^\square = (\underline{R}(X), \overline{R}(X))$$

诱导属性集  $A$  上的一个三划分

$$\{\underline{R}(X), \overline{R}(X) - \underline{R}(X), A - \overline{R}(X)\}$$

因此粗糙近似概念的构造是三支决策的一种具体表现，粗糙近似概念可以看成是一种三支概念。而在经典概念格模型中， $\forall X \subseteq U$ ， $X^*$  和  $A - X^*$  构成  $A$  的一个二划分，因此经典形式概念的构造可以看成是二支决策的一种具体表现，经典形式概念可以看成是一种二支概念。

接下来探究粗糙近似概念格模型与经典概念格模型之间的内在联系，为此首先构造一个经典形式背景用于近似不完备形式背景。

**定义 7：**粗糙近似形式背景

不完备形式背景  $IC = (U, A, V, T)$  的粗糙近似形式背景是一个经典形式背景

$$C_{\text{Rough}} = (U, A \times \{1, 2\}, R),$$

其中， $\forall a \in A$

$$(x, (a, 1)) \in R \Leftrightarrow a \in \underline{R}(X)$$

$$(x, (a, 2)) \in R \Leftrightarrow a \in \overline{R}(X)$$

表 2 给出表 1 的粗糙近似形式背景。它可以看成是 2 个经典形式背景  $C_1$  和  $C_2$  的下位形式背景，即

$$C_{\text{Rough}} = C_1 \cup C_2,$$

其中  $C_1$  可以看成是表 1 的下近似， $C_2$  可以看成是

表 1 的上近似。

表 2 粗糙近似形式背景  $C_{\text{Rough}}$

$U$	$(a,1)$	$(b,1)$	$(c,1)$	$(d,1)$	$(e,1)$	$(a,2)$	$(b,2)$	$(c,2)$	$(d,2)$	$(e,2)$
1	0	1	1	1	0	0	1	1	1	0
2	0	0	1	0	0	1	0	1	0	1
3	1	0	0	0	0	1	0	0	0	1
4	1	0	0	0	1	1	1	1	0	1

### 3.3 三支思想与经典概念格的比较

文献<sup>[10]</sup>中给出一种通过  $L(C_1)$  和  $L(C_2)$  构造粗糙近似概念格的方法, 利用同构映射的方法可以证得不完备形式背景的粗糙近似概念格同构于它的粗糙近似形式背景的经典概念格。

图 2 和图 3 分别给出表 1 和表 2 对应的粗糙近似概念格和经典概念格的 *Hasse* 图, 可以发现, 两者同构。

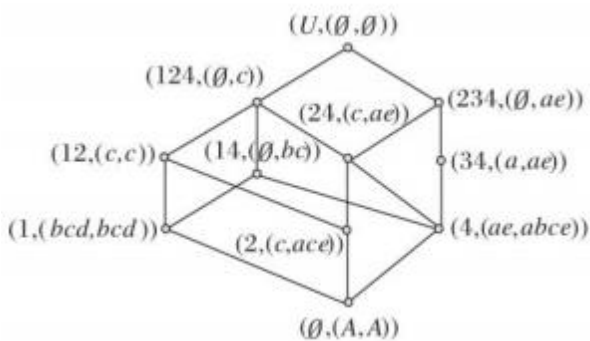


图 2 IC 的粗糙近似概念格  $RL(IC)$

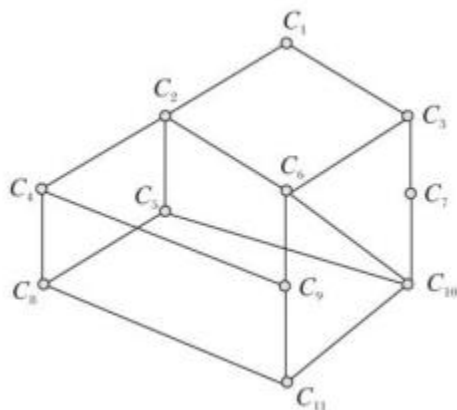


图 3  $C_{\text{Rough}}$  的经典概念格  $L(C_{\text{Rough}})$

在图 3 中, 各个概念分别如下:

- $C_1: (U, \emptyset)$ ,  $C_2: (124, (c, 2))$ ,
- $C_3: (234, (a, 2)(e, 2))$ ,
- $C_4: (12, (c, 1)(c, 2))$ ,
- $C_5: (14, (b, 2)(c, 2))$ ,
- $C_6: (24, (a, 2)(c, 1)(e, 2))$ ,
- $C_7: (34, (a, 1)(a, 2)(e, 2))$ ,
- $C_8: (1, \{b, c, d\} \times \{1, 2\})$ ,
- $C_9: (2, (c, 1)(a, 2)(c, 2)(e, 2))$ ,
- $C_{10}: (4, \{a, e\} \times \{1, 2\} \cup \{b, c\} \times \{2\})$ ,
- $C_{11}: (\emptyset, A \times \{1, 2\})$ .

尽管两者同构, 但与经典概念格相比, 粗糙近似概念格存在如下优势。

1) 数据存储所占空间小。假设不完备形式背景的大小为  $n \times m$ , 那么它诱导的经典形式背景大小为  $n \times 2m$ 。此外, 虽然经典概念格和粗糙近似概念格的构造算法框架相同, 两者的计算复杂度相同, 但是由于粗糙近似形式背景更大, 导致经典概念格  $L(C_{\text{Rough}})$  计算时间远大于粗糙近似概念格  $RL(IC)$ 。

2) 属性约简(保持概念格结构不变的最小属性子集)更简洁。例如通过计算可以得到表 1 中有 1 个属性约简:  $\{a, b, c\}$ , 而其对应的表 2 中有 2 个属性约简, 分别为:

- $\{(a, 1), (c, 1), (a, 2), (b, 2), (c, 2)\}$ ,
- $\{(a, 1), (c, 1), (b, 2), (c, 2), (e, 2)\}$ 。

对比可以发现,粗糙近似概念格的属性约简不但个数更少,而且约简更简洁。主要有 2 个原因:

- 不完备形式背景中的属性个数本身就少于其所诱导的经典形式背景中的属性个数;
- 构造概念格时,不完备形式背景中的一个属性  $a$  发挥的作用等价于其对应的经典形式背景中的 2 个属性  $(a, 1)$  和  $(a, 2)$  发挥的作用。进行属性约简时,删除不完备形式背景中的  $a$  相当于删除其所对应的经典形式背景中的  $(a, 1)$  和  $(a, 2)$ 。

### 3.4 总结与拓展

粗糙近似概念主要是从共同具有和可能共同具有的角度刻画知识。现实中有时需要从共同具有和共同不具有的角度刻画知识。例如在文献检索时,利用形如“关键词 1 and 关键词 2 and not 关键词 3”的检索公式可以更准确地获得检索结果。为此,在不完备形式背景中推广 Qi 提出的三支概念格模型<sup>[11]</sup>,得到对象诱导的近似概念格(OE 概念格)和属性诱导的近似概念格(AE 概念格)。它们与经典概念格相比也存在着类似于粗糙近似概念格相对于经典概念格的优势。

对此,祁建军在 2017 年 2 月借鉴形式概念分析中构建形式概念的 CbO 算法的思想,发表了一种构建三支概念的算法 CbO3C<sup>[12]</sup>。并通过使用基于部分闭包的正则检测和失效正则检测进行剪枝,使用约简条件排除非核心的三支概念,使用位操作实现集合运算等技术来提高效率,降低内存开销。

## 4 研究展望

通过前文对概念格与三支决策相结合的一些尝试及相关研究现状,从三支概念分析与三支概念学习两个层面对进一步研究方向进行讨论。

### 4.1 三支概念分析中的挑战性问题

#### 4.1.1 概念间相似性度量问题

相似性度量作为所有概念分析中的基本问题,对概念格的合并与简化有非常大的影响,但三支概念之间的相似性度量与经典概念<sup>[13]</sup>中的有所不同,

由于边界域的引入而带来数据的不一致性以及模糊值与缺失值带来的不确定因素,都使相关算法设计变得非常困难。

#### 4.1.2 各种三支概念之间的比较

三支概念分析可进一步区分为完备形式背景下的三支概念分析<sup>[14]</sup>与不完备形式背景下的三支概念分析<sup>[15]</sup>。他们具有不同的语义,前者强调共性属性(共同具有)与共性非属性(共同不具有),上一节我们详细讨论了前者,而后者侧重缺失信息的近似描述。而作为重点难点的不完备形式背景下的三支概念分析目前有三种模型,即不确定概念模型,近似概念模型以及部分已知概念模型。他们之间的比较以及思想结合是非常有意义的。

#### 4.1.3 复杂数据下补背景难以确定

由于不完备信息环境下复杂数据对象与属性间关系并不全都是直观确定的,所以它的补背景也难以确定,由于大数据时代概念的膨胀,穷举法构建概念格效率低下,应更多的利用失效检验进行优先筛选,如今已经提出的对象诱导的近似概念格(OE 概念格)和属性诱导的近似概念格(AE 概念格)不一定是解决这个问题的理想方式,三支概念模型有很强的实用背景,这个问题值得深入研究。

### 4.2 三支概念学习中的挑战性问题

#### 4.2.1 “噪声”影响下容错率的提升

理想情况下,随着知识的不断增加,三支决策中有了更多的判断标准,其不确定(不承诺)部分会逐渐减少,这是非常令人满意的结果。然而,实际的认知学习过程中,往往受到“噪声”影响,导致接受域,拒绝域,不承诺域 3 部分相互侵入,结构遭到破坏,在三支概念学习过程中如何减小甚至根除噪声影响变得十分有意义。

#### 4.2.2 异构数据环境下的三支概念学习

异构数据下的三支概念学习分为确定异构数据各自的三支决策和对三支决策结果的集成两部分。所以其核心问题是如何实现对异构数据各自论域的三分以及采用何种方式进行有效的集成?在这一点上智慧来<sup>[16]</sup>在异构形式背景下的概念格构造经验很有大的参考价值。

#### 4.2.3 大数据背景下,与粒计算<sup>[17]</sup>相结合的认知计算

粒就是指一些个体通过不分明关系、相似关系、邻近关系或功能关系等所形成的块。这种处理信息的过程,称信息粒化。可分为 2 大类:一类侧重于不确定性处理;另一类则注重于多粒度计算。粒

计算体现了一种聚类思想，又有专门用于处理不确定性的模型，所以是不是可以与三支决策以及三支概念学习结合，用于优化决策过程，提高决策及学习效率，是一个值得探讨的问题。

## 5 结束语

本文通过将传统的概念格思想与新兴的三支决策思想相结合，举例论述了三支决策思想下的概念格较经典概念格确实有一定的优势，也是更贴近于现实问题的真正模型。但是，目前概念格与三支决策相结合的研究不是很多，不论是模型创建还是算法设计都处于初级研究阶段，很大程度上还依赖于经典概念格的思想。不过，在思想进行拓展的道路上所遇到的新的挑战性问题，一定会推动这门学科不但完善和发展，进而开辟出一片新的前景。

## 参考文献

- [1] Škopljanac-Maćina F, Blašković B. Formal concept analysis—overview and applications[J]. *Procedia Engineering*, 2014, 69: 1258- 1267.
- [2] Andrews S. A 'Best-of-Breed' approach for designing a fast algorithm for computing fixpoints of Galois Connections[J]. *Information Sciences*, 2015, 295: 633-649.
- [3] Yao Y. An Outline of a Theory of Three-Way Decisions[C]// *International Conference on Rough Sets and Current Trends in Computing*. Springer, Berlin, Heidelberg, 2012: 1- 17.
- [4] Wille R. Restructuring lattice theory: An approach based on hierarchies of concepts[C]// *International Conference on Formal Concept Analysis*. Springer, Berlin, Heidelberg, 2009: 314-339.
- [5] Ganter B, Wille R. *Formal Concept Analysis: mathematical foundations*. (translated from the German by Cornelia Franzke) Springer-Verlag[J]. 1999.
- [6] Yao Y. Three-way decisions and cognitive computing[J]. *Cognitive Computation*, 2016, 8(4): 543-554.
- [7] 刘 盾, 李天瑞, 苗夺谦, 等. 三支决策与粒计算. 北京: 科学出版社, 2013
- [8] Burmeister P, Holzer R. On the treatment of incomplete knowledge in formal concept analysis[C]// *International Conference on Conceptual Structures*. Springer, Berlin, Heidelberg, 2000: 385-398.
- [9] Li J, Mei C, Lv Y. Incomplete decision contexts: approximate concept construction, rule acquisition and knowledge reduction[J]. *International Journal of Approximate Reasoning*, 2013, 54( 1): 149- 165.
- [10] 张慧雯, 刘文奇, 李金海. 不完备形式背景下近似概念格的公理化方法[J]. *计算机科学*, 2015, 42(6): 67-70.
- [11] Qi J, Wei L, Yao Y. Three-way formal concept analysis[C]// *International Conference on Rough Sets and Knowledge Technology*. Springer, Cham, 2014: 732-741.
- [12] 汪文威, 祁建军. 三支概念的构建算法[J]. *西安电子科技大学学报 (自然科学版)*, 2017 (2017 年 01): 71-76.
- [13] Wang L, Liu X. A new model of evaluating concept similarity[J]. *Knowledge-Based Systems*, 2008, 21(8): 842-846.
- [14] Qi J, Qian T, Wei L. The connections between three-way and classical concept lattices[J]. *Knowledge-Based Systems*, 2016, 91: 143- 151.
- [15] Yao Y. Interval sets and three-way concept analysis in incomplete contexts[J]. *International Journal of Machine Learning and Cybernetics*, 2017, 8( 1): 3-20.
- [16] 智慧来. 面向异构数据分析的形式概念分析扩展模型[J]. *电子学报*, 2013, 41( 12): 2451-2455.
- [17] Zadeh L A. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic[J]. *Fuzzy sets and systems*, 1997, 90(2): 111- 127.