

基于粗集知识推理的 IDSS 中的知识表示^{*}

钱振伟, 高怀雁, 罗艳琳

(云南大学 系统工程中心, 云南 昆明 650091)

摘要: 探讨在智能决策支持系统中的知识表达, 讨论了属性在知识表达系统中的作用, 决策表格形式的知识表达系统的性质、作用以及根据粗集理论(其基本观点是: 人类的知识表现为对对象的分类能力) 分析处理海量信息中信息的有用特征, 通过分析、推理产生最小决策规则.

关键词: 智能决策支持系统; 粗集; 依赖度; 知识推理

中图分类号: TP 391 文献标识码: A 文章编号: 0258- 7971(2003)01- 0022- 03

知识推理是智能决策支持系统中的核心, 即根据所获得的信息通过数据分析、推理, 从而产生合理的决策规则, 形成有用知识的过程. 为了处理智能数据, 就需要对知识进行符号表示. 知识表达系统就是研究将对象的知识通过指定的对象的基本特征和特征值来描述, 以便通过一定的方法从大量浩如烟海的数据中发现有用的知识或决策规则. 粗集理论(Rough Set) 作为智能信息处理技术的一个新成果, 是由波兰科学家 Z. Pawlak 教授提出来的对不完整数据进行分析、推理、学习、发现的新方法. 使用粗集算法能在保留关键信息的前提下对数据进行简约并求得知识的最小表达; 能识别数据间的依赖关系; 揭示概念内涵与推理模式, 能从经验数据中获取控制规则从而对系统进行智能控制. 根据粗集理论的方法, 知识推理就是给定知识表达系统的条件属性和结果(决策) 属性, 求出所有符合该知识的最小决策算法.

1 粗集理论的基本概念^[1,2]

设 U 是论域, $R \subseteq U \times U$ 为二元关系, 它可以用函数给出

$$R(x, y) = \begin{cases} 1 & (x, y) \in R; \\ 0 & (x, y) \notin R. \end{cases}$$

设 $x \in U$ 定义 x 的 R^- 邻域为

$$r(x) := \{y \mid (x, y) \in R\}.$$

定义广义逼近算子 $\underline{\text{apr}}_R$ 和 $\overline{\text{apr}}_R$ 如: $\forall X \subseteq U$

$$\underline{\text{apr}}_R(X) := \{x \mid r(x) \subseteq X\},$$

$$\overline{\text{apr}}_R(X) := \{x \mid r(x) \cap X \neq \emptyset\}.$$

$X \subseteq U$ 假定给定一个感兴趣的对象的论域 U , 对于任何子集, 可称为 U 中的概念或范畴, 并且 U 中的任何概念族称为关于 U 的知识. 这些概念也构成了特定论域 U 的分类. 一个 U 上的分类族定义为一个 U 上的知识库, 这样, 知识库表达了一个或一组智能机构的各种基本分类方式. 通常情况下, 用等价关系来代替分类的概念.

令 $X \subseteq U$, 且 R 为一等价关系, 当 X 为某些 R 基本范畴的并时, 称 X 是 R 可定义的, 否则 X 为 R 不可定义的. R 可定义集是论域的子集, 它可在知识库 K 中被精确定义, 而 R 不可定义集不能在这个知识库中被定义. R 可定义集称为 R 精确集, 而 R 的不可定义集称为 R 粗集. 粗集可以近似地定义, 为达到这个目的, 使用 2 个精确集(粗集的上近似和下近似) 和边界来描述.

$R_-(X) = \underline{\text{apr}}_R(X)$ 为 X 关于 R 的下近似^[3,4];

$R^-(X) = \overline{\text{apr}}_R(X)$ 为 X 关于 R 的上近似.

则 $bn_R(X) = R^-(X) - R_-(X)$ 为 X 关于 R

* 收稿日期: 2001- 12- 02

基金项目: 云南省 2001 年软科学重点资助项目(2002ZK30).

作者简介: 钱振伟(1974-), 男, 福建人, 硕士, 主要从事管理信息系统方面的研究.

的边界.

如果 $R_-(X) = R^-(X)$ 则称为明确集, 否则 X 为 $R_-(x)$ 的粗集.

$\text{pos}_R(X) = R_-(X)$ 称为 X 的 R 正域. 把 $\text{neg}_R(X) = U/R_-(X)$ 称为 X 的 R 负域. 简单地说, 正域 $\text{pos}_R(X)$ 或 X 的下近似就是那些对于知识 R 能完全确定地归入集合 X 的对象的集合. 类似地, 负域 $\text{neg}_R(X)$ 是那些对于知识 R 不属于集合 X 的元素的集合, 它们是 X 的补集. 边界域是从某种意义上论域的不确定域, 对于知识 R 属于边界域的对象不能确定地划分是属于 X 或 $\neg X$. X 的上近似是由那些对于知识 R 不能排除它们属于 X 的可能性的对象构成, 从形式上, 上近似就是正域和边界域的并集.

2 知识的表示及基于粗集的知识推理及化简知识的表示、简化及核

2.1 知识表示可通过知识表达系统来完成, 知识表达系统的基本成分是被研究对象的集合, 关于这些对象的知识是通过指定对象的属性和它们的属性值来描述的.

一个数据表知识表达系统 S 可表示为 $S = \langle U, C, D, V, F \rangle$, 其中: U 表示论域;

$C \cup D = A$ 是属性集合, 子集 C 和 D 分别称为条件属性和结果(决策)属性;

$V = V_{\alpha_1} \times V_{\alpha_2} \times V_{\alpha_3} \times \dots \times V_{\alpha_n}$ 表示属性 A 的值域, V_{α} 表示原子属性 α 的值域;

$F: U \times A \rightarrow V$ 表示从 $V \times A$ 到 V 的信息映射, 定义 $F\alpha: U \rightarrow V_{\alpha}$.

设属性集合

$$B = \{b_1, b_2, b_3, \dots, b_m\} \subseteq A,$$

$$V_B = B_{b_1} \times B_{b_2} \times B_{b_3} \times \dots \times B_{b_m}.$$

定义映射 $F_B: U \rightarrow V_B$ 表示关于属性 B 的属性值.

论域 U 关于条件属性 C 上的 R 的商集, 记为 U/R_C ;

论域 U 关于决策属性 D 上的 R 的商集, 记为 U/R_D ;

定义 U/R_B 中的等价类为事件, 则 U/R_C 为条件事件, U/R_D 为决策事件.

$C^-(D_j) = \{C_i \mid C_j \in U/R_C \text{ 且 } C_i \cap D_j \neq \emptyset\}$ 为决策事件关于条件属性的上近似;

$$C_-(D_j) = \{C_i \mid C_j \in U/R_C \text{ 且 } C_i \subseteq D_j\}$$

为决策事件关于条件属性的下近似.

设有两集合族 G, R , 其中 r 是 R 中的某一等价关系, 如 $\text{pos}_{(R-\{r\})}(G) = \text{pos}_R(G)$, 则称 r 是关于 G 可省略的, 否则为 G 不可省略的.

设有两个集合 H 和 $R, H \subseteq R$, 如 R 中的任意元素是不可省略的, 称 R 是独立的. 设 H 是独立的, 若 $\text{pos}_H(G) = \text{pos}_R(G)$, 则称 H 为 R 的 G 简化(Reduction), 从定义可知, G 关于 H 和 R 的下近似是相同的, 即维持了与 R 相同的分类能力. R 中所有不可省略关系的交集, 称为 R 的核(Core), 记为 $\text{core}(R)$, 即 $\text{core}(R) = \bigcap \text{red}(R)$. 核中的属性是影响分类的重要属性.

2.2 事件依赖性的度量 C_i 为 U/R_C 中的条件事件, D_j 为 U/R_D 中决策事件, 设决策事件依赖于条件事件的程度为映射 $CF_{ij}: C_i \rightarrow D_j$, 且 $CF_{ij} = \text{card}(C_i \cap D_j) / \text{card}(C_i)$.

如条件事件 C_j 属于或包含于决策事件 D_j 的下近似 $C_-(D_j)$ 时, $CF_{ij} = 1$;

如条件事件 C_j 属于或包含于决策事件 D_j 的上近似 $C^-(D_j)$ 时, $CF_{ij} = 0$.

对于属性集 C 导出的分类的属性子集 $B' \subseteq B$ 的重要性, 我们用两者的依赖度量的差来度量, 即

$$r_B(C) - R_{B-B'}(C) = \frac{\text{card}(\text{pos}_B(C))}{\text{card } U} - \frac{\text{card}(\text{pos}_{B-B'}(C))^{[5]}}{\text{card } U},$$

这表示当从集合 B 中去掉某些属性子集 B' 对象分类时, 分类 $U \mid \text{ind}(C)$ 的正域受到怎样的影响.

对于属性集 C 倒出的分类集的重要性, 我们用两者的依赖度量的差来度量.

2.3 基于粗集的知识推理 根据前面的介绍, 知识表达系统将论域描述为一个二维表格, 每一行描述一个对象, 每一列描述一个属性, 属性分别为条件属性和决策属性. 知识推理的过程, 首先要进行条件属性的化简, 消去重复行, 然后对每一决策规则进行冗余属性的简化. 一般情况下, 一个知识表达系统的简化不止一种, 这些简化都维持了与原有条件属性相同的分类能力, 因此要得到使用意义上的最小决策规则就要合理地选择有效属性来正确或近似地表征研究的论域.

普通情况下, 决策者会拥有对各条件对象的属

性权重的先验知识, 权重用来衡量属性的相对重要性. 在不同的决策环境下, 相同的属性对决策输出会有不同的影响, 即权重对环境敏感. 粗集理论中的属性依赖度即表达了在当前的数据环境下属性对决策规则的影响, 但它不能反映决策者的先验知识, 因此, 将二者结合作为选择有效属性的准则不失为一种合理的解决方案^[6].

2.4 具体实现算法如下

① 提出论域中各条件属性和决策属性组成二维数据视图即决策规则表;

② 确定数据分类标准, 将各属性值以标准化方式表达, 消去冗余属性;

如果该知识表达系统的决策规则不相容, 则可将它分为两个子表, 一个表为相容决策表; 另一个表为不相容决策表. 不相容决策表是由当前信息不能被推理的知识, 所以只处理相容决策表.

③ 根据 $\alpha = 1 - \frac{\text{card}(\text{pos}_{c-a}(D))^{1/7}}{\text{card}U}$ 计算各属性在当前数据信息环境下的重要性, 及属性的依赖度;

对于每一子属性的依赖度, 可由前面定义的决策条件事件依赖度取得.

当然, 也可以考察 $\text{pos}_{B-\{a\}}(C)$ 与 $\text{pos}_B(C)$ 之商的形式来表达属性 a 的重要性^[8];

④ 依赖度为 0 的属性表示去掉该属性时, 分类 U/C 的正域不受影响, 因此, 根据先验权重的排序, 消去依赖度为 0 的且先验权重最小属性;

⑤ 计算每一决策规则的核和可能的简化;

⑥ 根据一定规则选择有效决策规则的属性简化表, 取得最简规则.

在实际系统中, 每一规则可能都会有几种简化形式, 它们的组合可能会是一个很大的规则集合, 对于这样庞大的解集, 在实际系统中使用起来非常麻烦, 除非是针对特定案例进行决策. 因此, 须考虑选择最有效的属性子集来进行简正确或化近似地表达该论域. 从实践经验中可知, 人们在现实中考察一个对象时, 往往最愿意取得的属性是根据先验知识认为的权重最大的属性, 所以, 应从各规则的简化规则中尽可能选择包含的权重较大的属性的简化来表征该论域的决策规则. 这里给出这样一种实用有效的求解办法

设简化后的决策表属性集为 $\{a_1, a_2, \dots, a_m\}$ ^[9], 它们先验权重为

$$p(a_1), p(a_2), \dots, p(a_m),$$

$$P_j = \sum_{j=1}^m (O(a_j) \times p(a_j)),$$

规则 i 有 k 种可能的简化形式, 定义每种简化形式的权重为

$$p(a_1), p(a_2), \dots, p(a_m),$$

其中如果简化形式中 a_j 为指定值, 则 $O(a_j) = 1$, a_j 为非指定值则 $O(a_j) = 0$. 取每种权重最大的简化形式组合得到实用有效的简化决策规则集.

3 结束语

本文探讨了智能决策支持系统中通过条件-决策表来表达一个信息系统的知识, 在此基础上利用粗集理论结合决策者的先验知识进行分析、推理, 得到可能的简化决策规则, 然后应用实用有效的原则求得一组合理的决策规则集, 从而有效地解决了智能决策支持系统中决策规则的获取问题.

参考文献:

- [1] PAWLAK Z. Rough sets[J]. International Journal of Information and Computer Science, 1982, 1(1): 62-66.
- [2] PAWLAK Z. Rough sets: theoretical aspects of reasoning about data[J]. Kluwer Academic, 1991, 6: 19-25.
- [3] 曾黄麟. 粗集理论及其应用[M]. 重庆: 重庆大学出版社, 1996.
- [4] LIN T Y, CERCONI N. Eds. Rough sets and data mining[J]. Kluwer Academic, 1997, 2: 76-80.
- [5] 王 珏, 苗夺谦, 周育健. 关于 RoughSet 理论与应用综述[J]. 模式识别与人工智能, 1996, 7: 32-36.
- [6] 钱振伟, 高怀雁, 罗艳琳, 等. 基于在 BP 嵌入动量因子的 IDSS 函数逼近器[J]. 云南大学学报(自然科学版), 2002, 24(6): 414-416.
- [7] MROZEK A, PLOŃKA L. Rough sets in industrial applications[J]. Rough Sets in Knowledge Discovery, 1995, 11: 36-41.
- [8] 韩祯祥, 张 琦, 文福论. 粗集理论及其应用综述[J]. 控制理论与应用, 1999, 4: 67-82.
- [9] POLKOWSKI L, SKOWRON A. Eds. physica verlag [J]. Rough Sets in Knowledge Discovery, 1998, 10: 51-59.

programming codes [M] . New York: Springer-Verlag,

1987.

Predictor-corrector technique in process system optimization

LIANG Xīming

(College of Information Science & Engineering, Central South University, Changsha 410083, China)

Abstract: Sequence quadratic programming is widely used in process system optimization. We propose a predictor-corrector interior point technique to solve quadratic subproblems in sequence quadratic programming and give the equivalence between this technique and perturbed composite Newton method. Numerical experiments are made.

Key words: process system optimization; sequence quadratic programming; predictor-corrector method

* * * * *

(上接第 24 页)

The intellectual expression of IDSS based on rough set inference

QIAN Zhenwei, GAO Huayan, LUO Yalin

(Center of Systems Engineering Research, Yunnan University, Kunming 650091, China)

Abstract: Nowadays, the research for the realization of IDSS has already become the focus of various subjects. Especially, with the development of the artificial intelligence, there are many new theories and methods used in the realization of IDSS. The writers explore hereinafter the intellectual expression in IDSS, discuss the application of property in the intellectual expression system and the nature and function of the intellectual expression system of decision form and the useful character of the information which is analyzed and dealt with numerous information and based on rough set theory. Through analysis and speculation, the minimum decision rules come into being.

Key words: intelligent decision support system; rough set; dependency; intellectual inference