
《智能信息处理》课程作业

三支概念分析的研究现状与展望

赵茂昌

作业	分数[20]
得分	

2020 年 11 月 12 日

三支概念分析的研究现状与展望

赵茂昌

(大连海事大学 信息科学与技术学院 大连 116026)

摘要 形式概念分析(Formal Concept Analysis, FCA) [1]是由 R. Wille 于 1982 年提出一种从形式背景进行数据分析和规则提取的强有力工具, 在信息检索、数据挖掘、软件工程等领域上的应用十分广泛。概念格的每个节点是一个形式概念, 由两部分组成: 外延, 即概念所覆盖的实例; 内涵, 即概念的描述, 该概念覆盖实例的共同特征。另外, 概念格通过 Hasse 图生动和简洁地体现了这些概念之间的泛化和特化关系。因此, 概念格被认为是进行数据分析的有力工具。从数据集中(概念格中称为形式背景)中生成概念格的过程实质上是一种概念聚类过程; 然而, 概念格可以用于许多机器学习的任务。

关键词 形式概念分析; 规则提取; 概念格; 概念聚类

Research status and Prospect of three branch concept analysis

Maochang Zhao

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

Abstract Formal concept analysis (FCA) [1] is a powerful tool for data analysis and rule extraction from formal background proposed by R. Wille in 1982. It is widely used in information retrieval, data mining, software engineering and other fields. Each node of concept lattice is a formal concept, which is composed of two parts: extension, that is, the instance covered by the concept; connotation, which is the description of concept, which covers the common characteristics of instances. In addition, the concept lattice vividly and concisely reflects the generalization and specialization relationship between these concepts through Hasse graph. Therefore, concept lattice is considered to be a powerful tool for data analysis. The process of generating concept lattice from data set (called formal context in concept lattice) is essentially a concept clustering process; however, concept lattice can be used for many machine learning tasks.

Keywords formal concept analysis; rule extraction; concept lattice; concept clustering.

1 引言

形式概念分析是一种基于形式背景进行数据处理和知识表示的数学工具, 是 Ganter and Wille 在 1982 年提出的一种表示形式概念的模式, 被广泛应用于数据挖掘、信息检索、机器学习、人工智能等领域。然而, 在形式概念分析中通常只能得到二支决策, 而非三支决策, 即, 对于每一个形式概念, 只能确定一个对象是否拥有内涵中的所有元素或者一个属性是否被外延中的所有元素共享[2]。实际上, 对于一个形式背景的补背景也能得到一个对象是否不

拥有任何内涵中的任意元素或者一个属性不被外延中任何对象具有等信息, 但是这些信息并没有反应在形式概念中, Qi et al 将三支决策理论应用到形式概念分析中, 提出了三支概念分析理论。三支概念分析的两个关键组成部分是三支概念与三支概念格。由三支决策思想可知, 正域、负域、边界域为全集的一个划分, 因此三支决策只需要刻画正域和负域两部分。属性导出三支概念的外延由正域和负域两部分组成, 这两部分在形式背景中表示内涵中的属性被对象“共同拥有”和“共同不拥有”。对象导出三支概念的内涵也由正域和负域两部分

组成,这两部分在形式背景中表示对象“共同拥有”和“共同不拥有”内涵中的属性。

2 形式概念分析

形式概念分析是由 Wille 提出的一种从形式背景中进行数据分析和规则提取的有用的工具。一个形如 (U, A, I) 的数据表被称作一个形式背景,其中 U 是对象集, A 是属性集, I 是 U 和 A 之间的二元关系。在一个 (U, A, I) 中, $\forall x \in U, a \in A, (x, a) \in I$ 说明 x 拥有 a 或者 a 被 x 拥有。 $(x, a) \notin I$ 说明 x 没有 a 或者 a 不被 x 拥有。为了简便,通常分别使用“ $I(x, a)=1$ ”和“ $I(x, a)=0$ ”表示 $(x, a) \in I$ 和 $(x, a) \notin I$ [3]。

定义 1 设 (U, A, I) 是一个形式背景, $X \subseteq U, B \subseteq A$, 一对对偶算子“ $*$ ”被定义为:

$$X^* = \{a \in A \mid (x, a) = 1, \forall x \in X\},$$

$$B^* = \{x \in U \mid (x, a) = 1, \forall a \in B\}.$$

同时, X^* 和 B^* 的补集被定义为:

$$\sim X^* = \{a \in A \mid (x, a) = 0, \exists x \in X\},$$

$$\sim B^* = \{x \in U \mid (x, a) = 0, \exists a \in B\}.$$

性质 1 设 (U, A, I) 是一个形式背景, $X, X_1, X_2 \subseteq U, B, B_1, B_2 \subseteq A$, 对偶算子“ $*$ ”存在以下性质。

$$1) X_1 \subseteq X_2 \Rightarrow X_2^* \subseteq X_1^*, B_1 \subseteq B_2 \Rightarrow B_2^* \subseteq B_1^*;$$

$$2) X \subseteq X^{**}, B \subseteq B^{**};$$

$$3) X^* = X^{***}, B^* = B^{***};$$

$$4) X \subseteq B^* \Leftrightarrow B \subseteq X^*;$$

$$5) (X_1 \cup X_2)^* = X_1^* \cap X_2^*, (B_1 \cup B_2)^* = B_1^* \cap B_2^*;$$

$$6) (X_1 \cap X_2)^* \supseteq X_1^* \cup X_2^*, (B_1 \cap B_2)^* \supseteq B_1^* \cup B_2^*.$$

在形式背景 (U, A, I) 中, 对任意的 $X \subseteq U$ 和 $B \subseteq A$, 如果 $X^* = B$ 且 $X = B^*$, 那么 (X, B) 被称为一个概念, 其中 X 是概念的外延, B 是概念的内涵。

一般的, (U, A, I) 中所有概念构成了概念格 $L(U, A, I)$ 。 $L(U, A, I)$ 中任意的 $(X_1, B_1), (X_2, B_2)$ 存在排序 $(X_1, B_1) \leq (X_2, B_2) \Leftrightarrow X_1 \subseteq X_2 \Leftrightarrow B_1 \supseteq B_2$, 其中 (X_1, B_1) 是 (X_2, B_2) 的子概念, (X_2, B_2) 是 (X_1, B_1) 的超概念。如果 $L(U, A, I)$ 中的概念满足法则

$$(X_1, B_1) \wedge (X_2, B_2) = (X_1 \cap X_2, (B_1 \cup B_2)^*),$$

$$(X_1, B_1) \vee (X_2, B_2) = ((X_1 \cup X_2)^*, B_1 \cap B_2).$$

那么, $L(U, A, I)$ 被称作完备的概念格。

在文献[45]中, 为了构造三支概念, 对偶算子“ $*$ ”被称作正算子, 下面给出负算子“ $\bar{*}$ ”的定义。

定义 2 设 (U, A, I) 是一个形式背

景, $X \subseteq U, B \subseteq A$, 负算子“ $\bar{*}$ ”的定义如下

$$X^{\bar{*}} = \{a \in A \mid (x, a) = 0, \forall x \in X\},$$

$$B^{\bar{*}} = \{x \in U \mid (x, a) = 0, \forall a \in B\}.$$

显然, (U, A, I) 的负算子即为 (U, A, I) 的正算子。同时, “ $\bar{*}$ ”拥有与“ $*$ ”相同的性质。

定义 3 设 (U, A, I) 是一个形式背景, $X, Y \subseteq U, B, C \subseteq A$ 。结合正算子“ $*$ ”和负算子“ $\bar{*}$ ”给出三支算子及其逆算子的定义如下

$$X \uparrow = (X^*, X^{\bar{*}}), B \uparrow = (B^*, B^{\bar{*}});$$

$$(X, Y) \downarrow = \{a \in A \mid a \in X^* \text{ 且 } a \in Y^{\bar{*}}\},$$

$$(B, C) \downarrow = \{x \in U \mid x \in B^* \text{ 且 } x \in C^{\bar{*}}\}.$$

将 Zadeh[4] 的模糊集合理论与三支形式概念分析相结合即可形成模糊三支形式概念分析, 进而构造模糊三支概念格[5]。

定义 4 模糊形式背景设 $F = (G, M, \Gamma)$ 为一个模糊形式背景, 其中, G 为所有对象的集合, M 为所有属性的集合, Γ 为一个在域 $G \times M$ 上定义的模糊集。对于 $\forall (x, a) \in G \times M$ 有 $0 \leq \mu(x, a) \leq 1$ 。

定义 5 给定一个模糊形式背景 $F = (G, M, \Gamma)$ 以及一个阈值 α , $X \subseteq G, A \subseteq M$, $\varphi(X)$ 和 $\varphi(A)$ 为 X 与 A 上的模糊集。定义模糊正算子 $\bar{*}$ 为:

$$X^{\bar{*}} = \{(a, \mu(X, a)) \mid a \in M, \forall x \in X, \mu(x, a) \geq \alpha\}$$

$$\varphi(X)^{\bar{*}} = \{a \mid a \in M, \forall x \in X, \mu(x, a) \geq \alpha\}$$

$$A^{\bar{*}} = \{(x, \mu(X, a)) \mid x \in G, \forall a \in A, \mu(x, a) \geq \alpha\}$$

$$\varphi(A)^{\bar{*}} = \{x \mid x \in G, \forall a \in A, \mu(x, a) \geq \alpha\}$$

模糊负算子为:

$$X^{\bar{*}} = \{(a, \mu(X, a)) \mid a \in M, \forall x \in X, \mu(x, a) < \alpha\}$$

$$\varphi(X)^{\bar{*}} = \{a \mid a \in M, \forall x \in X, \mu(x, a) < \alpha\}$$

$$A^{\bar{*}} = \{(x, \mu(X, a)) \mid x \in G, \forall a \in A, \mu(x, a) < \alpha\}$$

$$\varphi(A)^{\bar{*}} = \{x \mid x \in G, \forall a \in A, \mu(x, a) < \alpha\}$$

结合正算子 $\bar{*}$ 与负算子 $\bar{*}$, 定义在 X 与 A 上的一

对模糊三支算子 \leftarrow :

$$X^{\leftarrow} = (X^*, X^{\bar{*}}), A^{\leftarrow} = (A^*, A^{\bar{*}})$$

对于以上三支算子, 可以逆定义算子 \rightarrow 为:

$$(\varphi(X), \varphi(Y))^{\rightarrow} = \{a \in M | a \in \varphi(X)^{\bar{*}}, a \in \varphi(Y)^{\bar{*}}\}$$

$$(\varphi(A), \varphi(B))^{\rightarrow} = \{x \in G | x \in \varphi(A)^{\bar{*}}, x \in \varphi(B)^{\bar{*}}\}$$

3 三支概念分析研究现状及展望

3.1 三支概念分析研究现状

由于3WCA是FCA与3WD结合的结果,形式概念分析的优劣势等问题以及研究热点也被带入到3WCA中。

因此,3WCA的研究内容主要包括:与FCA的关系,以及FCA中相关研究的平行研究,比如三支概念格的建格算法、属性约简、规则提取等;与三支概念格自身特点相关的研究内容包括:不完备背景上的知识获取、冲突分析问题的解决方法、对边界域中的对象的深入研究、序贯三支概念格等[6]。

3.1.1 3WCA基础理论研究

Qi等提出三支概念分析后,从元素、集合和序的角度分别分析了对象/属性导出三支概念格与经典概念格之间的关系,并给出了在经典概念基础上构造三支概念的充要条件以及相应算法。

Ren等研究了对象/属性导出三支概念格的4种属性约简问题及计算方法,并讨论了其间的关系。

陈雪等针对两种三支概念格,在AE(OE)协调的决策形式背景下,研究了保持AE(OE)非冗余规则信息不丢失的属性约简理论,并给出了计算方法。

林洪等提出了三支粒协调决策形式背景以及三支粒协调集的概念,结合区分矩阵和区分函数给出了三支粒约简方法,并讨论了决策形式背景下三支粒约简、粒约简、分类约简之间的关系。

Yu等主要对可以用三支概念格表示的完

备格的特征进行了研究,包括原子和不可约元等特殊元素的性质,完备格可被表示为三支概念格的条件等,并对三支粗糙概念格进行了类似的讨论。

3.1.2 模糊三支形式概念分析

在模糊形式背景 (U, A, I^*) 中,模糊概念有三种情形:模糊概念的外延模糊,但是其内涵是经典的;模糊概念的内涵模糊,但是其外延是经典的;模糊概念的外延和内涵均是模糊的[3]。

定义 6 设 (U, A, I^*) 是一个模糊形式背景, $X^* \subseteq F(U)$, $B^* \subseteq F(A)$, 一对正对偶算子“ \uparrow ”定义为

$$\begin{aligned} \bar{X}^{\uparrow} &= \{ \langle a, \mu_X(a) \rangle | \forall x \in X, \\ &\quad \tilde{I}(x, a) \geq \bar{X}(x) \}, \\ \tilde{B}^{\uparrow} &= \{ \langle x, \mu_B(x) \rangle | \forall b \in B, \\ &\quad \tilde{I}(x, b) \geq \tilde{B}(b) \}. \end{aligned}$$

其中:

$$\mu_X(a) = \bigwedge_{x \in X} \tilde{I}(x, a), \mu_B(x) = \bigwedge_{b \in B} \tilde{I}(x, b).$$

性质 2 设 (U, A, I^*) 是一个模糊形式背景, $X^*, X_1^*, X_2^* \subseteq F(U)$, $B^*, B_1^*, B_2^* \subseteq F(A)$, 正对偶算子“ \uparrow ”存在以下性质:

- 1) $\bar{X}_1 \subseteq \bar{X}_2 \Rightarrow \bar{X}_2^{\wedge} \subseteq \bar{X}_1^{\uparrow}$,
 $\bar{B}_1 \subseteq \bar{B}_2 \Rightarrow \bar{B}_2^{\wedge} \subseteq \bar{B}_1^{\uparrow}$;
- 2) $\bar{X} \subseteq \bar{X}^{\uparrow\uparrow}$, $\bar{B} \subseteq \bar{B}^{\uparrow\uparrow}$;
- 3) $\bar{X}^{\uparrow} = \bar{X}^{\uparrow\uparrow\uparrow}$, $\bar{B}^{\uparrow} = \bar{B}^{\uparrow\uparrow\uparrow}$;
- 4) $\bar{X} \subseteq \bar{B}^{\wedge} \Leftrightarrow \bar{B} \subseteq \bar{X}^{\uparrow}$;
- 5) $(\bar{X}_1 \cup \bar{X}_2)^{\wedge} = \bar{X}_1^{\uparrow} \cap \bar{X}_2^{\uparrow}$,
 $(\bar{B}_1 \cup \bar{B}_2)^{\wedge} = \bar{B}_1^{\uparrow} \cap \bar{B}_2^{\uparrow}$;
- 6) $(\bar{X}_1 \cap \bar{X}_2)^{\wedge} \supseteq \bar{X}_1^{\uparrow} \cup \bar{X}_2^{\uparrow}$,
 $(\bar{B}_1 \cap \bar{B}_2)^{\wedge} \supseteq \bar{B}_1^{\uparrow} \cup \bar{B}_2^{\uparrow}$.

3.2 三支概念分析研究展望

由于3WCA具有语义详实、可解释性强等特点,在前述研究成果的基础上还有很多可以进一步拓展研究的内容[6]。

不完备背景的深入研究:相对于完备性的基础信息,不完备的情况在我们的生活和科研领域更为多见。利用部分已知概念这一特殊的

三支概念,以及区间集所起的三支作用,可以先从理论上寻求表述和分析不完备背景的方法,探讨不同类型的部分已知概念族的结构,进而应用于实际问题当中。

决策形式背景的规则获取:FCA中基于决策形式背景进行的规则提取,往往需要设定决策形式背景的协调性。类似地,对于3WCA而言,协调性的设定也是必须的,而且会更为复杂一些,但是其含有的信息也更丰富。由于牵涉到了决策形式背景的补背景及其上的三支规则,也就是负规则的引入,这使得基于决策形式背景的三支角度的规则获取变得复杂,而更有意义。但是这部分工作还没有很好很深刻的成果。

与三支决策理论的深入结合三分而治再分析效用,是三支决策理论发展到目前为止较为完全的思想框架,而三支概念分析仅利用3WCA中的导出算子考虑了三分,对于治略与效用尚未考虑。即使是三分,目前的研究也多是对完备背景的对象集和属性集进行精确的三分,而对于模糊背景或者不完备背景来说,三分方法的研究是一个非常有意义的问题,而且才刚刚起步。另外,三分以后,处于边界域中的个体其信息是不够明确的,如何对它们进行更进一步的持续深入的分析,也是不容忽视的问题。

4 总 结

由于三分支决策理论和形式概念分析的优势,以及其描述语义的精确性和精细性,从认知角度获得新的研究增长点成为可能,这一点已经得到相关领域学者特别是国内学者的认可,并在短短五年内迅速发展。本文的研究成果和未来可行的研究前景是我们下一步工作的指导方针。然而,三个分支的概念分析仍然存在着不可回避的问题。例如,三个概念格的结构复杂性决定了先构建格再获取信息的方法很难推广和自适应,也很难在当前大数据趋势下应用,这与FCA一样。因此,对这三个分支的概念分析的研究还需要做更深入的工作。

参 考 文 献

- [1] Ganter B, Stumme G, Wille R. Formal Concept Analysis: Theory and applications - J. UCS Special Issue. Journal of Universal Computer Science, 2004, 10(8): 926-926.
- [2] 汪文威, 祁建军. 三支概念的构建算法. 西安电子科技大学学报, 2017, 44 (1) :71-76.
- [3] 徐伟华, 杨蕾, 张晓燕. 模糊三支形式概念分析与概念认知学习[J]. 西北大学学报(自然科学版), 2020, 50(04):516-528.
- [4] Zadeh L.A. Fuzzy logic and approximate reasoning. Synthese, 1975, 30 (3-4) :407-428.
- [5] 龙柄翰, 徐伟华. 模糊三支概念分析与模糊三支概念格[J]. 南京大学学报(自然科学), 2019, 55(04):537-545.
- [6] 魏玲, 高乐, 祁建军. 三支概念分析研究现状与展望[J]. 西北大学学报(自然科学版), 2019, 49(04):527-537.