

粗集

上下近似运算

逻辑性质

②

计算机科学2000Vol. 27No. 11

布尔代数

粗集中上下近似运算的逻辑性质*)

The Logical Properties of Lower and Upper Approximation Operations in Rough Sets

79-81, 78

祝峰 何华灿

(西北工业大学计算机系 西安710072)

TP18

TP301.6

Abstract In this paper, we discuss the logical properties of rough sets through topological boolean algebras and closure topological boolean algebras. We get representation theorems of finite topological boolean algebras and closure topological boolean algebras under the upper-lower relation condition, which establish the relationship between topological boolean algebras or closure topological boolean algebras and rough sets in the general sets are similar to the Stone's representation theorem of boolean algebras.

Keywords Topological boolean algebra, Closure topological boolean algebra, Rough set relation condition

1 引论

近年来,粗集理论的实际应用与理论探讨已成为计算机科学中的一个热点问题。1995年 Pawlak 曾在文[6]中指出,粗集的逻辑性质研究将是今后粗集理论的一个重要问题。本文正是通过深入研究拓扑布尔代数与粗集的关系,给出了关于有限拓扑布尔代数的表示定理,从逻辑上全面刻画了粗集中上下近似运算这一核心概念。

2 基本概念

我们首先给出本文所要用到的一些基本概念。

定义1 满足下列条件的集合 A 称为关于运算 $U, \cap, -$ 和 I 的拓扑布尔代数:

- (1) A 是关于 $U, \cap, -$ 的布尔代数;
- (2) A 对于一元运算 I 封闭;
- (3) 对 $x \in A$, 有 $Ix \leq x$;
- (4) 对 $x \in A$, 有 $IIx = Ix$;
- (5) 对 $x, y \in A$, 有 $I(x \cap y) = Ix \cap Iy$;
- (6) $I1 = 1$

由定义易证 C, I 是单调增的。

粗集的定义 设 $U = \{u_1, u_2, \dots, u_n\}$ 是一个有限集, R 是 U 上的一个等价关系。我们称序偶 (U, R) 为近似空间 (approximation space)。 R 是等价关系, 它

必然对 U 产生一个剖分, 由此得到的等价类称为 (U, R) 的初等集 (elementary set)。为方便起见, 我们将空集也认为是初等集, 若干初等集的并集称之为 (U, R) 的可定义集 (definable set)。等价类将全集 U 剖分成若干个不交子集 $[\omega_1], [\omega_2], \dots, [\omega_m]$ 。这里 $[\omega_i]$ 表示是由描述词 ω_i 所标识的等价类。又设 Ω 是所有描述词 ω_i 的集合, 即 $\Omega = \{\omega_1, \omega_2, \dots, \omega_m\}$ 。 U 的子集 A 表示一个概念, 则 A 的下近似集 (lower approximation) IA 是包含在 A 中的最大可定义集:

$$IA = \bigcup_{[\omega_i] \subseteq A} [\omega_i]$$

而 A 的上近似集 (upper approximation) CA 是包含 A 的最小可定义集:

$$CA = \bigcup_{[\omega_i] \cap A \neq \emptyset} [\omega_i]$$

3 有限拓扑布尔代数的表示定理

我们知道, 对于有限的布尔代数有 Stone 表示定理, 即任何有限布尔代数同构于某幂集代数 $(P(S), U, \cap, -, \Phi, S)$ 。对于一个非空集合 S 及 S 上的等价关系 R , 我们可以导出 S 的一个一元运算 I 下近似。容易证明, $(P(S), U, \cap, -, I, \Phi, S)$ 是一个拓扑布尔代数。从这里我们自然想到一个有限的拓扑布尔代数是否也同构于某个集合 S 的幂集 $P(S)$ 关于运算 $U, \cap, -,$ 及 I 的拓扑布尔代数, 其中 $U, \cap, -$ 是普通集合的

*) 本文得到国家教委博士点基金、陕西省自然科学基金资助。祝峰 博士生, 主要研究人工智能与计算机基础理论。何华灿 教授, 博士生导师, 主要研究人工智能基础理论和应用、泛逻辑学等。

并、交、补运算, I 是由 S 上的某个等价关系 R 所导出的 $P(S)$ 上的下近似运算, 当然, C 也就是 S 上的由 I 所导出的上近似运算。下面我们主要讨论这个问题。

对于一般情况而言, 上面的猜想是不成立的。举例如下:

对于一个非空集合 S , $(P(S), \cup, \cap, -, \Phi, S)$ 是一个布尔代数, 任取 $a \in S$, 令 $B = S - \{a\}$, 作 $P(S)$ 上的一元运算 I 为 $IS = S, IB = B$, 对其它 $C \subseteq S, I(C) = \Phi$ 。则 $(P(S), \cup, \cap, -, 1, \Phi, S)$ 是一个拓扑布尔代数, 但 S 中没有一个等价关系 R , 使得 I 是由 R 导出的下近似运算, 因为要是有这样的 R 的话, 由上面的定义, B 必是一个等价类, 故 $\{a\}$ 也是一个等价类, 根据下近似运算的定义, 应有 $I\{a\} = \{a\}$, 而我们所定义的则是 $I\{a\} = \Phi$ 。

但对于满足一定条件的拓扑布尔代数, 上面的猜想确实是成立的。

定义2 设 $\langle A, \cup_1, \cap_1, -, 1_1, 0_1, I_1 \rangle$ 和 $\langle B, \cup_2, \cap_2, -, 1_2, 0_2, I_2 \rangle$ 是两个拓扑布尔代数, 若存在一个从 A 到 B 的一一映射 f , 使得对任意 $x, y \in A$, 有:

$$(2.1) f(x \cup_1 y) = f(x) \cup_2 f(y);$$

$$(2.2) f(x \cap_1 y) = f(x) \cap_2 f(y);$$

$$(2.3) f(-_1 x) = -_2 f(x);$$

$$(2.4) f(I_1 x) = I_2 f(x)$$

则称 $\langle A, \cup_1, \cap_1, -, 1_1, 0_1, I_1 \rangle$ 和 $\langle B, \cup_2, \cap_2, -, 1_2, 0_2, I_2 \rangle$ 是同构的。

上下关联条件: 设 $\langle A, \cup, \cap, -, 1, 0, I \rangle$ 是拓扑布尔代数, 若 $Iu = u \Rightarrow Cu = u$, 亦即 $I - u = -u$, 则称拓扑布尔代数 $\langle A, \cup, \cap, -, 1, 0, I \rangle$ 是满足上下关联条件。

$\langle P(S), \cup, \cap, -, I, \Phi, S \rangle$ 显然满足上下关联条件。

命题 对于满足上下关联条件的拓扑布尔代数 $\langle A, \cup, \cap, -, 1, 0, I \rangle$, 若 $Iu = u, Iv = v$, 则 $I(u \cup v) = u \cup v, I(u \cap v) = u \cap v$ 。

证明: 由 $Iu = u, Iv = v$ 和上下关联条件得: $I - u = -u, I - v = -v$ 。而 $-(u \cup v) = -u \cap -v$, 于是, $I - (u \cup v) = I(-u \cap -v) = (I - u) \cap (I - v) = -u \cap -v = -(u \cup v)$ 。再由上下关联条件得:

$$I(u \cup v) = u \cup v$$

$$\text{同理可得: } I(u \cap v) = u \cap v$$

推论 对于满足上下关联条件的拓扑布尔代数 $\langle A, \cup, \cap, -, I, 0, 1 \rangle$, 若 $Iu_1 = u_1, \dots, Iu_m = u_m$, 则 $I(u_1 \cup \dots \cup u_m) = u_1 \cup \dots \cup u_m$ 。

定理1 设有限集合 A 是关于运算 $\cup, \cap, -$, 和 I 的拓扑布尔代数, 且满足上下关联条件, 那么存在一个有限集 S 及 S 上的一个等价关系 R , 使得 $\langle A, \cup, \cap, -, I, 0, 1 \rangle$ 同构于 $\langle P(S), \cup, \cap, -, I, \Phi, S \rangle$, 其中

I 是由 R 导出的下近似运算。

证明: 令 $B = \{u \in A: I_1 u = u, u \neq 0\}$, 由 $I_1 1 = 1$, 有 $1 \in B$, 即 B 非空。又 A 有限, 得 B 是有限非空集合。于是 B 中必有极小元, 设 $u_1, u_2, \dots, u_m$ 是 B 中所有不同极小元。下面我们证明:

(a) 设 $u \in A$, 若对某个 $i (1 \leq i \leq m)$, 有 $u < u_i$, 则 $I_1 u = 0$;

(b) $u_i \cap_j u_j = 0, 1 \leq i, j \leq m, i \neq j$;

(c) $u_1 \cup_2 u_2 \cup_3 \dots \cup_m u_m = 1$;

(d) $u_i \leq (u_1 \cup_2 \dots \cup_{j-1} u_{j-1} \cup_{j+1} u_{j+1} \dots \cup_m u_m)$ 不成立;

(e) 对于 $u \in A, I_1 u$ 必是 $u_1, u_2, \dots, u_m$ 中若干个元之并。

由 $u < u_i$, 得 $I_1 u \leq u < u_i$, 另外 $I_1 1 = I_1 u$ 。由 B 的定义和 u_i 的极小性, 必有 $I_1 u = 0$ 。因此 (a) 成立。

由于 u_i, u_j 不相等且无大小关系, 所以 $u_i \cap_1 u_j < u_i$ 。于是由 (a) 证得 (b)。

设 u 是 $u_1 \cup_2 u_2 \cup_3 \dots \cup_m u_m$ 在 A 中的补, 即 $u = -_1 (u_1 \cup_2 u_2 \cup_3 \dots \cup_m u_m) = -_1 u_1 \cap_1 -_1 u_2 \cap_1 \dots \cap_1 -_1 u_m$ 。于是由上下关联条件, $I_1 u = (I_1 -_1 u_1) \cap_1 (I_1 -_1 u_2) \cap_1 \dots \cap_1 (I_1 -_1 u_m) = -_1 u_1 \cap_1 -_1 u_2 \cap_1 \dots \cap_1 -_1 u_m = -_1 (u_1 \cup_2 u_2 \cup_3 \dots \cup_m u_m)$, 即 $I_1 u = u$ 。故要么 $u = 0$, 这说明 $u_1 \cup_2 u_2 \cup_3 \dots \cup_m u_m = 1$; 要么 $u \neq 0$, 这时 $u \in B$, 由 $u_1, u_2, \dots, u_m$ 的极小性和 B 的有限性, 必有某个 $u_i, 1 \leq i \leq m$ 使得 $u_i \leq u$ 。这样的话, 由 $u \cap_1 (u_1 \cup_2 u_2 \cup_3 \dots \cup_m u_m) = 0$, 得 $(u_i \cap_1 u_1) \cup_2 \dots \cup_m (u_i \cap_1 u_m) = 0$ 。得 $u \cap_1 u_i = 0$, 而 $u \cap_1 u_i = u_i$ 矛盾。故必有 $u = 0$, 也就是 $u_1 \cup_2 u_2 \cup_3 \dots \cup_m u_m = 1$ 。这样就证明了 (c)。

对于 (d), 事实上有: $u_i \cap_1 (u_1 \cup_2 \dots \cup_{j-1} u_{j-1} \cup_{j+1} u_{j+1} \dots \cup_m u_m) = 0$

对于 (e), 当 $I_1 u = 0$ 时显然成立。若 $I_1 u \neq 0$, 则 $I_1 u \in B$, 于是由 B 的有限性和 $u_1, u_2, \dots, u_m$ 极小性, 其中有若干个元小于等于 $I_1 u$, 我们不妨设 $u_1, u_2, \dots, u_m (m \leq n)$ 是 $u_1, u_2, \dots, u_m$ 中所有小于等于 $I_1 u$ 的元。

我们现在证明: $I_1 u = u_1 \cup_2 u_2 \cup_3 \dots \cup_m u_m$

显然 $u_1 \cup_2 u_2 \cup_3 \dots \cup_m u_m \leq I_1 u$ 。要是等式不成立的话, 取 $v = I_1 u -_1 (u_1 \cup_2 u_2 \cup_3 \dots \cup_m u_m)$, 由命题及其推论, $I_1 v = v$, 因此 $v \in B$ 。故必有某个 $u_i (1 \leq i \leq n)$, 使得 $u_i \leq v$ 。即 $u_i \leq I_1 u -_1 (u_1 \cup_2 u_2 \cup_3 \dots \cup_m u_m) = I_1 u \cap_1 (u_{m+1} \cup_2 \dots \cup_m u_n)$, 也就是 $u_i \leq I_1 u$ 且 $u_i \leq u_{m+1} \cup_2 \dots \cup_m u_n$ 。若 $1 \leq i \leq m$, 则后一个式子不成立。若 $m+1 \leq i \leq n$, 则前一个式子不成立。这个矛盾证明了结论 (e)。

由 Stone 表示定理, 存在 S , 使得布尔代数 $\langle A, \cup, \cap, -, 0, 1 \rangle$ 同构于布尔代数 $\langle P(S), \cup, \cap, -, \Phi, S \rangle$, 也就是存在一个从 A 到 B 的一一映射 f , 使得对任意 $x, y \in A$, 有:

$$(T1) \quad f(x \cup y) = f(x) \cup f(y)$$

$$(T2) \quad f(x \cap y) = f(x) \cap f(y)$$

$$(T3) \quad f(-_j x) = -_j f(x)$$

由(b), (c)及f的性质,有

$$(\alpha) \quad f(u_i) \cap f(u_j) = \Phi, 1 \leq i, j \leq n, i \neq j$$

$$(\beta) \quad f(u_1) \cup \dots \cup f(u_n) = S$$

于是存在S上的一个等价关系R,使得 $f(u_1), \dots, f(u_n)$ 是其初等集.这样就导出了一个S上的下近似运算I.现证: $f(I_1 u) = If(u)$

由下近似的特性, $If(u)$ 必是 $f(u_1), f(u_2), \dots, f(u_n)$ 中若干个的并.为方便起见,不妨设 $If(u) = f(u_1) \cup f(u_2) \cup \dots \cup f(u_m)$, $m \leq n$,于是由f的一一对应性及下粗集的定义,有 $v_{m+1}, \dots, v_n \in A, v_i < u_i, i = m+1, \dots, n$,使得:

$$f(u) = f(u_1) \cup \dots \cup f(u_m) \cup f(v_{m+1}) \cup \dots \cup f(v_n)$$

根据上式及f的性质,有

$$u = u_1 \cup \dots \cup u_m \cup v_{m+1} \cup \dots \cup v_n$$

$$\text{于是 } u_1 \cup \dots \cup u_m \leq u$$

$$u_1 \cup \dots \cup u_m = I_1(u_1 \cup \dots \cup u_m) \leq I_1 u$$

由(e), $I_1 u$ 是 $u_1, u_2, \dots, u_n$ 中若干个元的并.我们现在证明 $u_{m+1}, \dots, u_n$ 不是这些元之一.要不然,有 $m+1 \leq i \leq n, u_i \leq I_1 u \leq u$,于是, $u_i = u_i \cap I_1 u = u_i \cap (u_1 \cup \dots \cup u_m \cup v_{m+1} \cup \dots \cup v_n) = v_i$ 矛盾.因此 $I_1 u = u_1 \cup \dots \cup u_m$.这就证明了 $f(I_1 u) = If(u)$.

至此定理证毕.

结论 本文对粗集与拓扑布尔代数之间的关系进行了深入的讨论,定理1是在添加了上下关联条件后得到的关于有限拓扑布尔代数的表示定理,该定理全面刻画了有限拓扑布尔代数和粗集的内在关系,它类似于Stone表示定理所体现的布尔代数与普通集合及其运算的内在关系,逻辑上刻画了粗集中的下近似运算.

(下转第78页)

《计算机科学》参考文献表的著录格式

《计算机科学》采用顺序编码制.在文后参考文献表中,各条文献按在论文中的文献序号排列顺序,项目应完整,内容、著录格式与符号应正确,所用参考文献应列出主要者,一般尽量限制在10篇内,下面列出著录原则和常见的几种著录格式:

1 著录原则

原则1 只著录最必要、最新的文献,最必要者直接引用的精选文献,最新者非一般知识性内容或某些陈旧史料.

原则2 只著录公开发表的文献.

原则3 采用标准化的著录格式.应遵循已有的国际标准和国家标准(BG7714-87《文后参考文献著录规则》).

2 著录格式

a. 专著

标引项顺序号 著者.书名.版本.其他责任者.出版地:出版者,出版年.文献数量(选择项)

b. 专著中析出的文献

标引项顺序号 作者.题名.见(In):原文献责任者.书名.版本.出版地:出版者,出版年.在原文献中的位置

c. 论文集中析出的文献

标引项顺序号 作者.题名.见(In):编者.文集名.出版地:出版者,出版年.在原文献中的位置

d. 期刊中析出的文献

标引项顺序号 作者.题名.其他责任者.刊名,年,卷(期):在原文献中的位置

e. 报纸中析出的文献

标引项顺序号 作者.题名.报纸名.年-月-日(版次)

f. 专利文献

标引项顺序号 专利申请者.专利题名.专利国别.专利文献种类,专利号.出版日期

g. 学位论文或技术报告

标引项顺序号 作者.题名:[学位论文或技术报告].保存地:保存者,年份

h. 会议论文

标引项顺序号 作者.题名.会议名称,会址,会议年份

4. 实例验证

实验的目的主要是为了验证本算法的有效性,并从运行效率、约简率等方面来验证本算法是否优于其它的特征属性选择算法。我们从UCI机器学习数据集中挑选几个有典型意义的数据集,并选择RS做为参照来验证算法的优良。另外,我们为了科学地验证本算法的有效性,防止由于先验知识所带来的干扰,设评价函数中 $A(X)=0$ 。

在实验结果中,我们发现两种算法所获得的特征属性个数差不多,可见GA的特征属性选择能力并不明显强于RS的属性约简能力,这主要是由于我们采用的属性集评价标准一致,也没有加入用户的经验知识。然而由于其内在的并行性,其效率明显优于RS,特别是对于属性数目较多、实例个数较多的FAS问题,GA的运算效率明显优于单纯的RS算法。

结束语 本文以粗糙集理论为基础,给出了一个基于属性依赖程度的属性重要度标准,并结合遗传算法,提出了一个特征属性选择算法。该算法具有编码简单、运行效率高的特点,特别对于解决巨维数据的特征属性选择问题取得了一定的效果,克服了粗糙集约简方法计算量大、效率低的不足。本算法不仅可以应用于人工智能、模式识别,而且在数据库知识发现的数据预处理中也有广泛的应用。

但是,本算法受粗糙集属性重要度的影响,只能处理离散属性,如何基于粗糙集理论构造能够评测连续属

性的测度标准,还留待以后的工作进一步研究。另外,如何运用并行遗传算法来进一步提高运算效率,也是特征属性选择的一个可以进一步研究的问题。

参考文献

- 1 Hu X H, Cercone N. Learning in relation database: a Rough Set approach. *Computational Intelligence*, 1995, 11(2): 323~338
- 2 Jelonek J, et al. Rough Set reduction of attributes and their domains for neural networks. *Computational Intelligence*, 1995, 11(2): 339~347
- 3 韩祯祥, 张琦, 等. 粗糙集理论及其应用综述. *控制理论与应用*, 1999, 16(2): 153~157
- 4 王珏, 王任, 等. 基于 Rough Set 理论的“数据浓缩”. *计算机学报*, 1998, 21(5): 393~399
- 5 王珏, 苗夺谦, 等. 关于 Rough Set 理论与应用的综述. *模式识别与人工智能*, 1996, 9(4): 337~344
- 6 曾黄麟. 粗糙集理论及其应用. 重庆: 重庆大学出版社, 1998
- 7 王志海, 胡可云, 等. 基于粗糙集理论的知识发现综述. *模式识别和人工智能*, 1998, 11(2): 346~351
- 8 刘明吉, 王秀峰. 数据挖掘中的数据预处理. *计算机科学* (已录用)
- 9 赛英, 陈文伟. 从数据库中发现知识的方法研究与应用. *管理科学学报*, 1999, 2(3): 92~96
- 10 常犁云, 王国胤, 等. 一种基于 Rough Set 理论的属性约简及规则提取方法. *软件学报*, 1999, 10(11): 1206~1211
- 11 Pawlak Z. Rough Sets. *Theoretical Aspects of Reasoning About Data*. Kluwer, Dordrecht, 1992
- 12 Pawlak Z, et al. Rough Sets. *Comm. ACM*, 1995, 38(11)
- 13 何华灿, 等. 经验性思维中的泛逻辑. 北京: 中国科学(E辑), 1996, 1
- 14 何华灿, 等. 人工智能导论. 西安: 西北工业大学出版社, 1988
- 15 祝峰, 何华灿. 粗糙集的公理化. *计算机学报* (已录用)
- 16 祝峰, 何华灿. Logical Properties of Rough Sets. *HPC-Asia2000* (已录用), Beijing, May, 2000
- 17 祝峰, 何华灿, 等. 粗糙集中粗元的结构及其拓广. *计算机科学*, 2000, 27(6)

(上接第81页)

参考文献

- 1 Chuchro M. On Rough Sets in Topological Boolean Algebras. In: *Rough sets, Fuzzy Sets and Knowledge Discovery*. Springer-Verlag, 1994
- 2 Rosiowa H, Sikorski R. *The Mathematics of the Metamathematics*. PWN, Warsaw, 1968
- 3 Bonkowski Z. A Certain Conception of Calculus of Rough Sets. *Notre Dame Journal of Formal Logic*, 1992, 33(3)
- 4 Pawlak Z. Rough Sets: An Algebraic and Topological Approach. [ICS PAS Report] Pol. Ac. Sc., Warsaw, 1982