基于三支区间集前概念的知识提取及应用

毛 华[1,2]， 胥德华[1,2]， 刘 川[1,2]， 郑博雅[1]， 袁晓垒[1]， 杨兰珍[1]

(1. 河北大学 数学与信息科学学院 河北 保定 071002；
2. 河北省机器学习与计算机智能重点实验室 河北 保定 071002)

摘要： 前概念,作为形式概念“弱化”的知识表达形式,常用来提取对象共有的属性与属性所共有的对象这两类知识。区间集概念格是将区间集理论与形式概念相结合后产生的,利用区间集能够反映不确定信息的特性把形式概念拓展为区间集概念。将前概念理论结合三支决策引入区间集概念,提出了对象诱导的三支区间集前概念,即 OE-区间集前概念。在此基础上,讨论了 OE-区间集前概念的格结构,并与已有成果进行比较,给出了相应的构造算法。这些所得理论成果,通过医生总结病例的实践内容加以验证,说明了所定义知识表示形式的有效性和实用性。这将有助于拓展区间集概念理论以及前概念理论的研究深度和应用范围,同时也为知识提取和信息挖掘提供了一种可行的方法。

关键词： 前概念；区间集；区间集概念；三支决策；三支区间集前概念

中图分类号： TP18 **文献标志码：** A **文章编号：** 1671-6841(2025)03-0057-08

DOI： 10.13705/j.issn.1671-6841.2023190

Knowledge Extraction and Application Based on Three-way Interval-set Preconcept

MAO Hua[1,2], XU Dehua[1,2], LIU Chuan[1,2], ZHENG Boya[1], YUAN Xiaolei[1], YANG Lanzhen[1]

(1. *College of Mathematics and Information Science*, *Hebei University*, *Baoding* 071002, *China*;
2. *Hebei Key Laboratory of Machine Learning and Computational Intelligence*, *Baoding* 071002, *China*)

Abstract: Preconcepts, as a form of knowledge expression that weakened formal concepts in formal concept analysis, were used to study practical problems to extract two types of knowledge: common attributes of objects and objects shared by attributes, which were two important types of knowledge that formal concepts cannot extract. The interval-set concept lattice combined the interval-set theory with the formal concept, so the formal concept could be extended to the interval-set concept by using the characteristics of the uncertain information that the interval set can reflect. By introducing the interval-set concept with preconcept theory and three-way decisions, the object-induced three-way interval-set preconcept (namely, OE-interval-set preconcept) was proposed. On this basis, the lattice structure of OE-interval-set preconcept was discussed and then compared with preconcepts, three-way preconcepts and OE-interval-set preconcepts. The algorithm for finding all OE-interval-set preconcepts was constructed using the relationship among preconcepts, three-way preconcepts and OE-interval-set preconcepts. These theoretical results were validated by doctors in summarizing the practical content of cases, demonstrating the effectiveness and practicality of the defined knowledge representation form. This would help expand the research depth and application scope of interval-set concept and preconcept theory and provide a feasible method for knowledge extraction and information mining.

收稿日期： 2023-08-16
基金项目： 国家自然科学基金项目(61572011)
第一作者： 毛华(1963—),女,教授,主要从事三支决策、粒计算、概念格和粗糙集研究,E-mail: mh@hbu.edu.cn。
通信作者： 胥德华(1998—),男,硕士研究生,主要从事形式概念分析、区间集、三支决策研究,E-mail:1486360343@qq.com。

Key words: preconcept; interval-set; interval-set concept; three-way decision; three-way interval-set preconcept

0 引言

由 Wille[1]提出的形式概念分析,已被成功应用到许多领域[2-4]。形式概念分析(formal concept analysis,FCA)通过在形式背景上构建出派生算子来挖掘形式概念,实现对实际问题进行知识提取的平台搭建。然而,形式概念所要求满足的条件过于苛刻。为此,Stahl 等[5]弱化了其条件并提出前概念。如今,前概念的研究已取得不少成果[6-7]。

在许多问题的研究中,由于信息的不确定,一些知识往往不能被精确定义,需要借助一对上下界构成闭区间来对该概念的外延(内涵)进行表示。因此,Yao 提出了区间集理论[8]。区间集概念格是 FCA 和区间集相结合形成的。集合是退化的区间集,因此形式概念也是特殊的区间集概念,形式概念可挖掘到的知识同样适用区间集概念,反之却不一定成立。目前,区间集概念成果丰硕[9-10]。

Yao[11]提出的三支决策(three-way decision, TWD)是基于二支决策发展而来的决策模型。Qi 等[12]首次将三支决策的方法应用到 FCA 中,在形式概念所能提取知识基础上,还能挖掘出对象不共有的属性并且属性不共有的对象,即三支概念。随后,三支前概念[13]和三支区间集概念[14]相继被提出。

从以上各种成果来看,对区间集概念格的研究,已有学者成功将区间集概念和三支概念结合[14],但由于形式概念可弱化为前概念,故可尝试弱化区间集概念为区间集前概念。此外,由三支前概念可知,TWD 可结合前概念,所以区间集前概念可与三支前概念结合得到三支区间集前概念(这其实也可以看成将前概念、区间集和 TWD 相结合得到的结果)。本文主要思想路线如图 1 所示。

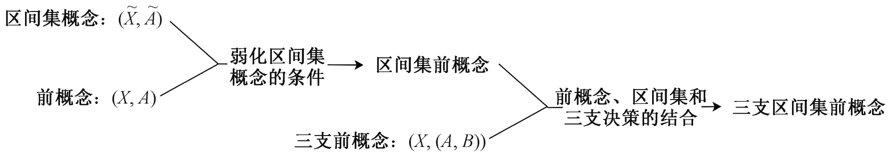

图 1 思想路线

Figure 1 Ideological line

在实际生活中,也的确存在将前概念、区间集和 TWD 相结合进行知识提取的需要。例如,医生在总结病例的时候,会挑选有代表性的病患,将相关治疗过程进行记录。由于病案描述等原因,选取通常在某个范围进行,这也是使用区间集来表述的原因。对这些病患进行症状分析时,医生希望得到这些病患共同具有的关键病症,以便总结出相应的治疗方法。因此,前概念更适合对这类知识进行表示和提取。此外,总结出他们都不具有的病症也是有必要的。因此,还需要借助 TWD 来对病例进行信息挖掘。

为了实践图 1 中的思想路线,将上述例子中涉及提取知识的信息挖掘出来,需要建立统一的数学模型,以便解决更多类似的问题。

1 预备知识

定义 1[8] 设 U 为非空有限论域,对于 $X_l \subseteq X_u \subseteq U$, $\tilde{X}=[X_l,X_u]=\{X\in P(U)\mid X_l\subseteq X\subseteq X_u\}$,称 $[X_l,X_u]$(也可写成 $\tilde{X}$)为定义在 U 上的区间集。记 $P(U)$ 为 U 的幂集,$IP(U)$ 为定义在 U 上区间集全体。

对区间集 $[X_l,X_u],[Y_l,Y_u]\in IP(U)$:

$$[X_l,X_u]\tilde{\cap}[Y_l,Y_u]=[X_l\cap X_u,Y_l\cap Y_u],$$

$$[X_l,X_u]\tilde{\cup}[Y_l,Y_u]=[X_l\cup X_u,Y_l\cup Y_u],$$

$$[X_l,X_u]=[Y_l,Y_u]\Leftrightarrow X_l=X_u \text{ 且 } Y_l=Y_u,$$

$$[X_l,X_u]\tilde{\subseteq}[Y_l,Y_u]\Leftrightarrow X_l\subseteq Y_l \text{ 且 } X_u\subseteq Y_u。$$

定义 2[1] 设 L 是非空集合,称定义在 L 上的二元关系 $\leqslant$ 为偏序关系,若它满足下面三条性质,$\forall x,y,z\in L$,

1)自反性:$x\leqslant x$。

2)反对称性:$x\leqslant y$ 和 $y\leqslant x\Rightarrow x=y$。

3)传递性:$x\leqslant y$ 且 $y\leqslant z\Rightarrow x\leqslant z$。

则集合 L 和偏序关系$\leqslant$构成一个偏序集 $(L,\leqslant)$。$\forall x,y\in L,x\wedge y=\inf\{x,y\}$ 和 $x\vee y=\sup\{x,y\}$ 都存在,则 $(L,\leqslant)$ 构成格。

定义 3[1] 设 $K=(G,M,I)$ 为形式背景,G 为非空对象集,M 为非空属性集,I 为定义在 G 和 M 上的二元关系。对 $\forall g\in G,m\in M,gIm$ 表示对象 g 和属性 m 具有关系 I;而 gI^cm 则表示对象 g 和属性

m 具有关系 I^c, 这里 $(g,m) \in I \Leftrightarrow (g,m) \notin I^c$。

定义 4[1]　在形式背景 $K = (G,M,I)$ 中,对 $\forall X \in P(G)$ 和 $\forall A \in P(M)$,定义一对正算子 $^*: P(G) \to P(M)$ 和 $^*: P(M) \to P(G)$ 为

$$X^* = \{a \mid a \in M, \forall x \in X, xIa\};$$

$$A^* = \{x \mid x \in G, \forall a \in A, xIa\}。$$

负算子 $^{\overline{*}}: P(G) \to P(M)$ 和 $^{\overline{*}}: P(M) \to P(G)$ 为

$$X^{\overline{*}} = \{a \mid a \in M, \forall x \in X, xI^c a\};$$

$$A^{\overline{*}} = \{x \mid x \in G, \forall a \in A, xI^c a\}。$$

定义 5[5]　在形式背景 $K = (G,M,I)$ 中,对 $\forall X \in P(G)$ 和 $\forall A \in P(M)$,(X,A) 是一个前概念,当且仅当 $X \subseteq A^*(\Leftrightarrow A \subseteq X^*)$。记 K 下所有前概念为 $P_c(K)$,$(P_c(K), \leqslant)$ 为前概念格。

定义 6[13]　在形式背景 $K = (G,M,I)$ 中,对 $\forall X \in P(G)$ 和 $\forall A,B \in P(M)$,$(X,(A,B))$ 是一个对象诱导的三支前概念,当且仅当 $X \subseteq A^* \cap B^{\overline{*}}(\Leftrightarrow A \subseteq X^*$ 且 $B \subseteq X^{\overline{*}})$,简称 OE-前概念。记 K 下所有 OE-前概念为 $T_{OE}(K)$,$(T_{OE}(K), \leqslant)$ 为OE-前概念格。

注 1　为方便描述,给定非空有限集合 U 和 V,若它们之间存在双射 $\varphi: U \to V$, 则称这两个集合是同构的,用 $U \cong V$ 来表示。

定义 7[14]　在形式背景 $K = (G,M,I)$ 中,定义一对对象诱导的三支区间集算子 $^{\triangleright}: IP(G) \to IP(M) \times IP(M)$ 和 $^{\triangleleft}: IP(M) \times IP(M) \to IP(G)$ 为对 $\forall \tilde{X} = [X_l, X_u] \in IP(G)$,$\tilde{A} = [A_l, A_u] \in IP(M)$ 和 $\tilde{B} = [B_l, B_u] \in IP(M)$,

$$\tilde{X}^{\triangleright} = (\tilde{X}^*, \tilde{X}^{\overline{*}}) = ([X_u{}^*, X_l{}^*], [X_u{}^{\overline{*}}, X_l{}^{\overline{*}}]),$$

$$(\tilde{A},\tilde{B})^{\triangleleft} = \tilde{A}^* \tilde{\cap} \tilde{B}^{\overline{*}} = [A_u{}^* \cap B_u{}^{\overline{*}}, A_l{}^* \cap B_l{}^{\overline{*}}]。$$

2　*OE*-区间集前概念

2.1　基本定义

定义 8　在形式背景 $K = (G,M,I)$ 中,三元区间集序对 $(\tilde{X},(\tilde{A},\tilde{B}))$ 称为一个对象诱导的三支区间集前概念,当且仅当有 $\tilde{X} \tilde{\subseteq} (\tilde{A},\tilde{B})^{\triangleleft}(\Leftrightarrow (\tilde{A},\tilde{B}) \tilde{\subseteq} \tilde{X}^{\triangleright})$,简称 OE-区间集前概念。其中,$\tilde{X}$ 和 $(\tilde{A},\tilde{B})$ 分别为 OE-区间集前概念 $(\tilde{X},(\tilde{A},\tilde{B}))$ 的外延和内涵。记 K 下所有 OE-区间集前概念为 $W_{OE}(K)$。

下面,通过一个例子来解释定义 8。

例 1　某医生对一天当中接诊的 4 位患者进行病例总结,主要症状有:头疼、呼吸困难、发烧、头晕和咳嗽。该医生通过表 1 展示这 4 位患者的相关信息,若某患者具有某症状,用"×"来标记,否则用"—"标记。

表 1　患者症状信息统计

Table 1　Patient symptom information statistics

姓名	头疼	呼吸困难	发烧	头晕	咳嗽
小芳	×	—	—	×	—
小兰	×	—	×	×	—
小明	—	×	—	×	—
小智	×	—	×	—	×

将表 1 中的 4 位患者从上到下依次用数字 1、2、3、4 来代表,即有对象集 $G_1 = \{1,2,3,4\}$;5 种症状从左到右分别用字母 a、b、c、d、e 来代表,即有属性集 $M_1 = \{a,b,c,d,e\}$;二元关系 I_1 如表 2 所示。因此,将表 1 转化为数学表达,即表 2 所示形式背景。

表 2　一个形式背景 $K_1 = (G_1, M_1, I_1)$

Table 2　A formal context $K_1 = (G_1, M_1, I_1)$

序号	a	b	c	d	e
1	×	—	—	×	—
2	×	—	×	×	—
3	—	×	—	×	—
4	×	—	×	—	×

假设该医生选择了个别有代表性的患者,如 $[2,23]$,表示该医生一定要分析小兰的症状,可能还会分析小明的症状。

由定义 8 可知,$[2,23]^{\triangleright} = ([2,23]^*, [2,23]^{\overline{*}}) = ([d,acd],[e,be])$,表明小兰和小明都有头晕的症状,小兰则有头疼、发烧和头晕的症状,而小兰和小明都没有咳嗽的症状,小兰则没有呼吸困难和咳嗽的症状。

由定义 8,属性区间集 $\tilde{A} \tilde{\subseteq} [d,acd]$,$\tilde{B} \tilde{\subseteq} [e,be]$,从而可得对象区间集 $[2,23]$ 对应所有 OE-区间集前概念,有 $([2,23],([d,acd],[e,be]))$,$([2,23],([\varnothing,acd],[e,be]))$,$([2,23],([d,cd],[e,be]))$,$([2,23],([d,ad],[e,be]))$,$([2,23],([d,d],[e,be]))$,$([2,23],([d,d],[\varnothing,b]))$ 等。

2.2　*OE*-区间集前概念格

定理 1　在形式背景 $K = (G,M,I)$ 中,对 $\forall (\tilde{X}_1,(\tilde{A}_1,\tilde{B}_1)),(\tilde{X}_2,(\tilde{A}_2,\tilde{B}_2)) \in W_{OE}(K)$,定义二元关系"$\leqslant$"为

$$(\tilde{X}_1,(\tilde{A}_1,\tilde{B}_1)) \leqslant (\tilde{X}_2,(\tilde{A}_2,\tilde{B}_2)) \Leftrightarrow \tilde{X}_1 \tilde{\subseteq} \tilde{X}_2 \text{ 且 } (\tilde{A}_1,\tilde{B}_1) \tilde{\supseteq} (\tilde{A}_2,\tilde{B}_2),$$

则 $(W_{OE}(K), \leqslant)$ 为一个偏序集。

证明　由定义 2,易证定理 1 成立。

例 2　在表 2 所示的形式背景 $K_1=(G_1, M_1, I_1)$ 中,序对 $([23,23],([d,d],[e,e]))$ 和 $([234,G_1],([\varnothing,\varnothing],[\varnothing,\varnothing]))$ 都是 OE-区间集前概念。由于 $[23,23]\tilde{\subseteq}[234,G_1]$ 且 $([d,d],[e,e])\tilde{\supseteq}([\varnothing,\varnothing],[\varnothing,\varnothing])$,从而 $([23,23],([d,d],[e,e]))\leqslant([234,G_1],([\varnothing,\varnothing],[\varnothing,\varnothing]))$ 成立。

定理 2　在形式背景 $K=(G,M,I)$ 中,偏序集 $(W_{OE}(K),\leqslant)$ 构成格,称为 OE-区间集前概念格。其中,上确界"$\vee$"和下确界"$\wedge$"分别定义为

$$(\tilde{X}_1,(\tilde{A}_1,\tilde{B}_1))\vee(\tilde{X}_2,(\tilde{A}_2,\tilde{B}_2))=$$
$$(\tilde{X}_1\tilde{\cup}\tilde{X}_2,(\tilde{A}_1\tilde{\cap}\tilde{A}_2,\tilde{B}_1\tilde{\cap}\tilde{B}_2));$$
$$(\tilde{X}_1,(\tilde{A}_1,\tilde{B}_1))\wedge(\tilde{X}_2,(\tilde{A}_2,\tilde{B}_2))=$$
$$(\tilde{X}_1\tilde{\cap}\tilde{X}_2,(\tilde{A}_1\tilde{\cup}\tilde{A}_2,\tilde{B}_1\tilde{\cup}\tilde{B}_2))。$$

证明　由定义 2 可知,需要证明所定义的上下确界都存在,现在以上确界为例进行证明。

1) $(\tilde{X}_1\tilde{\cup}\tilde{X}_2,(\tilde{A}_1\tilde{\cap}\tilde{A}_2,\tilde{B}_1\tilde{\cap}\tilde{B}_2))\in W_{OE}(K)$。

因为 $(\tilde{A}_1\tilde{\cap}\tilde{A}_2,\tilde{B}_1\tilde{\cap}\tilde{B}_2)^{\triangleleft}=(\tilde{A}_1\tilde{\cap}\tilde{A}_2)^{*}\tilde{\cap}(\tilde{B}_1\tilde{\cap}\tilde{B}_2)^{\overline{*}}\tilde{\supseteq}(\tilde{A}_1{}^{*}\tilde{\cup}\tilde{A}_2{}^{*})\tilde{\cap}(\tilde{B}_1{}^{\overline{*}}\tilde{\cap}\tilde{B}_2{}^{\overline{*}})$,又因为 $\tilde{X}_1\tilde{\subseteq}\tilde{A}_1{}^{*}\tilde{\cap}\tilde{B}_1{}^{\overline{*}}$,$\tilde{X}_2\tilde{\subseteq}\tilde{A}_2{}^{*}\tilde{\cap}\tilde{B}_2{}^{\overline{*}}$,从而 $\tilde{X}_1\tilde{\cup}\tilde{X}_2\tilde{\subseteq}(\tilde{A}_1{}^{*}\tilde{\cup}\tilde{A}_2{}^{*})\tilde{\cap}(\tilde{B}_1{}^{\overline{*}}\tilde{\cap}\tilde{B}_2{}^{\overline{*}})\tilde{\subseteq}(\tilde{A}_1\tilde{\cap}\tilde{A}_2,\tilde{B}_1\tilde{\cap}\tilde{B}_2)^{\triangleleft}$。由定义 8,有 $(\tilde{X}_1\tilde{\cup}\tilde{X}_2,(\tilde{A}_1\tilde{\cap}\tilde{A}_2,\tilde{B}_1\tilde{\cap}\tilde{B}_2))\in W_{OE}(K)$。

2) $(\tilde{X}_1\tilde{\cup}\tilde{X}_2,(\tilde{A}_1\tilde{\cap}\tilde{A}_2,\tilde{B}_1\tilde{\cap}\tilde{B}_2))$ 是 $(\tilde{X}_1,(\tilde{A}_1,\tilde{B}_1))$ 和 $(\tilde{X}_2,(\tilde{A}_2,\tilde{B}_2))$ 的一个上界。

因 $\tilde{X}_1,\tilde{X}_2\tilde{\subseteq}\tilde{X}_1\tilde{\cup}\tilde{X}_2$ 且 $(\tilde{A}_1,\tilde{B}_1),(\tilde{A}_2,\tilde{B}_2)\tilde{\supseteq}(\tilde{A}_1\tilde{\cap}\tilde{A}_2,\tilde{B}_1\tilde{\cap}\tilde{B}_2)$,则 $(\tilde{X}_1,(\tilde{A}_1,\tilde{B}_1)),(\tilde{X}_2,(\tilde{A}_2,\tilde{B}_2))\leqslant(\tilde{X}_1\tilde{\cup}\tilde{X}_2,(\tilde{A}_1\tilde{\cap}\tilde{A}_2,\tilde{B}_1\tilde{\cap}\tilde{B}_2))$。

3) $(\tilde{X}_1\tilde{\cup}\tilde{X}_2,(\tilde{A}_1\tilde{\cap}\tilde{A}_2,\tilde{B}_1\tilde{\cap}\tilde{B}_2))$ 是 $(\tilde{X}_1,(\tilde{A}_1,\tilde{B}_1))$ 和 $(\tilde{X}_2,(\tilde{A}_2,\tilde{B}_2))$ 的上确界。

假设 $(\tilde{X}_3,(\tilde{A}_3,\tilde{B}_3))$ 是 $(\tilde{X}_1,(\tilde{A}_1,\tilde{B}_1))$ 和 $(\tilde{X}_2,(\tilde{A}_2,\tilde{B}_2))$ 的一个上界,则 $\tilde{X}_1,\tilde{X}_2\tilde{\subseteq}\tilde{X}_3$ 且 $(\tilde{A}_1,\tilde{B}_1),(\tilde{A}_2,\tilde{B}_2)\tilde{\supseteq}(\tilde{A}_3,\tilde{B}_3)$,从而有 $\tilde{X}_1\tilde{\cup}\tilde{X}_2\tilde{\subseteq}\tilde{X}_3$ 且 $(\tilde{A}_1\tilde{\cap}\tilde{A}_2,\tilde{B}_1\tilde{\cap}\tilde{B}_2)\tilde{\supseteq}(\tilde{A}_3,\tilde{B}_3)$。因此,$(\tilde{X}_1\tilde{\cup}\tilde{X}_2,(\tilde{A}_1\tilde{\cap}\tilde{A}_2,\tilde{B}_1\tilde{\cap}\tilde{B}_2))\leqslant(\tilde{X}_3,(\tilde{A}_3,\tilde{B}_3))$。

因此,$(W_{OE}(K),\leqslant)$ 构成格。

3　前概念、OE-前概念与 OE-区间集前概念

3.1　前概念与 OE-前概念

定理 3　在形式背景 $K=(G,M,I)$ 中,对 $\forall X\in P(G)$,$\forall A,B\in P(M)$,下面结论成立:

1) $(X,A)\in P_c(K)\Rightarrow(X,(A,X^{\overline{*}}))\in T_{OE}(K)$;

2) $(X,B)\in P_c(K^c)\Rightarrow(X,(X^{*},B))\in T_{OE}(K^c)$。

证明　根据定义 5 和 6,定理 3 易证。

定理 4　在形式背景 $K=(G,M,I)$ 中,对 $\forall X\in P(G)$,$\forall A,B\in P(M)$,下面结论成立:

1) $(X,(A,B))\in T_{OE}(K)\Rightarrow(X,A)\in P_c(K)$;

2) $(X,(A,B))\in T_{OE}(K)\Rightarrow(X,B)\in P_c(K^c)$。

证明　根据定义 5 和 6,定理 4 易证。

在形式背景 $K=(G,M,I)$ 中,令一个新的集合 $V(K)=\{(X,(A,\varnothing))\in T_{OE}(K)\mid\forall X\in P(G),\forall A\in P(M)\}$,然后得到定理 5。

定理 5　对于形式背景 $K=(G,M,I)$,必有 $V(K)\cong P_c(K)$ 成立。

证明　定义映射 $\varphi:V(K)\rightarrow P_c(K)$ 为 $\varphi(X,(A,\varnothing))=(X,A)$。

首先,证明 φ 为单射:对 $\forall(X_1,(A_1,\varnothing)),(X_2,(A_2,\varnothing))\in V(K)$,满足 $(X_1,(A_1,\varnothing))\neq(X_2,(A_2,\varnothing))$,则必有 $\varphi(X_1,(A_1,\varnothing))=(X_1,A_1)\neq(X_2,A_2)=\varphi(X_2,(A_2,\varnothing))$。

其次,证明 φ 为满射:对于 $\forall(X,A)$,$(X,A)=\varphi(X,(A,\varnothing))$,从而可知 φ 为满射。

综上所述,φ 为双射。因此,在集合同构的意义下,$V(K)$ 与 $P_c(K)$ 可看成同一数学结构,即 $V(K)\cong P_c(K)$。

定理 6　在形式背景 $K=(G,M,I)$ 中,必有一个映射 $\delta:(P_c(K),\leqslant)\rightarrow(T_{OE}(K),\leqslant)$,使得 $\delta(P_c(K),\leqslant)$ 是 $(T_{OE}(K),\leqslant)$ 的一个子格。

证明　定义映射 $\delta:(\mathrm{P_c(K)},\leqslant)\rightarrow(\mathrm{T_{OE}(K)},\leqslant)$ 为 $\delta(X,A)=(X,(A,\varnothing))$。

显然,$\varnothing\subseteq V(K)\subseteq T_{OE}(K)$。

又有 $\forall(X_1,(A_1,\varnothing)),(X_2,(A_2,\varnothing))\in V(K)$,$(X_1,(A_1,\varnothing))\wedge(X_2,(A_2,\varnothing))=(X_1\cap X_2,(A_1\cup A_2,\varnothing\cup\varnothing))\in V(K)$;

$(X_1,(A_1,\varnothing))\vee(X_2,(A_2,\varnothing))=(X_1\cup X_2,(A_1\cap A_2,\varnothing\cap\varnothing))\in V(K)$。

从而,$V(K)$ 对于 $\wedge$ 和 $\vee$ 都封闭。

因此,定理 6 结论成立。

3.2　OE-前概念与 OE-区间集前概念

定理 7　在形式背景 $K=(G,M,I)$ 中,对 $\forall X,X_1,X_2\in P(G)$,$\forall A,A_1,A_2,B,B_1,B_2\in P(M)$,结论 1)~5)成立。

1) $(X,(A,B))\in T_{OE}(K)\Rightarrow([X,X],([A,A],[B,B]))\in W_{OE}(K)$;

2) 若 $X_1\subseteq X_2$ 且 $(A_1,B_1)\supseteq(A_2,B_2)$,则 $(X_1,$

$(A_1,B_1)),(X_2,(A_2,B_2)) \in T_{OE}(K) \Rightarrow ([X_1,X_2],([A_2,A_1],[B_2,B_1])) \in W_{OE}(K)$;

3) 若 $X_1 \subseteq X_2$ 且 $(A_1,B_1) \subseteq (A_2,B_2)$, 则 $(X_1,(A_1,B_1)),(X_2,(A_2,B_2)) \in T_{OE}(K) \Rightarrow ([X_1,X_2],([A_1,A_2],[B_1,B_2])) \in W_{OE}(K)$;

4) 若 $X_1 \subseteq X_2, A_1 \subseteq A_2$ 且 $B_1 \supseteq B_2$, 则 $(X_1,(A_1,B_1)),(X_2,(A_2,B_2)) \in T_{OE}(K) \Rightarrow ([X_1,X_2],([A_1,A_2],[B_2,B_1])) \in W_{OE}(K)$;

5) 若 $X_1 \subseteq X_2, A_1 \supseteq A_2$ 且 $B_1 \subseteq B_2$, 则 $(X_1,(A_1,B_1)),(X_2,(A_2,B_2)) \in T_{OE}(K) \Rightarrow ([X_1,X_2],([A_2,A_1],[B_1,B_2])) \in W_{OE}(K)$。

证明　根据定义 6 和 8,定理 7 易证。

例 3　在表 2 所示的形式背景 $K_1=(G_1, M_1, I_1)$ 中,

1) $(123,(d,e)) \in T_{OE}(K_1)$, 且有 $([123,123],([d,d],[e,e])) \in W_{OE}(K_1)$。

2) $(2,(ad,be)),(23,(d,e)) \in T_{OE}(K_1)$, 满足 $2 \subseteq 23$ 和 $(ad,be) \supseteq (d,e)$, 且有 $([2,23],([d,ad],[e,be])) \in W_{OE}(K_1)$。

3) $(12,(a,e)),(12,(ad,be)) \in T_{OE}(K_1)$, 满足 $12 \subseteq 12$ 和 $(a,e) \subseteq (ad,be)$, 且有 $([12,12],([a,ad],[e,be])) \in W_{OE}(K_1)$。

4) $(4,(a,b)),(24,(ac,\varnothing)) \in T_{OE}(K_1)$, 满足 $4 \subseteq 24, a \subseteq ac$ 和 $b \supseteq \varnothing$, 且有 $([4,24],([a,ac],[\varnothing,b])) \in W_{OE}(K_1)$。

5) $(3,(d,e)),(23,(\varnothing,e)) \in T_{OE}(K_1)$, 满足 $3 \subseteq 23, d \supseteq \varnothing$ 和 $e \subseteq e$, 且有 $([3,23],([\varnothing,d],[e,e])) \in W_{OE}(K_1)$。

定理 8　在形式背景 $K=(G,M,I)$ 中,对 $\forall X_1, X_2 \in P(G), \forall A_1,A_2,B_1,B_2 \in P(M)$, 必有 $([X_1,X_2],([A_1,A_2],[B_1,B_2])) \in W_{OE}(K) \Rightarrow (X_1,(A_2,B_2)),(X_2,(A_1,B_1)) \in T_{OE}(K)$ 成立。

证明　根据定义 6 和 8,定理 8 易证。

例 4　在表 2 所示的形式背景 $K_1=(G_1, M_1, I_1)$ 中, $([2,23],([d,acd],[e,e])) \in W_{OE}(K_1)$, 而且有 $(2,(acd,e))$, $(23,(d,e)) \in T_{OE}(K_1)$。

定理 9　在形式背景 $K=(G,M,I)$ 中,必有一个映射 $\xi:(T_{OE}(K), \leqslant) \to (W_{OE}(K), \leqslant)$, 使得 $\xi(T_{OE}(K), \leqslant)$ 是 $(W_{OE}(K), \leqslant)$ 的一个子格。

证明　定义映射 $\xi:(T_{OE}(K), \leqslant) \to (W_{OE}(K), \leqslant)$ 为 $\xi(X,(A,B))=([X,X],([A,A],[B,B]))$。设 $H(K)=\{([X,X],([A,A],[B,B])) \in W_{OE}(K) \mid \forall X \in P(G), \forall A,B \in P(M)\}$。显然, $\varnothing \subseteq H(K) \subseteq W_{OE}(K)$。又对于 $\forall([X_1,X_1],([A_1,A_1],[B_1,B_1])),([X_2,X_2],([A_2,A_2],[B_2,B_2])) \in H(K)$, $([X_1,X_1],([A_1,A_1],[B_1,B_1])) \wedge ([X_2,X_2],([A_2,A_2],[B_2,B_2]))=([X_1 \cap X_2, X_1 \cap X_2],([A_1 \cup A_2, A_1 \cup A_2],[B_1 \cup B_2, B_1 \cup B_2])) \in H(K)$; $([X_1,X_1],([A_1,A_1],[B_1,B_1])) \vee ([X_2,X_2],([A_2,A_2],[B_2,B_2]))=([X_1 \cup X_2, X_1 \cup X_2],([A_1 \cap A_2, A_1 \cap A_2],[B_1 \cap B_2, B_1 \cap B_2])) \in H(K)$。从而, $H(K)$ 对于 $\wedge$ 和 $\vee$ 都封闭。因此,定理 9 结论成立。

3.3　前概念、*OE*-前概念与 *OE*-区间集前概念

定理 10　在形式背景 $K=(G,M,I)$ 中, $P_c(K)$、$T_{OE}(K)$ 和 $W_{OE}(K)$ 在集合同构意义下有 $P_c(K) \subseteq T_{OE}(K) \subseteq W_{OE}(K)$ 成立。

证明　根据定理 5 和定理 10 易证。

例 5　在表 2 所示的形式背景 $K_1=(G_1, M_1, I_1)$ 中,列举下面 3 个三元区间集序对: $([4,4],([ac,ac],[\varnothing,\varnothing]))$, $([12,12],([ad,ad],[be,be]))$ 和 $([3,123],([d,bd],[e,ae]))$。下面,通过表 3 对这三者进行比较。

表 3　对比表

Table 3　A comparison table

三元区间集序对	$P_c(K)$	$T_{OE}(K_1)$	$W_{OE}(K_1)$
$([4,4],([ac,ac],[\varnothing,\varnothing]))$	是	是	是
$([12,12],([ad,ad],[be,be]))$	不是	是	是
$([3,123],([d,bd],[e,ae]))$	不是	不是	是

接着,在表 4 中,将前概念、*OE*-前概念和 *OE*-区间集前概念比较并分析。

表 4　对比分析表

Table 4　A comparison analysis table

名称	提取知识的表达形式	序对维数	提取知识的规则	提取知识的平台
前概念	(X,A)	2	$X \subseteq A^*(\Leftrightarrow A \subseteq X^*)$	格
OE-前概念	$(X,(A,B))$	3	$X \subseteq A^* \cap B^{\overline{*}}(\Leftrightarrow A \subseteq X^*$ 且 $B \subseteq X^{\overline{*}})$	格
OE-区间集前概念	$(\tilde{X},(\tilde{A},\tilde{B}))$	3	$\tilde{X} \tilde{\subseteq} (\tilde{A},\tilde{B})^{\lhd}(\Leftrightarrow(\tilde{A},\tilde{B}) \tilde{\subseteq} \tilde{X}^{\rhd})$	格

从表 4 可以看出,*OE*-区间集前概念能提取 *OE*-前概念和前概念都不能提取的知识,并且 *OE*-前概念和前概念能提取的知识也都能通过 *OE*-区间集前概念来提取,从而 *OE*-区间集前概念能够提取

的知识量比另外两个都多。

这三者在知识提取时,都是通过构造不同规则然后建立在相应的格上进行的。也就是说,OE-区间集前概念将前概念和 OE-前概念的知识表示形式进行拓展,在知识提取方面具有明显优势。

3.4 构造算法

本节从形式背景 $K=(G,M,I)$ 出发,先寻找全体前概念,再寻找全体 OE-前概念,最后可找到全体 OE-区间集前概念。

因此,可以给出相应的构造路径。

$$K=(G,M,I)\xrightarrow{\text{算法 1}}P_c(K)\xrightarrow{\text{算法 2}}T_{OE}(K)\xrightarrow{\text{算法 3}}W_{OE}(K)。$$

下面给出寻找全体前概念的算法 1。

算法 1 寻找全体前概念的算法

输入:形式背景 $K=(G,M,I)$。

输出:$P_c(K)$。

1) Let $P_c(K)=\varnothing$ and $i=1$
2) for each $X\subseteq G$ do
3) $\quad C=X^*$ and $P_i=(X,C)$
4) $\quad$ for each $A\subseteq C$ do
5) $\quad\quad P_i=P_i\cup(X,A)$
6) $\quad$ end for
7) $\quad P_c(K)=P_c(K)\cup P_i, i=i+1$
8) end for
9) Generate $P_c(K)$

定义 5 和文献[8,10]保证了算法 1 正确。

算法 1 的复杂度分析。

1) 将 $K=(G,M,I)$ 下的集合 $P_c(K)$ 设为空集,这个过程的时间复杂度为 $O(1)$。

2) 列出对象集 G 中的所有子集 X 并对每一个 X 都进行下面的操作:把经过算子 $^*:P(G)\to P(M)$ 运算得到的 X^* 设为 C, 将所得到的二元序对 (X,A) 放到集合 P_i 中。这个过程的时间复杂度为 $O(2^{|M|})$。其中:$|M|$ 为属性集 M 包含属性的个数。

3) 接着把 C 中所有子集 A 与这个 X 结合成二元序对 (X,A) 并放到集合 $P_c(K)$ 中。这个过程的时间复杂度为 $O(2^{|G|+|M|})$。其中:$|G|$ 为属性集 G 包含对象的个数。

4) 当每一个 X 都完成这样的操作后,所生成的 $P_c(K)$ 便包含了 K 下的外延为 X 且内涵为 A 的所有前概念。

通过算法 1 的复杂度分析 1)~4)可知,算法 1 的时间复杂度为

$$O(1+2^{|M|}+2^{|G|+|M|})=O(2^{|G|+|M|})。$$

下面给出寻找全体 OE-前概念的算法 2。

算法 2 寻找全体 OE-前概念算法

输入:形式背景 $K=(G,M,I)$ 和 $P_c(K)$。

输出:$T_{OE}(K)$。

1) Let $T_{OE}(K)=\varnothing$ and $j=1$
2) for each $(X,A)\subseteq P_c(K)$ do
3) $\quad D=X^{\overline{*}}$ and $T_j=(X,(A,D))$
4) $\quad$ for each $B\subseteq D$ do
5) $\quad\quad T_j=T_j\cup(X,(A,B))$
6) $\quad$ end for
7) $\quad T_{OE}(K)=T_{OE}(K)\cup T_j, j=j+1$
8) end for
9) Generate $T_{OE}(K)$

定义 6 和定理 3 保证了算法 2 正确性。

算法 2 的复杂度分析。

1) 将 $K=(G,M,I)$ 下的集合 $T_{OE}(K)$ 设为空集,这个过程的时间复杂度为 $O(1)$。

2) 列出 $P_c(K)$ 中的所有前概念 (X,A), 把由算子 $^{\overline{*}}:P(G)\to P(M)$ 运算得到的 $X^{\overline{*}}$ 设为 D, 将所得到的三元序对 $(X,(A,D))$ 放到集合 T_j 中。这个过程的时间复杂度为 $O(2^{|G|+|M|})$。

3) 把 D 中所有子集 B 与这个 (X,A) 结合成三元序对 $(X,(A,B))$ 并放到集合 $T_{OE}(K)$ 中。此过程时间复杂度为 $O(2^{|G|+2|M|})$。

4) 当每一个 (X,A) 都完成这样的操作后,所生成的 $T_{OE}(K)$ 便包含了 K 下的外延为 X 且内涵为 (A,B) 的所有 OE-前概念。

通过算法 2 的复杂度分析 1)~4)可知,算法 2 的时间复杂度为

$$O(1+2^{|G|+|M|}+2^{|G|+2|M|})=O(2^{|G|+2|M|})。$$

在给出寻找全体 OE-区间集前概念的算法之前,先要给出定理 11。

首先假设通过定理 7 中的 5 个结论得到的所有 OE-区间集前概念构成的集合为 Δ, 然后对下面的定理 11 展开讨论。

定理 11 在形式背景 $K=(G,M,I)$ 中,全体 OE-区间集前概念 $W_{OE}(K)$ 与集合 Δ 相同。

证明 首先,$\Delta\subseteq W_{OE}(K)$ 是显然成立的。

其次,由于全体 OE-区间集前概念 $W_{OE}(K)$ 中的元素都是形如 $(\tilde{X},(\tilde{A},\tilde{B}))$ 且满足 $\tilde{X}\tilde{\subseteq}(\tilde{A},\tilde{B})^{\triangleleft}$ 的三元区间集序对,然后可分为下面 8 种具体情况。

① $\tilde{X}=[X_1,X_2]$, $\tilde{A}=[A_1,A_2]$ 和 $\tilde{B}=[B_1,B_2]$;

② $\tilde{X} = [X_1, X_2]$,$\tilde{A} = [A_2, A_1]$ 和 $\tilde{B} = [B_2, B_1]$;

③ $\tilde{X} = [X_1, X_2]$,$\tilde{A} = [A_2, A_1]$ 和 $\tilde{B} = [B_1, B_2]$;

④ $\tilde{X} = [X_1, X_2]$,$\tilde{A} = [A_1, A_2]$ 和 $\tilde{B} = [B_2, B_1]$;

⑤ $\tilde{X} = [X_2, X_1]$,$\tilde{A} = [A_1, A_2]$ 和 $\tilde{B} = [B_1, B_2]$;

⑥ $\tilde{X} = [X_2, X_1]$,$\tilde{A} = [A_2, A_1]$ 和 $\tilde{B} = [B_2, B_1]$;

⑦ $\tilde{X} = [X_2, X_1]$,$\tilde{A} = [A_2, A_1]$ 和 $\tilde{B} = [B_1, B_2]$;

⑧ $\tilde{X} = [X_2, X_1]$,$\tilde{A} = [A_1, A_2]$ 和 $\tilde{B} = [B_2, B_1]$。

对于上面这 8 种具体情况来说,与定理 8 的证明方法类似,并根据定义 6 和 8 可知,②和⑤都可以推出 $(X_1,(A_1,B_1)),(X_2,(A_2,B_2)) \in T_{OE}(K)$,因此②与⑤等价;同理,①与⑥等价;④与⑦等价;③与⑧等价。

因此,这 8 种具体情况可用①、②、③和④来代表,从而有 $\Delta \supseteq W_{OE}(K)$。

综上所述,$\Delta = W_{OE}(K)$。下面,给出寻找全体 OE-区间集前概念的算法 3。

算法 3　寻找全体 OE-区间集前概念算法

输入:形式背景 $K = (G, M, I)$ 和 $T_{OE}(K)$。

输出:$W_{OE}(K)$。

1) Let $W_{OE}(K) = \varnothing$,

2) while $((X_1,(A_1,B_1)),(X_2,(A_2,B_2)) \in T_{OE}(K))$ do

3) if $X_1 \subseteq X_2$ and $(A_1,B_1) \supseteq (A_2,B_2)$ then

4) $W_{OE}(K) = W_{OE}(K) \cup ([X_1,X_2],([A_2,A_1],[B_2,B_1]))$

5) elif $X_1 \subseteq X_2$ and $(A_1,B_1) \subseteq (A_2,B_2)$ then

6) $W_{OE}(K) = W_{OE}(K) \cup ([X_1,X_2],([A_1,A_2],[B_1,B_2]))$

7) elif $X_1 \subseteq X_2, A_1 \subseteq A_2$ and $B_1 \supseteq B_2$ then

8) $W_{OE}(K) = W_{OE}(K) \cup ([X_1,X_2],([A_1,A_2],[B_2,B_1]))$

9) elif $X_1 \subseteq X_2, A_1 \supseteq A_2$ and $B_1 \subseteq B_2$ then

10) $W_{OE}(K) = W_{OE}(K) \cup ([X_1,X_2],([A_2,A_1],[B_1,B_2]))$

11) end if

12) end while

13) Generate $W_{OE}(K)$

定理 7 和定理 11 保证了算法 3 的正确性。

算法 3 的时间复杂度分析。

1) 将 $K = (G,M,I)$ 下的集合 $W_{OE}(K)$ 设为空集,这个过程的时间复杂度为 $O(1)$。

2) 对集合 $T_{OE}(K)$ 中的任意两个 OE-前概念 $((X_1,(A_1,B_1))$ 和 $(X_2,(A_2,B_2))$ 分别进行相应的判断和操作。当满足其中任一条件时,都可以构成相应的 OE-区间集前概念。这个过程的时间复杂度为 $O(2^{2^{|G|+4|M|}})$。

3) 当每两个 OE-前概念都完成判断和操作后,所生成的 $W_{OE}(K)$ 便包含了 K 下的所有 OE-区间集前概念。

通过算法 3 的复杂度分析 1)~3)可知,算法 3 的时间复杂度为

$$O(1 + 2^{2^{|G|+4|M|}}) = O(2^{2^{|G|+4|M|}})。$$

因此,构造 $W_{OE}(K)$ 的算法由算法 1~3 组成,总的时间复杂度为

$$O(2^{|G|+|M|}) + O(2^{|G|+2|M|}) + O(2^{2^{|G|+4|M|}}) = O(2^{2^{|G|+4|M|}})。$$

下面,通过实际例子对上述算法 3 进行具体实现与验证。

例 6　在表 2 所示的形式背景 $K_1 = (G_1, M_1, I_1)$ 中,下面将构造 OE-区间集前概念的算法进行实际应用并展示出部分结果。

按照算法 1 的步骤 1),可得到初始值 $P_c(K) = \varnothing$。

按照算法 1 的步骤 2)~3)和步骤 8)的循环语句,比如当 $X = 12 \in P(G_1)$ 时,可以得到 $C = X^* = ad$ 和 $P_i = (12, ad)$。

按照算法 1 的步骤 4)~6)的循环语句以及步骤 7),比如当 $A = d \subseteq C$,此时的 $P_i = P_i \cup (12, d)$。相应地,此时 i 值对应的 P_i 放到集合 $P_c(K)$ 中。

按照算法 1 的步骤 9),输出 $P_c(K)$ 的值。

由算法 2 步骤 1),有初始值 $T_{OE}(K) = \varnothing$。

按照算法 2 的步骤 2)~3)和步骤 8)的循环语句,当 $(12, d) \in P_c(K)$ 时,可以得到 $D = X^{\bar{*}} = be$ 和 $T_j = (12,(d, be))$。

按照算法 2 的步骤 4)~6)的循环语句以及步骤 7),比如当 $B = e \subseteq D$,此时的 $T_j = T_j \cup (12,(d,e))$。相应地,此时 j 值对应的 T_j 放到集合 $T_{OE}(K)$ 中。

按照算法 2 的步骤 9),输出 $T_{OE}(K)$。

由算法 3 步骤 1),初始值 $W_{OE}(K) = \varnothing$。

按照算法 3 的步骤 2)和 12),对于 $T_{OE}(K)$ 中的任意两个元素,比如有 $(12,(d,e))$ 和 $(2,(acd,e))$,进行相应的判断和操作。

按照算法 3 的步骤 3)~11)的判断语句,由于满足 $2 \subseteq 12$,$d \subseteq acd$ 和 $e \subseteq e$,从而可以得到 $([2,12],([a,acd],[e,e])) \in W_{OE}(K)$。

按照算法3的步骤13),输出 $W_{OE}(K)$。

4 结论

本文在区间集概念格的基础上,通过与前概念、三支决策相结合提出 OE-区间集前概念与其格结构,并用实际案例说明 OE-区间集前概念在提取知识过程中的应用。

未来工作包括在模糊形式背景或不完备形式背景下展开对 OE-区间集前概念的研究,并探索其可应用的其他领域。

参考文献:

[1] WILLE R. Restructuring lattice theory: an approach based on hierarchies of concepts[M]. Berlin: Springer Press, 2009: 314-339.

[2] 徐健锋, 辛朋, 薛国泽, 等. 面向概念漂移集成分类的三支决策优化方法[J]. 郑州大学学报(理学版), 2021, 53(1):22-28.
XU J F, XIN P, XUE G Z, et al. Three-way decision optimization method for concept drift ensemble classification[J]. Journal of Zhengzhou university (natural science edition), 2021, 53(1):22-28.

[3] CHUNDURI R K, CHERUKURI A K. Scalable algorithm for generation of attribute implication base using FP-growth and spark[J]. Soft computing, 2021, 25(14): 9219-9240.

[4] HU Q, YUAN Z, QIN K Y, et al. A novel outlier detection approach based on formal concept analysis[J]. Knowledge-based systems, 2023, 268: 110486.

[5] STAHL J, WILLE R. Preconcepts and set representations of contexts[M]//Classification as a Tool of Research. Amsterdam: North-Holland Press, 1986: 431-438.

[6] ŠOSTAK A, UĻJANE I, KRASTIŅŠ M. Gradation of fuzzy preconcept lattices[J]. Axioms, 2021, 10(1): 41.

[7] BURGMANN C, WILLE R. The basic theorem on preconcept lattices[M]//Formal Concept Analysis. Berlin: Springer Berlin Heidelberg, 2006: 80-88.

[8] YAO Y Y. Interval-set algebra for qualitative knowledge representation[C]//Proceedings of International Conference on Computing and Information. Piscataway: IEEE Press, 2002: 370-374.

[9] MA J M, HU L L, QIAN Y H. Object-oriented interval-set concept lattices[J]. International journal of approximate reasoning, 2019, 110: 64-81.

[10] 马建敏, 景嫄. 乐观多粒度区间集粗糙集[J]. 郑州大学学报(理学版), 2018, 50(3):87-93.
MA J M, JING Y. Optimistic multi-granulation interval-set rough sets[J]. Journal of Zhengzhou university (natural science edition), 2018, 50(3):87-93.

[11] YAO Y Y. An outline of a theory of three-way decisions[M]//Rough Sets and Current Trends in Computing. Berlin: Springer Press, 2012: 1-17.

[12] QI J J, WEI L, YAO Y Y. Three-way formal concept analysis[M]//Rough Sets and Knowledge Technology. Cham: Springer International Publishing, 2014: 732-741.

[13] MAO H, CHENG Y L, LIU X Q. Three-way preconcept and two forms of approximation operators[J]. Soft computing, 2023, 27(2): 855-865.

[14] 刘营营, 米据生, 梁美社, 等. 三支区间集概念格[J]. 山东大学学报(理学版), 2020, 55(3): 70-80.
LIU Y Y, MI J S, LIANG M S, et al. Three-way interval-set concept lattice[J]. Journal of Shandong university (natural science), 2020, 55(3): 70-80.