

粗集

粗元

知识发现

数据挖掘

数据库

10

38-39, 34

粗集中粗元的结构及其拓广*

The Rough Elements in the Rough Set; Their Structures and Generalizations

祝峰 何华灿 周延泉 胡麒 王辉

(西北工业大学 西安 710072)

TP18

TP311.13

Abstract This paper gives a deep discussion about the structures of rough elements in rough sets and generalizes the rough element concept. With the concepts of the lower rough element and upper rough element, it presents the role and the nature of rough elements.

Keywords Rough set, Proper rough element, Rough element, Upper rough element, Lower rough element

1 引言

在决策过程中, 如何处理不确定性, 已吸引了不少学者的注意, 对此, 人们提出了各种各样关于不确定性信息的表示模型, 及不确定性推理方法, 如 Shortliffe 和 Buchanan 的不确定因子法, Duda 的概率逻辑, Dempster 和 Shafer 提出的证据理论方法, Zadeh 的模糊集理论, Dubois 和 Prade 提出的可能度推理方法等。近年来, 波兰人 Pawlak 建立的粗集理论在学术界影响很大, 这是因为粗集理论为数据挖掘和知识发现领域提供了不同于常规数据库方法的一种有效而新颖的理论, 粗集理论的实际应用与理论探讨是当前计算机科学中的一个热点问题。本文打算深入研究粗集的结构及其拓广。

2 基本概念

首先给出粗集的定义。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。我们称序偶对 (U, R) 为近似空间, 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)。为了方便起见, 我们将空集也认为是初等集。若干个初等集的并集称之为 (U, R) 的可定义集。等价类将全集 U 剖分成若干个不交子集 $[\omega_1], [\omega_2], \dots, [\omega_n]$ 。记 U/R 为所有初等集构成的集合。设 A 是 U 的子集, 则 A 的下近似集 $L(A)$ 是包含在 A 中的最大可定义集:

$$L(A) = \bigcup_{[\omega_i] \subseteq A} [\omega_i]$$

而 A 的上近似集 $U(A)$ 是包含 A 的最小可定义集:

$$U(A) = \bigcup_{[\omega_i] \cap A \neq \emptyset} [\omega_i]$$

以下的定义1到定义6引自文[1]。

定义1 设 X 是 U 的一个子集。(1) 如果 $Y \subseteq X, U(Y) = U(X)$, 则称 Y 是 X 的上样板; (2) 如果 Y 是 X 的一个样板且没有 X 的样板 Z , 使得 $Z \subset Y$, 则称 Y 是 X 的一个最小上样板。

定义2 若 $L(X) = L(Y), U(X) = U(Y)$, 则称 X, Y 是粗相等的, 记为 $X \approx Y$ 。

显然 \approx 是一个等价关系, 记 X 所在的等价类为 $[X]_{\approx}$ 。

定义3 若 $L(X) \subseteq L(Y), U(X) \subseteq U(Y)$, 则称 $[X]_{\approx}$ 粗小于等于 $[Y]_{\approx}$, 记为 $[Z]_{\approx} \leq [Y]_{\approx}$ 。

粗集的运算给出如下:

定义4 设 $[X]_{\approx}, [Y]_{\approx}$ 是两个粗集。(1) (粗集交运算) $[X]_{\approx} \cap [Y]_{\approx} = [L(X) \cap L(Y) \cup P]_{\approx}$, 其中 P 是 $U(X) \cap U(Y)$ 一个最上小样板。(2) (粗集并运算) $[X]_{\approx} \cup [Y]_{\approx} = [L(X) \cup L(Y) \cup P]_{\approx}$, 其中 P 是 $U(X) \cup U(Y)$ 一个最小上样板。

定义5 如果下列条件成立, 则称集合 $Y \subseteq U$ 是粗集 $[X]_{\approx}$ 的一个专有粗元 (Proper rough element), 记为 $Y \in [X]_{\approx}$: (1) $Y \neq \emptyset$; (2) $[Y]_{\approx} \leq [X]_{\approx}$; (3) 不存在 $Z \subseteq U$, 使得 $Z \neq \emptyset, [Z]_{\approx} \leq [Y]_{\approx}, [Z]_{\approx} \neq [Y]_{\approx}$ 。

定义6 如果下列条件成立, 则称集合 $Y \subseteq U$ 是粗

*) 本文得到国家教委博士点基金(98069923)、陕西省自然科学基金(98X15)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。周延泉 博士生, 主要研究人工智能与数据挖掘, 胡麒、王辉 博士生, 主要研究人工智能。

集 $[X]_{\sim}$ 的一个粗元,记为 $Y \in_{\sim} [X]$, (1) $Y \neq \Phi$; (2) $[Y]_{\sim} \leq_{\sim} [X]_{\sim}$; (3)存在 $Z \in U/R$ 使得 $U(Y) = Z$.

3 专有粗元和粗元的结构

对于任一 $X \subset U$,由粗集的定义有

$$U(X) = L(X) \cup X_1$$

$$U = U(X) \cup X_2$$

这里 $L(X)$, X_1 与 X_2 是三个不同的可定义集,因此有下标集 A, B 和 C 使得:

$$L(X) = \bigcup_{a \in A} \omega_a \quad X_1 = \bigcup_{a \in B} \omega_a \quad X_2 = \bigcup_{a \in C} \omega_a$$

更进一步,将 A 分成 A_1 与 A_2 的两个不交部分,使得当 $a \in A_1$ 时, ω_a 是单点集, $a \in A_2$ 时, ω_a 多于一个点,而由上、下近似集的定义,对于 $a \in B, \omega_a$ 必为多于一点的集合,下面是关于专有粗元的结构性质:

定理1 对于如上所述的 $[X]_{\sim}$ 有:

- (1)当 $a \in A_1$ 时, ω_a 是专有粗元
- (2)当 $a \in A_2$ 时, ω_a 的非空真子集是专有粗元
- (3)当 $a \in B$ 时, ω_a 的非空真子集是专有粗元
- (4)除(1)、(2)、(3)所述的以外,无其它专有粗元.

证明:(1)该结论比较明显。(2)、(3)设 Y 是某个 ω_a 的非空真子集,其 $a \in A_2$ 或 $a \in B$,显然 $L(Y) = \Phi \subseteq L(X)$, $U(Y) = \omega_a \subseteq X$,即 $[Y]_{\sim} \leq_{\sim} [X]_{\sim}$.若有 $Z \neq \Phi$ 使得 $[Z]_{\sim} \leq_{\sim} [Y]_{\sim}$,则有 $L(Z) \subseteq L(Y) = \Phi$, $U(Z) \subseteq U(Y) = \omega_a$,而 ω_a 是初等集,因此 $L(Z) = \Phi$, $U(Z) = \omega_a$,于是 $[Z]_{\sim} = [Y]_{\sim}$,故 Y 是专有粗元.

(4)对于任一个 $Y \subseteq U$,若 Y 含有 $U(X)$ 以外的元,则 $Y \subseteq X$ 不成立,因此 Y 不是专有粗元.若 Y 含有 $\omega_a (a \in A \text{ 或 } a \in B)$ 中的不同初等集中的元 x, y ,则有:

若 $x \in \omega_a, y \in \omega_\beta, a \in A_1, \beta \in A_1$.显然 $L(Y) \supseteq \omega_a \cup \omega_\beta$,取 $Z = \{x\}$,有 $L(Z) = U(Z) = \omega_a \subseteq L(Y) \subseteq U(Y)$,即 $Z \neq \Phi, [Z]_{\sim} \leq_{\sim} [Y]_{\sim}$ 且 $[Z]_{\sim} \neq [Y]_{\sim}$.

若 $x \in \omega_a, y \in \omega_\beta, a \in A_1, \beta \in A_2$.显然 $L(Y) \supseteq \omega_a, U(Y) \supseteq \omega_a \cup \omega_\beta$,取 $Z = \{x\}$,有 $L(Z) = U(Z) = \omega_a \subseteq L(Y) \subseteq U(Y)$,即 $Z \neq \Phi, [Z]_{\sim} \leq_{\sim} [Y]_{\sim}$ 且 $[Z]_{\sim} \neq [Y]_{\sim}$.

对于其它几种情况,同样可证.

定理2 对于如上所述的 X ,有:

- (1)当 $a \in A$ 时, ω_a 的非空子集是粗元
- (2)当 $a \in B$ 时, ω_a 的非空真子集是粗元
- (3)除(1)、(2)所述的以外,无其它粗元.

证明:(1)设 $\Phi \subset Y \subseteq \omega_a, a \in A, L(Y) \subseteq \omega_a \subseteq L(X), U(Y) = \omega_a \subseteq U(X)$,因此 $[Y]_{\sim} \leq_{\sim} [X]_{\sim}$.取 $Z = \omega_a$,即有 $Z \in U/R$ 且 $U(\omega_a) = Z$.

(2)设 $\Phi \subset Y \subset \omega_a, a \in B, L(Y) = \Phi \subseteq L(X), U(Y) = \omega_a \subseteq U(X)$,因此 $[Y]_{\sim} \leq_{\sim} [X]_{\sim}$.取 $Z = \omega_a$,即有 $Z \in U/R$ 且 $U(\omega_a) = Z$.

(3)设 Y 是粗元,由 $[Y]_{\sim} \leq_{\sim} [X]_{\sim}, Y$ 一定不含 $U(X)$

以外的元,又由 $Z \in U/R$,使得 $U(Y) = Z, Y$ 一定不可能含有不同初等集中的元.由 $L(Y) \subseteq L(X)$,可知 $Y \neq \omega_a, a \in B$,因为 $L(\omega_a) = \omega_a$,而对 $a \in B, \omega_a \subseteq L(X)$ 不成立.证毕.

从定理1和定理2可以看出,专有粗元和粗元的主要区别在于当 $a \in A_2$ 时, ω_a 的非空真子集是专有粗元,特别地,当 $a \in A_2$ 时, ω_a 不是专有粗元,而当 $a \in A_2$ 时, ω_a 的非空子集都是粗元,特别地当 $a \in A_2$ 时, ω_a 本身是粗元.这样,作为专有粗元,它无法区分 $a \in A_2$ 还是 $a \in B$ 时的 ω_a ,而粗元则可以区分.

4 粗元的拓广

下面我们将粗元的概念进行拓广:

定义7(下粗元的定义) 如果下列条件成立,则称集合 $Y \subseteq U$ 是粗集 $[X]_{\sim}$ 的一个下粗元,记为 $Y \in_{\sim} [X]_{\sim}$. (1) $Y \neq \Phi$; (2) $L(Y) \subseteq L(X)$; (3)存在 $Z \in U/R$ 使得 $U(Y) = Z$.

定理3(下粗元的结构) 对于如上所述的 $[X]_{\sim}$,有:

- (1)当 $a \in A$ 时, ω_a 的非空子集是下粗元;
- (2)当 $a \in B \cup C$ 时, ω_a 的非空真子集是下粗元;
- (3)除(1)、(2)所述的以外,无其它下粗元.

证明:(1)设 $\Phi \subset Y \subseteq \omega_a, a \in A, L(Y) \subseteq \omega_a \subseteq L(X)$,取 $Z = \omega_a$,即有 $Z \in U/R$ 且 $U(\omega_a) = Z$.

(2)设 $\Phi \subset Y \subset \omega_a, a \in B \cup C, L(Y) = \Phi \subseteq L(X)$,取 $Z = \omega_a$,即有 $Z \in U/R$ 且 $U(\omega_a) = Z$.

(3)设 Y 是粗元,由存在 $Z \in U/R$,使得 $U(Y) = Z, Y$ 一定不可能含有不同初等集中的元.由 $L(Y) \subseteq L(X)$,可知 $Y \neq \omega_a, a \in B \cup C$,因为 $L(\omega_a) = \omega_a$,而对 $a \in B \cup C, \omega_a \subseteq L(X)$ 不成立.证毕.

定理4 $L(X) = L(Y)$ 的充要条件是它们有相同的下粗元.

证明:由定理3即可得.

定理5 设 $[X]_{\sim}$ 和 $[Y]_{\sim}$ 是两个粗集,则:(1) $Z \in_{\sim} [X]_{\sim} \cup_{\sim} [Y]_{\sim}$ 当且仅当 $Z \in_{\sim} [X]_{\sim}$ 或 $Z \in_{\sim} [Y]_{\sim}$; (2) $Z \in_{\sim} [X]_{\sim} \cap_{\sim} [Y]_{\sim}$ 当且仅当 $Z \in_{\sim} [X]_{\sim}$ 且 $Z \in_{\sim} [Y]_{\sim}$.

证明:由定理3和相交、粗并的定义即可得证.

定义8(上粗元的定义) 如果下列条件成立,则称集合 $Y \subseteq U$ 是粗集 $[X]_{\sim}$ 的一个上粗元,记为 $Y \in_{\sim} [X]_{\sim}$: (1) $Y \neq \Phi$; (2) $U(Y) \subseteq U(X)$; (3)存在 $Z \in U/R$ 使得 $U(Y) = Z$.

定理6(上粗元的结构) 对于如上所述的 $[X]_{\sim}$,有:(1)当 $a \in A \cup B$ 时, ω_a 的非空子集是上粗元;(2)除(1)所述的以外,无其它上粗元.

证明:类似于定理2的证明.

(下转第34页)

有效性检验以及永久存储等功能。

客户层和临时服务器层之间的通信是通过 HTTP 和 CORBA 来完成的,即利用 HTTP 调用相关的 servlet,从而得到所需的 CORBA 对象的引用,然后就可以使用 CORBA 机制了。相同的机制也适用于临时服务器层与应用服务器层之间的通信。而应用服务器层和数据库层之间的通信则是通过 JDBC 来完成的。

由于采用 CORBA 作为通信机制,在未来,我们也可以将其它使用 CORBA 的服务器(如用 C++实现的服务器)结合进我们的系统。

4 运行过程

当用户第一次访问应用程序时,相关的 applets 和 servlets 将从应用服务器层被下载到临时服务器层,对最终用户而言,applet 是从临时服务器直接下载的。一旦 applet 最终下载到客户端,客户就可以向临时服务器上的某个 servlet 发出服务请求(必要时,servlet 可以从应用服务器动态下载)。

servlet 接收到客户的请求后,或者在本地进行处理,或者将该请求再发送到应用服务器上。一般来说,诸如打印和文件访问之类的请求由临时服务器上的服务模块处理;而应用数据的处理请求则发送到应用服务器上,该模型有如下优点:

a)高度的可扩展性;采用 XML 和 DOM 技术使程序中的处理功能和应用数据完全分离;通过中间各层使前端客户不直接和数据库打交道以及基于 CORBA

的分布式处理环境都大大增加了系统的可扩展性。

b)大多数逻辑事务处理都在应用服务器和数据库层完成,所以客户是“瘦”客户。

c)充分利用了网络资源。对于一些静态的应用数据,可以直接缓存到临时服务器上。

d)显著地提高了系统的效率。这表现在以下几个方面:applet 下载时间减少、打印和文件访问之类的服务就可直接在离客户最近的临时服务器上实现、服务模块都是可重用模块、不影响客户就可以修改应用服务器。

e)因为客户端 applet 和服务器方的 servlet 都可以被动态下载,所以系统的安装将大大简化。

结束语 由于采用了最新的一些相关技术并借鉴了高速缓存的思想,使得基于服务的 Java 多层计算模型在系统开发效率、系统运行效率以及系统的未来扩展性方面都比原先的三层计算模型有了较大的提高。随着 Java 技术、网络技术和其它相关技术的进一步发展,Java 计算模型还会有进一步发展。至少从现在来说,我们可以预见引入 JavaBean 技术将是该计算模型进一步发展的必然趋势。

参考文献

- 1 Available at: <http://java.sun.com/products/jdk/2/docs/guide/extensions/spec.html>
- 2 Available at: <http://java.sun.com/products/jdk1.1/docs/guide/rmi/spec/>
- 3 Available at: <http://java.sun.com/products/java-server/servlets>

(上接第39页)

定理7 $U(X)=U(Y)$ 的充要条件是它们有相同的上粗元。

证明:由定理6即可得证。

定理8 设 $[X]_{\sim}$ 和 $[Y]_{\sim}$ 是两个粗集,则:(1) $Z \in U[X]_{\sim} \cup U[Y]_{\sim}$ 当且仅当 $Z \in U[X]_{\sim}$ 或 $Z \in U[Y]_{\sim}$; (2) $Z \in U[X]_{\sim} \cap U[Y]_{\sim}$ 当且仅当 $Z \in U[X]_{\sim}$ 且 $Z \in U[Y]_{\sim}$

证明:由定理6和粗交、粗并的定义即可得证。

5 上粗元、下粗元与粗元的关系

定理9 Y 是 $[X]_{\sim}$ 的粗元当且仅当 Y 是 $[X]_{\sim}$ 的下粗元和上粗元。

证明:由定理2、定理3和定理6即可得证。

定理10 $[X]_{\sim}=[Y]_{\sim}$ 的充要条件是它们有相同的粗元。

证明:由定理4和定理7即可得证。

定理11^[1] 设 $[X]_{\sim}$ 和 $[Y]_{\sim}$ 是两个粗集,则:(1) $Z \in [X]_{\sim} \cup [Y]_{\sim}$ 当且仅当 $Z \in [X]_{\sim}$ 或 $Z \in [Y]_{\sim}$;

(2) $Z \in [X]_{\sim} \cap [Y]_{\sim}$ 当且仅当 $Z \in [X]_{\sim}$ 且 $Z \in [Y]_{\sim}$

证明:由定理5、8、9和10即得证。

结论 本文对文[1]进行了更深入的讨论,拓展了粗集中的粗元概念,通过引入的上粗元和下粗元概念,对粗集的基本结构元素进行了细致的描述和刻画。

参考文献

- 1 Bonkowski Z. Algebraic Structures of Rough Sets, Rough Sets, Fuzzy Sets and Knowledge Discovery Springer-Verlag, 1994
- 2 Bonkowski Z. A Certain Conception of Calculus of Rough Sets Notre Dame Journal of Formal Logic, 1992, 33
- 3 Bryniarski E. A Calculus of Rough Sets of the first order. Bull. Pol. Ac. Math., 1989, 37
- 4 Gehrke M, Walker E. On the Structures of Rough Sets. Bull. Pol. Ac. Math., 1992, 40
- 5 苗夺谦,王珏.粗糙集理论中知识粗糙性与信息熵关系的讨论.模式识别与人工智能,1998,1
- 6 王志海,等.基于粗糙集合理论的知识发现综述.模式识别与人工智能,1998,2
- 7 何华灿,等.经验性思维中的泛逻辑.北京:中国科学(E辑),1996,1
- 8 何华灿,等.人工智能导论.西安:西北工业大学出版社,1988