

DOI: 10.11992/tis.201607018

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20170626.1740.012.html>

集族等价与基于粒的下近似算子研究

胡霞¹, 费鹏², 杜卫锋³

(1. 苏州工业职业技术学院 软件与服务外包学院, 江苏 苏州 215104; 2. 苏州市创采软件有限公司, 江苏 苏州 215128; 3. 嘉兴学院 数理与信息工程学院, 浙江 嘉兴 314001)

摘 要: 基于覆盖的粗集是推广经典粗集理论的方法之一, 有基于元素、基于粒和基于子系统的 3 类定义上下近似的途径, 以往大多数的文献往往从基于元素的角度出发进行定义。为了研究基于粒的近似算子特别是下近似算子的性质, 借鉴格论中既约元、可约元等概念, 提出了集族约简的概念。从集族约简出发, 探讨了集族等价的概念与性质, 并设计了集族约简的算法, 得到了两个集族等价是两个集族生成相同的下近似运算的充要条件这一结果, 为进一步开展一般二元关系下基于粒的近似算子的公理化方法的研究做了初步的理论方面的准备工作。

关键词: 近似算子; 约简; 粗集; 既约元; 可约元; 覆盖; 粒; 集族约简

中图分类号: TP18 **文献标志码:** A **文章编号:** 1673-4785(2018)02-0327-04

中文引用格式: 胡霞, 费鹏, 杜卫锋. 集族等价与基于粒的下近似算子研究[J]. 智能系统学报, 2018, 13(2): 327-330.

英文引用格式: HU Xia, FEI Peng, DU Weifeng. On collections equivalence and the granule based lower approximation operators[J]. CAAI transactions on intelligent systems, 2018, 13(2): 327-330.

On collections equivalence and the granule based lower approximation operators

HU Xia¹, FEI Peng², DU Weifeng³

(1. School of Software and Service Outsourcing, Suzhou Institute of Industrial Technology, Suzhou 215104, China; 2. Suzhou Chuangcai Software Co., Ltd., Suzhou 215128, China; 3. School of Mathematics, Physics and Information Engineering, Jiaxing University, Jiaxing 314001, China)

Abstract: Covering based rough set is one of the methods to extend the classical rough set theory. There are three kinds of approaches, the element based definition, the granule based definition, and the subsystem based definition, to define upper and lower approximation. Most of the literature in the past tends to define based on element. In order to study the properties of the granule based approximation operators, especially the lower approximation operator, referring the concepts of irreducible element and reducible element from lattice theory, the concept of collections reduct is put forward. Starting from the concept of collections reduct, the concept and properties of collections equivalence are discussed, and collections reduction algorithm is designed. The result that collections equivalence is the necessary and sufficient condition for generating the same lower approximation by collections is given here. The preliminary theoretical preparation is done here to further develop the axiomatization of the granule based approximation operators under general binary relation.

Keywords: approximation operators; reduct; rough sets; irreducible element; reducible element; covering; granule; collections reduct

粗集理论是一种处理不协调、不完备和不精确信息的数学工具^[1], 自 1982 年波兰数学家 Pawlak 首

次提出以来, 经过 30 余年的研究与发展, 在理论和应用上均取得了长足的进步。

收稿日期: 2016-07-19. 网络出版日期: 2017-06-26.

基金项目: 国家自然科学基金项目 (61202109).

通信作者: 杜卫锋. E-mail: 23031520@qq.com.

粗集理论的基础之一是从近似空间诱导出一对近似算子——上近似算子和下近似算子。经典的

Pawlak 粗集模型基于等价关系,这种严格的要求在一定程度上限制了粗集理论的应用。为了推广粗集模型,可以把等价关系放宽为一般二元关系,在此情况下,由一般二元关系决定的论域上的集簇不再“划分”。因此,另一种推广粗集模型的思路就是将原来由等价关系决定的“划分”放宽为“覆盖”,从而建立基于覆盖的粗集理论。但不同于经典粗集理论,基于覆盖的粗集理论不存在唯一的定义上下近似算子的方法。

已有很多文献对基于覆盖的粗集理论开展了卓有成效的研究^[2-8]。姚一豫等^[3]系统地研究了基于覆盖的 20 对上下近似算子; Mauricio Restrepo 等^[4]通过某些算子的相等关系将上下近似算子缩减为 16 对,并提出了算子精细度的概念,给了 16 对近似算子的精细关系,并用哈斯图进行了描述。M. Restrepo 等^[5]则研究了基于覆盖的具有对偶性的 16 对上下近似算子的拓扑性质。在数据挖掘领域,去除冗余属性,获取属性约简,从而简化知识的表示,提升数据处理效率是一个重要的研究课题。李立峰等^[6]研究发现覆盖粗糙集与形式背景之间存在一一对应关系,并且证明了覆盖粗糙集的交约简可化为概念格的属性约简; C. Wang 等^[7]开发了一种基于覆盖粗糙集的属性约简方法,这种启发式方法可以比较高效地获得近似最优约简的属性集; Yang Bin 等^[8]则将包含度的概念引入覆盖粗糙集,探索了一种新的覆盖近似空间的若干性质。在覆盖粗集理论中,我们有基于元素、基于粒和基于子系统的 3 类定义上下近似的途径,以往大多数的文献往往从基于元素的角度出发进行定义,本文则以后继邻域作为基本研究对象“粒”,并以此为出发点,借鉴格论中既约元、可约元等概念^[9-10],探讨了(覆盖)集簇中的既约元、可约元、集簇的约简及其算法,另外,研究了集簇与其生成的下近似算子的关系,为下一步开展基于粒的公理化方法的研究做一些初步的理论方面的准备工作。

1 集簇约简

下面来分析一个例子。

例 1 设论域 $U = \{a, b, c, d\}$, 其上有 3 个不同的二元关系分别为

$$R_1 = \{(a, a), (a, b), (c, b), (c, c)\}$$

$$R_2 = \{(b, a), (b, b), (c, b), (c, c)\}$$

$$R_3 = \{(a, a), (a, b), (c, b), (c, c), (d, a), (d, b), (d, c)\}$$

基于粒的广义上下近似算子定义为后继邻域的广义并,而并不考虑该后继邻域是哪一个元素的后继邻域。因此我们只需研究二元关系 R 诱导的集

族 C , 而无需再去研究二元关系本身。此时,基于粒的广义上下近似算子可等效为

$$\underline{R}(X) = \bigcup \{C \in C | C \subseteq X\} \quad (1)$$

$$\bar{R}(X) = \bigcup \{C \in C | C \cap X \neq \emptyset\} \quad (2)$$

先对二元关系 R_1 进行分析,令 $r(x) = \{y \in U | (x, y) \in R\}$ 称为元素 x 的后继邻域,由此定义得, $r_1(a) = \{a, b\}$, $r_1(c) = \{b, c\}$, 其他元素的后继邻域为空集。非空后继邻域可以构成集簇为: $C_1 = \{r_1(a), r_1(c)\} = \{\{a, b\}, \{b, c\}\}$, 这就是二元关系 R_1 诱导的集簇。

同理,二元关系 R_2 诱导的集簇 $C_2 = \{\{a, b\}, \{b, c\}\}$ 。

由此发现, R_1 和 R_2 诱导了相同的集簇,因此它们定义了相同的下近似运算。

再看二元关系 R_3 , 诱导的集簇 $C_3 = \{\{a, b\}, \{b, c\}, \{a, b, c\}\}$ 。与 R_1 、 R_2 诱导的集簇并不相同,却也定义了相同的下近似运算。下面的分析将回答该问题。

定义 1 设 C 是论域 U 上的一个集簇, $C \in C$, 如果 C 不能表示成 $C - \{C\}$ 中若干个元之并, 则称 C 是 C 的既约元。否则, 称 C 是可约元。

由此可见, 集簇中的元素可以分成既约元和可约元两类。

定理 1 设 C 是论域 U 上的一个集簇, $C \in C$ 是一个可约元, $C' \in C - \{C\}$, 则 C' 是 C 的既约元当且仅当它是 $C - \{C\}$ 的既约元。

证明 1) 如果 C' 是 C 的既约元, 则 C' 不能表示成 $C - \{C'\}$ 中若干个元之并, 那么当然也就不能表示成 $C - \{C, C'\}$ 中若干个元之并, 因此 C' 是 $C - \{C\}$ 的既约元。

2) 反之, 如果 C' 是 $C - \{C\}$ 的既约元, 则 C' 不能表示成 $C - \{C, C'\}$ 中若干个元之并, 下面要证明 C' 同样也不能表示成 $C - \{C'\}$ 中若干个元之并, 而这只需证明 C' 不能表示成 $C - \{C, C'\}$ 中若干个元之并与 C 的并即可, 即 C' 不可能表示成 $C_1 \cup C_2 \cup \dots \cup C_n \cup C$, 其中 $C_1, C_2, \dots, C_n \in C - \{C, C'\}$ 。下面用反证法加以证明:

假设 C' 能表示成 $C - \{C, C'\}$ 中若干元与 C 之并, 即 $C' = C_1 \cup C_2 \cup \dots \cup C_n \cup C$, 其中 $C_1, C_2, \dots, C_n \in C - \{C, C'\}$ 。而 C 是一个可约元, 故 C 又可表示成 $C - \{C, C'\}$ 中若干个元之并, 即 $C = D_1 \cup D_2 \cup \dots \cup D_m$, 其中 $D_1, D_2, \dots, D_m \in C - \{C, C'\}$ 。由此可得 $C' = C_1 \cup C_2 \cup \dots \cup C_n \cup D_1 \cup D_2 \cup \dots \cup D_m$, 其中 $C_1, C_2, \dots, C_n, D_1, D_2, \dots, D_m \in C - \{C, C'\}$ 。 C' 表示成了 $C - \{C, C'\}$ 中若干元之并, 这与 C' 是 $C - \{C\}$ 的既约元矛盾, 故假设不成立, 因此 C' 是 C 的既约元。

由命题 (1)、(2) 得证。

定理 2 设 C 是论域 U 上的一个集簇, $C \in C$ 是一个可约元, $C' \in C - \{C\}$, 则 C' 是 C 的可约元当且仅当它是 $C - \{C\}$ 的可约元。

证明

1) 如果 C' 是 $C - \{C\}$ 的可约元, 则 C' 可以用 $C - \{C, C'\}$ 中若干个元之并表示出来, 那么当然也可表示成 $C - \{C'\}$ 中若干个元之并, 因此 C' 是 C 的可约元。

2) 反之, 如果 C' 是 C 的可约元, 则 C' 可用 $C - \{C'\}$ 中的若干元 C_1, C_2, \dots, C_n 表示出来。如果 $\forall C_i$, 都有 $C_i \neq C$, 则 C' 已经用 $C - \{C\}$ 中的若干个不等于 C' 的元之并表示出来了, 因此 C' 是 $C - \{C\}$ 的可约元。要是其中有一个元等于 C , 不妨设为 C_1 , 由于 C 是 C 中的可约元, 于是 C 可以表示成 $C - \{C\}$ 的若干元 D_1, D_2, \dots, D_m 的并。于是, $C' = C_1 \cup C_2 \cup \dots \cup C_n = D_1 \cup D_2 \cup \dots \cup D_m \cup C_2 \cup \dots \cup C_n$, 其中 $D_1, D_2, \dots, D_m, C_2, \dots, C_n$ 既不等于 C , 也不等于 C' , 这样证明了 C' 是 $C - \{C\}$ 中的可约元。

由命题 (1)、(2) 得证。

由定理 1 和定理 2 可以得出, 在一个集族中删除其中的一个可约元, 并不会改变其余元素是既约元还是可约元的性态。由此, 我们可以逐个删除集族中的所有可约元, 只剩下既约元。

定义 2 设 C 是论域 U 上的一个集族, 如果 C 中的每一个元都是既约元, 则称 C 是约简的。否则, 称 C 是可约的。

定义 3 对于 U 的一个集族 C , 通过删除集族中的所有可约元, 就得到了一个约简的集族, 我们将其称为集族 C 的约简, 记为 $\text{reduct}(C)$ 。

定理 1 和定理 2 实际上还保证了集族的约简是唯一的。

定义 4 设 C_1, C_2 是论域 U 上的两个集族, 如果 $\text{reduct}(C_1) = \text{reduct}(C_2)$, 则称集族 C_1 与 C_2 等价, 记为 $C_1 \sim C_2$ 。

集族等价形成集族空间 $\mathfrak{C} = \{C | C \text{ 是论域 } U \text{ 上的集族}\}$ 上的等价关系, 其中的等价类可记为 $[C]_R = \{X \in \mathfrak{C} | X \sim C\}$, 表示与集族 C 等价的所有集族的集合。

2 集族约简的算法

根据上节的结论, 我们可以给出求一个集族约简的算法, 该算法分为如下两大步骤:

- 1) 求集族的极小元 (极小元必定是既约元);
- 2) 由极小元求集族的非极小既约元。

将步骤 1)、2) 的结果并起来, 就是该集族的约简。

2.1 求集族的极小元

算法 1 求集族的极小元。

输入 论域 U 的一个集族 $C = \{C_1, C_2, \dots, C_n\}$;

输出 C 的极小元集合 $\text{minimal}(C)$ 。

1) 初始化, $\text{minimal}(C) = \emptyset$, 将集族 C 中的元素按基数从小到大排序;

2) 基数最小的元素一定是极小元, 设其基数为 i , 将其从集族 C 中移除并加入极小元集合 $\text{minimal}(C)$;

3) $i = i + 1$;

4) 如果集族 C 中已没有基数大于等于 i 的元素, 则转 5), 如果集族 C 中元素的基数均大于 i , 则转 3); 否则逐个检测基数为 i 的元素, 如果任何一个极小元均不是它的真子集, 则它是极小元, 将其从集族 C 中移除并加入极小元集合 $\text{minimal}(C)$, 转 3);

5) 算法结束, $\text{minimal}(C)$ 即为所求。

2.2 由极小元再求集族的非极小既约元

算法 2 求集族的非极小既约元。

输入 算法 1 结束时所得的集族 C 及 $\text{minimal}(C)$;

输出 非极小既约元的集合 $\text{irreducible}(C)$ 。

1) 初始化, $\text{irreducible}(C) = \emptyset$;

2) 逐个扫描集族 C 的元素 C , 找出 $\text{minimal}(C)$ 中每一个满足条件 $mC \subset C$ 的极小元 mC ;

3) 如果 $C \neq \cup mC$, 则 C 是一个既约元, 加入集合 $\text{irreducible}(C)$;

4) 算法结束, $\text{irreducible}(C)$ 即为所求。

最后, 将两步的结果并起来 $\text{reduct}(C) = \text{minimal}(C) \cup \text{irreducible}(C)$ 就是所求的集族 C 的约简。

例 2 设论域 $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}, x_{19}, x_{20}\}$, 计算集族 $C = \{C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9, C_{10}\}$ 的约简, 其中 $C_1 = \{x_1, x_2\}$, $C_2 = \{x_2, x_3\}$, $C_3 = \{x_1, x_2, x_3\}$, $C_4 = \{x_1, x_2, x_3, x_4\}$, $C_5 = \{x_4, x_5, x_6, x_7, x_8\}$, $C_6 = \{x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}\}$, $C_7 = \{x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}\}$, $C_8 = \{x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}\}$, $C_9 = \{x_{15}, x_{16}, x_{17}\}$, $C_{10} = \{x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}, x_{19}\}$ 。

2.3 根据求集族约简的算法, 先求集族的极小元

由算法 1, 将集族 C 中的元素按基数从小到大排序为

$$\begin{aligned} C_1 &= \{x_1, x_2\}, C_2 = \{x_2, x_3\} \\ C_3 &= \{x_1, x_2, x_3\}, C_9 = \{x_{15}, x_{16}, x_{17}\} \\ C_4 &= \{x_1, x_2, x_3, x_4\} \\ C_5 &= \{x_4, x_5, x_6, x_7, x_8\} \\ C_6 &= \{x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}\} \\ C_{10} &= \{x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}, x_{19}\} \\ C_8 &= \{x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}\} \end{aligned}$$

基数最小的元素 $C_1 = \{x_1, x_2\}$, $C_2 = \{x_2, x_3\}$ 是极小元。继续算法 1, 求得其余的极小元分别是 $C_9 = \{x_{15}, x_{16}, x_{17}, x_{18}\}$, $C_5 = \{x_4, x_5, x_6, x_7, x_8\}$, $C_6 = \{x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}\}$ 。

2.4 根据算法 2 由极小元再求集族的非极小既约元

$C_4 = \{x_1, x_2, x_3, x_4\}$, $C_{10} = \{x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}, x_{19}\}$, 最后结果为 $\text{reduct}(C) = \{C_1, C_2, C_4, C_5, C_6, C_9, C_{10}\}$ 。

3 集族等价与下近似运算

引理 1 设 C 是论域 U 上的一个集族, C 是 C 的可约元, $X \subseteq U$,则 X 在集族 $C - \{C\}$ 和集族 C 下生成的下近似相同。

证明 设 X 在集族 $C - \{C\}$ 和集族 C 下的下近似分别是 X_1, X_2 。显然有 $X_1 \subseteq X_2 \subseteq X$ 。另一方面,仍由下近似的定义,存在 C 中的元 C_1, C_2, \dots, C_n ,使得 $X_2 = C_1 \cup C_2 \cup \dots \cup C_n$ 。显然, C_1, C_2, \dots, C_n 都是 X 的子集。如果 C_1, C_2, \dots, C_n 均不等于 C 的话,则它们都属于 $C - \{C\}$,于是 C_1, C_2, \dots, C_n 都是 X_1 的子集。要是 C_1, C_2, \dots, C_n 中的某个等于 C ,不妨设为 C_1 ,由于 C 是 C 的可约元, C 可以表示成 $C - \{C\}$ 的若干个元 D_1, D_2, \dots, D_m 的并。于是, $X_2 = C_1 \cup C_2 \cup \dots \cup C_n = D_1 \cup D_2 \cup \dots \cup D_m \cup C_2 \cup \dots \cup C_n$,而 $D_1, D_2, \dots, D_m, C_2, \dots, C_n$ 都是 X 的子集,且都是 $C - \{C\}$ 中的元,因此它们都是 X_1 的子集。这样我们证明了 $X_2 \subseteq X_1$ 。

推论 1 设 C 是论域 U 上的一个集族,则 C 和 $\text{reduct}(C)$ 生成相同的下近似运算。

定理 3 设 C_1, C_2 是论域 U 的两个集族,如果 $\text{reduct}(C_1) = \text{reduct}(C_2)$,即 C_1 和 C_2 等价,则 C_1, C_2 生成相同的下近似运算。

4 结束语

本文从集族约简出发,探讨了关于集族的若干性质,得出了两个集族等价是两个集族生成相同的下近似运算的充要条件这一结论,为下一步开展基于粒的公理化方法的研究做了一些初步的理论方面的准备工作。本文借鉴格论中的概念来研究粗集,为研究粗集理论提供了一种新的思路。下一步的工作将把格论与粗集理论作更深入的结合,把格论中的一些方法和结论引入粗集理论,试图发现更多有趣的结果。另外,在此基础上将开展基于粒的粗集公理化方法的研究。

参考文献:

- [1] 张文修,梁怡,吴伟志. 信息系统与知识发现[M]. 北京: 科学出版社, 2003.
- [2] 祝峰,王飞跃. 关于覆盖广义粗集的一些基本结果[J]. 模式识别与人工智能, 2002, 15(1): 6-13.
ZHU Feng, WANG Feiyue. Some results on covering generalized rough sets[J]. Pattern recognition and artificial intelligence, 2002, 15(1): 6-13.
- [3] YAO Yiyu, YAO Bingxue. Covering based rough set approximations[J]. Information sciences, 2012, 200: 91-107.
- [4] RESTREPO M, CORNELIS C, GÓMEZ J. Partial order relation for approximation operators in covering based rough sets[J]. Information sciences, 2014, 284: 44-59.
- [5] RESTREPO M, GÓMEZ J. Topological properties for approximation operators in covering based rough sets[C]//Proceeding of the 15th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing. Tianjin, China, 2015: 112-123.
- [6] 李立峰,俞伟. 概念格约简与覆盖约简之间的关系[J]. 陕西理工学院学报: 自然科学版, 2014, 30(3): 37-40.
LI Lifeng, YU Wei. Relationships of reduction between concept lattice and covering[J]. Journal of Shaanxi university of technology: natural science edition, 2014, 30(3): 37-40.
- [7] WANG Changzhong, HE Qiang, CHEN Degang, et al. A novel method for attribute reduction of covering decision systems[J]. Information sciences, 2014, 254: 181-196.
- [8] YANG Bin, ZHU W. A new type of covering-based rough sets[C]//Proceedings of the 9th International Conference on Rough Sets and Knowledge Technology. Shanghai, China, 2014: 489-499.
- [9] 彭育威. 完全分配格的并一既约元的性质及分子格的代数结构[J]. 工程数学学报, 1985, 2(2): 113-117.
PENG Yuwei. Characterization of a join irreducible element of completely distributive lattice and algebraic structure of a molecular lattice[J]. Chinese journal of engineering mathematics, 1985, 2(2): 113-117.
- [10] 屈小兵,王学平. 完备格上并既约元的性质[J]. 模糊系统与数学, 2004, 18(S1): 176-179.
QU Xiaobing, WANG Xueping. Some properties of join-irreducible elements in complete lattice[J]. Fuzzy systems and mathematics, 2004, 18(S1): 176-179.

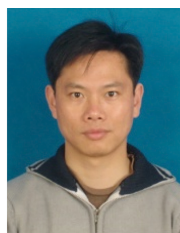
作者简介:



胡霞,女,1979年生,讲师,主要研究方向为粗糙集理论、信息处理。



费鹏,男,1978年生,高级工程师,主要研究方向为粗糙集、大数据处理。



杜卫锋,男,1977年生,副教授,博士,主要研究方向为粗糙集理论。