

基于相容关系的不完备模糊目标信息系统的分配约简^{*})

王福贵^{1,2} 李茹¹ 钱宇华¹ 阴志洲³ 曹付元¹

(山西大学计算机与信息技术学院 太原 030006)¹ (山西农业大学文理学院 山西太谷 030801)²

(北京航空航天大学 北京 100083)³

摘要 属性约简是粗糙集理论中的重要研究内容,本文在不完备模糊目标信息系统中,提出了分配约简的概念,给出了分配约简的判定定理、相应的辨识矩阵,最后给出了计算分配约简的算法,并通过实例说明了算法的有效性。这种约简是完备的目标信息系统的分配约简的推广,同时也是不完备单模糊目标信息系统的 (α, β) 精度约简的推广。

关键词 不完备模糊目标信息系统,分配约简,辨识矩阵

Tolerance Relation Based Assignment Reduction in Incomplete Information Systems with Fuzzy Objectives

WANG Fu-Gui^{1,2} LI Ru¹ QIAN Yu-Hua¹ YIN Zhi-Zhou³ CAO Fu-Yuan¹

(School of Computer and Information Technology, Shanxi University, Taiyuan 030006)¹

(School of Arts and Sciences, Shanxi Agricultural University, Taigu, Shanxi 030801)²

(Beijing University of Aeronautics and Astronautics, Beijing 100083)³

Abstract Attribute reduction is one of the important research topics in rough set theory. In this paper, we first propose a concept of assignment reduction which includes lower assignment reduction of level α , upper assignment reduction of level β and assignment reduction of level (α, β) in incomplete information systems with fuzzy objectives. We also present the judgement theorems and discernibility matrices for these reductions. Finally, we provide the algorithm of assignment reductions and illustrate its effectiveness. This assignment reduction is an extension of the assignment reduction in complete objective information systems as well as the precision reduction in incomplete information systems with single fuzzy objective.

Keywords Incomplete and fuzzy objective information system, Assignment reduction, Discernibility matrix

1 引言

粗糙集理论是波兰学者 Pawlak 于 1982 年提出的一种处理不精确、不确定和模糊知识的软计算工具。20 多年来,它已被成功地应用于人工智能、数据挖掘、模式识别与智能信息处理等领域,并越来越引起国际学术界的关注。知识约简是知识发现的重要课题,因而也是粗糙集理论的核心问题之一。众所周知,知识库中描述知识的属性并不是同等重要的,甚至其中某些属性是冗余的。所谓知识约简,就是在保持知识库在某性质不变的条件下,删除其中不相关或不重要的属性。通过知识约简,去掉不必要的属性,可以使知识表示简化,又不丢失基本信息,这正是人们所期望的。目前,许多学者对知识约简做了深入的研究,并取得了很多成果。但是,这些研究大多是在完备的信息系统、不完备的符号值信息系统^[1~5]、完备的模糊目标信息系统^[6~10]以及完备的模糊信息系统^[11,12]中进行的。由于种种原因,在现实生活中还存在一种不完备模糊目标信息系统,即这种信息系统的条件属性是不完备的,而目标属性是模糊的。对于这种信息系统至今很少学者做深入的研究。魏大宽等提出了基于相容关系、非对称相似关系以及概率相容关系下的不完备单模糊目标信息系统的精度约简等概念^[13~17]。

2 信息系统与不完备信息系统

信息系统是一个二元有序组 (U, AT) ,其中 U 是非空有限对象集, AT 是非空有限属性集,对于 AT 中的每一个属性

α 都有一个映射 $\alpha: U \rightarrow V_\alpha, V_\alpha$ 称为属性 α 的值域。

如果信息系统中至少有一个属性 $\alpha \in AT$ 的值域 V_α 包含空值(用 $*$ 表示空值),则称它是一个不完备信息系统,否则称它是完备信息系统。

对于不完备信息系统 (U, AT) , AT 的每个子集 A 确定 U 上的一个相容关系 T_A 如下:

$$T_A = \{(x, y) \subset U \times U \mid \forall a \in A, a(x) = a(y) \vee a(x) = * \vee a(y) = *\}$$

$T_A(x) = \{y \in U \mid (x, y) \in T_A\}$ 称为 x 的相容类。

相容关系 T_A 确定了 U 上的一个覆盖,记作 $U/T_A = \{T_A(x) \mid x \in U\}$,显然有 $T_A = \bigcap_{x \in U} T_A(x)$ 。

3 不完备模糊目标信息系统

3.1 不完备模糊目标信息系统的基本概念

称三元组 (U, AT, D) 是一个目标信息系统,其中 (U, AT) 是一个信息系统, (U, D) 是一个完备信息系统,非空集合 AT 称为条件属性集合,非空集合 D 称为目标属性集合。

对于目标信息系统 (U, AT, D) ,如果对于每个目标属性 $d \in D$ 是 U 到单位区间 $[0, 1]$ 上的映射, $d: U \rightarrow [0, 1]$ (即 d 为 U 上的模糊集, $d(x)$ 称为元素 x 隶属于 U 的隶属程度),则称它是一个模糊目标信息系统。

对于模糊目标信息系统 (U, AT, D) ,如果 (U, AT) 是不完备信息系统,则称它是一不完备模糊目标信息系统。

定义 1^[17] 设 (U, AT, D) 为不完备模糊目标信息系统,

^{*} 该课题受到以下项目资助:国家 863 高技术研究发展计划项目(No. 2006AA01Z142);山西省自然科学基金项目(No. 2006011028);太原市科技局创新项目。王福贵 硕士研究生,讲师,研究方向为人工智能;李茹 教授,研究方向为智能信息处理;钱宇华 讲师,博士研究生,研究方向为粗糙集理论、粒度计算;阴志洲 研究方向为人工智能;曹付元 讲师,研究方向为机器学习。

任给 $A \subseteq AT, x \in U, d \in D$ 记

$$d^A(x) = \max\{d(y) \mid y \in T_A(x)\}$$

$$d_A(x) = \min\{d(y) \mid y \in T_A(x)\}$$

则 d^A 和 d_A 均为 U 的上模糊集, 称 d^A 和 d_A 分别为模糊目标 d 关于条件属性子集 A 的上近似和下近似。

根据定义 1 显然有 $d_A(x) \leq d^A(x)$, 即 $d_A \subseteq d^A$ 。

定理 1 设 (U, AT, D) 为不完备模糊目标信息系统, $A \subseteq B \subseteq AT, d \in D, x \in U, x \in U$, 则 $d_A(x) \leq d_B(x) \leq d^B(x) \leq d^A(x)$ 。

证明: 因为 $A \subseteq B \subseteq AT$, 所以 $T_B \subseteq T_A$, 从而任给 $x \in U$ 都有 $T_B(x) \subseteq T_A(x)$, 进一步有 $d_A(x) = \min\{d(y) \mid y \in T_A(x)\} = \min\{\min\{d(y) \mid y \in T_B(x)\}, \min\{d(y) \mid y \in T_A(x) - T_B(x)\}\} = \min\{d_B(x), \min\{d(y) \mid y \in T_A(x) - T_B(x)\}\} \leq d_B(x)$ 。

另一方面 $d^A(x) = \max\{d(y) \mid y \in T_A(x)\} = \max\{\max\{d(y) \mid y \in T_B(x)\}, \max\{d(y) \mid y \in T_A(x) - T_B(x)\}\} = \max\{d^B(x), \max\{d(y) \mid y \in T_A(x) - T_B(x)\}\} \geq d^B(x)$ 。

定义 2 设 (U, AT, D) 为不完备模糊目标信息系统, 任给实数 $\alpha, \beta \in [0, 1]$, 对于任意的 $A \subseteq AT, x \in U$, 记

$$\delta_A^\alpha(x) = \{d \in D \mid d_A(x) \geq \alpha\},$$

$$\delta_A^\beta(x) = \{d \in D \mid d^A(x) \geq \beta\},$$

称 $\delta_A^\alpha(x)$ 和 $\delta_A^\beta(x)$ 分别为对象 x 的关于条件属性子集 A 的 α 水平下分配集和 β 水平的上分配集。

定理 2 设 (U, AT, D) 为不完备模糊目标信息系统, $A \subseteq B \subseteq AT, \alpha, \beta \in [0, 1], x \in U$, 则

$$(1) \delta_A^\alpha(x) \subseteq \delta_B^\alpha(x); (2) \delta_B^\beta(x) \subseteq \delta_A^\beta(x).$$

证明: (1) 任取 $d \in \delta_A^\alpha(x)$, 由定义 2 知 $d_A(x) \geq \alpha$, 由定理 1 知 $d_B(x) \geq d_A(x)$, 从而 $d_B(x) \geq \alpha$, 根据定义 2 得 $d \in \delta_B^\alpha(x)$, 所以 $\delta_A^\alpha(x) \subseteq \delta_B^\alpha(x)$ 。

(2) 任取 $d \in \delta_B^\beta(x)$, 由定义 2 知 $d^B(x) \geq \beta$, 由定理 1 知 $d^A(x) \geq d^B(x)$, 从而 $d^A(x) \geq \beta$, 根据定义 2 得 $d \in \delta_A^\beta(x)$, 所以 $\delta_B^\beta(x) \subseteq \delta_A^\beta(x)$ 。

3.2 不完备模糊目标信息系统的分配约简

定义 3 设 (U, AT, D) 为不完备模糊目标信息系统, $A \subseteq AT$ 任给实数 $\alpha, \beta \in [0, 1]$, 如果任取 $x \in U$ 都有 $\delta_A^\alpha(x) = \delta_{AT}^\alpha(x)$ ($\delta_A^\beta(x) = \delta_{AT}^\beta(x)$) 成立, 则称 A 是 (U, AT, D) 的 α 水平下分配协调集 (β 水平的上分配协调集); 如果任取 $x \in U$ 都有 $\delta_A^\alpha(x) = \delta_{AT}^\alpha(x)$ 和 $\delta_A^\beta(x) = \delta_{AT}^\beta(x)$ 成立, 则称 A 是 (U, AT, D) 的 (α, β) 水平分配协调集。

定义 4 设 (U, AT, D) 为不完备模糊目标信息系统, $A \subseteq AT, \alpha, \beta \in [0, 1]$, 如果 A 是 (U, AT, D) 的 α 水平下分配协调集 (β 水平上分配协调集, (α, β) 水平分配协调集), 并且 A 的任意真子集都不是 (U, AT, D) 的 α 水平下分配协调集 (β 水平上分配协调集, (α, β) 水平分配协调集), 则称 A 为 (U, AT, D) 的 α 水平下分配约简 (β 水平上分配约简, (α, β) 水平分配约简)。

定义 5 设 (U, AT, D) 为不完备模糊目标信息系统, 任给 $A \subseteq AT, \alpha, \beta \in [0, 1]$, 记

$$\rho_A(\alpha) = \sum_{x \in U} |\delta_A^\alpha(x)|, \rho^A(\beta) = \sum_{x \in U} |\delta_A^\beta(x)|,$$

称 $\rho_A(\alpha)$ 和 $\rho^A(\beta)$ 分别为 (U, AT, D) 关于条件属性 A 的 α 水平下分配指标和 β 水平上分配指标。

定理 3 设 (U, AT, D) 为不完备模糊目标信息系统, $A \subseteq B \subseteq AT, \alpha \in [0, 1]$, 则有 $\rho_A(\alpha) \leq \rho_B(\alpha)$ 。

证明: 任给 $x \in U$, 根据定理 2 有 $\delta_A^\alpha(x) \subseteq \delta_B^\alpha(x)$, 进一步有 $|\delta_A^\alpha(x)| \leq |\delta_B^\alpha(x)|$, 所以 $\rho_A(\alpha) \leq \rho_B(\alpha)$ 。

定理 3 说明在集合包含意义下, α 水平下分配指标 $\rho_A(\alpha)$ 关于 A 单调不减。

定理 4 设 (U, AT, D) 为不完备模糊目标信息系统, $A \subseteq B \subseteq AT, \beta \in [0, 1]$, 则有 $\rho^A(\beta) \geq \rho^B(\beta)$ 。

证明: 任给 $x \in U$, 根据定理 2 有 $\delta_B^\beta(x) \supseteq \delta_A^\beta(x)$, 进一步有 $|\delta_B^\beta(x)| \geq |\delta_A^\beta(x)|$, 所以 $\rho^A(\beta) \geq \rho^B(\beta)$ 。

定理 4 说明在集合包含意义下, β 水平上分配指标 $\rho^A(\beta)$ 关于 A 单调不减。

定理 5 设 (U, AT, D) 为不完备多模糊目标信息系统, $A \subseteq B \subseteq AT, \alpha, \beta \in [0, 1]$, 则有 $\frac{\rho_A(\alpha)}{\rho^A(\beta)} \leq \frac{\rho_B(\alpha)}{\rho^B(\beta)}$ 。

证明: 由定理 3 和定理 4 得证。

定理 6 设 (U, AT, D) 为不完备模糊目标信息系统, $A \subseteq AT, \alpha, \beta \in [0, 1]$, 则

(1) A 是 α 水平下分配协调集的充分必要条件是 $\rho_A(\alpha) = \rho_{AT}(\alpha)$;

(2) A 是 β 水平的上分配协调集的充分必要条件是 $\rho^A(\beta) = \rho^{AT}(\beta)$;

(3) A 是 (α, β) 水平分配协调集的充分必要条件是 $\frac{\rho_A(\alpha)}{\rho^A(\beta)} = \frac{\rho_{AT}(\alpha)}{\rho^{AT}(\beta)}$ 。

证明: (1) “ \Rightarrow ” 根据定义 3 和定义 5 必要性是显然的。

“ \Leftarrow ” 由定理 2 知 $\forall x \in U$ 都有 $\delta_A^\alpha(x) \subseteq \delta_{AT}^\alpha(x)$ 。假设 $\exists y \in U$ 使得 $\delta_A^\alpha(y) \subset \delta_{AT}^\alpha(y)$ 成立, 则一定有 $\rho_A(\alpha) < \rho_{AT}(\alpha)$, 这与 $\rho_A(\alpha) = \rho_{AT}(\alpha)$ 相矛盾, 所以 $\forall x \in U$ 都有 $\delta_A^\alpha(x) = \delta_{AT}^\alpha(x)$ 。得证。

(2) “ \Rightarrow ” 显然成立。

“ \Leftarrow ” 由定理 2 知 $\forall x \in U$ 都有 $\delta_B^\beta(x) \supseteq \delta_A^\beta(x)$ 。假设 $\exists y \in U$ 使得 $\delta_B^\beta(y) \supset \delta_A^\beta(y)$ 成立, 则一定有 $\rho^A(\beta) > \rho^B(\beta)$, 这与 $\rho^A(\beta) = \rho^{AT}(\beta)$ 相矛盾, 所以 $\forall x \in U$ 都有 $\delta_B^\beta(x) = \delta_A^\beta(x)$ 。得证。

(3) 由 (1) 和 (2) 的证明得证。

定理 7 设 (U, AT, D) 为不完备模糊目标信息系统, $A \subseteq B \subseteq AT, \alpha, \beta \in [0, 1]$, 如果 A 是 α 水平下分配协调集 (β 水平的上分配协调集, (α, β) 水平分配协调集), 则 B 也一定是 α 水平下分配协调集 (β 水平的上分配协调集, (α, β) 水平分配协调集)。

证明: 这里只证明当 A 是 α 水平下分配协调集时, B 也一定是 α 水平下分配协调集。

$\forall x \in U$, 由定义 3 知 $\delta_A^\alpha(x) = \delta_{AT}^\alpha(x)$, 再由定理 2 知 $\delta_A^\alpha(x) \subseteq \delta_B^\alpha(x) \subseteq \delta_{AT}^\alpha(x)$, 从而 $\delta_B^\alpha(x) = \delta_{AT}^\alpha(x)$, 所以 B 是 α 水平下分配协调集。

定理 8 设 (U, AT, D) 为不完备模糊目标信息系统, $A \subseteq AT, \alpha, \beta \in [0, 1]$, 则:

(1) A 是 (U, AT, D) 的 α 水平下分配协调集的充分必要条件是 $\forall x, y \in U$, 如果 $\exists d \in \delta_{AT}^\alpha(x), d(y) < \alpha$, 则 $y \notin T_A(x)$ 。

(2) A 是 (U, AT, D) 的 β 水平上分配协调集的充分必要条件是 $\forall x, y \in U$, 如果 $\exists d \in D - \delta_{AT}^\beta(x), d(y) \geq \beta$, 则 $y \notin T_A(x)$ 。

证明: (1) “ \Rightarrow ” A 是 α 水平下分配协调集, 即 $\forall x \in U$ 都有 $\delta_A^\alpha(x) = \delta_{AT}^\alpha(x)$ 。假设存在 $x, y \in U, d \in \delta_{AT}^\alpha(x), d(y) < \alpha$, 并且 $y \in T_A(x)$, 则 $d_A(x) = \min\{d(z) \mid z \in T_A(x)\} \leq d(y) < \alpha$, 从而 $d \notin \delta_A^\alpha(x)$, 所以 $\delta_A^\alpha(x) \neq \delta_{AT}^\alpha(x)$, 这与 $\delta_A^\alpha(x) = \delta_{AT}^\alpha(x)$ 相矛盾。必要性得证。

“ \Leftarrow ” $\forall x, y \in U$, 如果 $\exists d \in \delta_{AT}^\alpha(x), d(y) < \alpha$, 则 $y \notin T_A(x)$ 。即如果 $y \in T_A(x)$, 则 $\forall d \in \delta_{AT}^\alpha(x)$, 都有 $d(y) \geq \alpha$ 。从而 $\forall d \in \delta_{AT}^\alpha(x)$, 都有 $d_A(x) = \min\{d(z) \mid z \in T_A(x)\} \geq \alpha$, 进一步有 $d \in \delta_A^\alpha(x)$, 所以 $\delta_{AT}^\alpha(x) \subseteq \delta_A^\alpha(x)$ 。另一方面, 由定理 2 知 $\delta_A^\alpha(x) \subseteq \delta_{AT}^\alpha(x)$ 。所以有 $\delta_A^\alpha(x) = \delta_{AT}^\alpha(x)$ 。充分性得证。

(2)“ \Rightarrow ” A 是 β 水平上分配协调集,即任取 $x \in U$ 都有 $\delta_\beta^A(x) = \delta_\beta^{AT}(x)$ 。假设存在 $x, y \in U, d \in D - \delta_\beta^{AT}(x), d_i(y) \geq \beta$, 并且 $y \in T_A(x)$, 则 $d^A(x) = \max\{d(z) | z \in T_A(x)\} \geq d(y) \geq \beta$, 从而 $d \in \delta_\beta^A(x)$, 所以 $\delta_\beta^A(x) \neq \delta_\beta^{AT}(x)$, 这与 $\delta_\beta^A(x) = \delta_\beta^{AT}(x)$ 相矛盾。必要性得证。

“ \Leftarrow ” $\forall x, y \in U$, 如果 $\exists d \in D - \delta_\beta^{AT}(x), d(y) \geq \beta$, 则 $y \notin T_A(x)$ 。即如果 $y \in T_A(x)$, 则 $\forall d \in D - \delta_\beta^{AT}(x)$, 都有 $d_i(y)$

$$D_\alpha(x, y) = \begin{cases} \{a \in AT | a(x) \neq * \wedge a(y) \neq * \wedge a(x) \neq a(y)\}, \exists d \in \delta_{AT}(x) \wedge d(y) < \alpha \\ \phi, \text{ otherwise} \end{cases}$$

$$D^\beta(x, y) = \begin{cases} \{a \in AT | a(x) \neq * \wedge a(y) \neq * \wedge a(x) \neq a(y)\}, \exists d \in D - \delta_\beta^{AT}(x) \wedge d(y) \geq \beta \\ \phi, \text{ otherwise} \end{cases}$$

$$D_\alpha^\beta(x, y) = D_\alpha(x, y) \cup D^\beta(x, y)$$

则称 $D_\alpha(x, y)$ 称为 y 关于 x 的 α 水平下分配辨识属性集; $D^\beta(x, y)$ 称为 y 关于 x 的 β 水平上分配辨识属性集; $D_\alpha^\beta(x, y)$ 称为 y 关于 x 的 (α, β) 水平分配辨识属性集。记

$$D_\alpha = (D_\alpha(x, y) | x, y \in U),$$

$$D^\beta = (D^\beta(x, y) | x, y \in U),$$

$$D_\alpha^\beta = (D_\alpha^\beta(x, y) | x, y \in U),$$

则称 D_α, D^β 和 D_α^β 分别为 (U, AT, D) 的 α 水平下分配辨识矩阵、 β 水平上分配辨识矩阵和 (α, β) 水平分配辨识矩阵。

定义 7 设 (U, AT, D) 为不完备模糊目标信息系统, $\alpha, \beta \in [0, 1]$, 记

$$M_\alpha = \bigwedge_{x, y \in U} (\bigvee D_\alpha(x, y)),$$

$$M^\beta = \bigwedge_{x, y \in U} (\bigvee D^\beta(x, y)),$$

$$M_\alpha^\beta = \bigwedge_{x, y \in U} (\bigvee D_\alpha^\beta(x, y)),$$

其中当 $D_\alpha(x, y) = \phi$ 时, $\bigvee D_\alpha(x, y)$ 表示逻辑值“1”, 当 $D_\alpha(x, y) \neq \phi$ 时, $\bigvee D_\alpha(x, y)$ 表示 $D_\alpha(x, y)$ 中的所有元素的析取表达式, $\bigvee D^\beta(x, y)$ 和 $\bigvee D_\alpha^\beta(x, y)$ 的含义与 $\bigvee D_\alpha(x, y)$ 类似。

称 M_α, M^β 和 M_α^β 分别为 (U, AT, D) 的 α 水平下分配辨识公式、 β 水平上分配辨识公式和 (α, β) 水平分配辨识公式。

定理 9 设 (U, AT, D) 为不完备模糊目标信息系统, $\alpha, \beta \in [0, 1]$, 则辨识公式 $M_\alpha(M^\beta, M_\alpha^\beta)$ 的极小析取范式中的每一个极小项当中的元素构成的集合为 (U, AT, D) 的 α 水平下分配约简(β 水平上分配约简, (α, β) 水平分配约简)。

3.3 不完备模糊目标信息系统的分配约简算法

分配约简算法描述如下:

输入: 不完备模糊目标信息系统 $(U, AT, D), \alpha, \beta$ 。

输出: (U, AT, D) 的 α 水平下分配约简(β 水平上分配约简, (α, β) 水平分配约简)。

步骤 1: $\forall x \in U$, 计算相容类 $T_{AT}(x)$ 。

步骤 2: $\forall d \in D$, 计算 $d_{AT}(d^{AT}, d_{AT}$ 和 $d^{AT})$ 。

步骤 3: $\forall x \in U$, 计算 $\delta_{AT}^A(x) (\delta_\beta^{AT}(x), \delta_{AT}^A(x)$ 和 δ_β^{AT}

$$D_{0.5} = \begin{pmatrix} 0000 & 1001 & 1000 & 0000 & 0000 & 1100 & 0000 & 1000 & 0000 & 0000 \\ 1001 & 0000 & 0000 & 0011 & 1011 & 0000 & 1011 & 0000 & 1010 & 1001 \\ 1000 & 0000 & 0000 & 0100 & 1000 & 0000 & 1100 & 0000 & 0100 & 1000 \\ 0100 & 0011 & 0100 & 0000 & 0111 & 0010 & 0000 & 0010 & 0000 & 0110 \\ 0000 & 1011 & 1000 & 0000 & 0000 & 1110 & 0000 & 1011 & 0000 & 0000 \\ 1100 & 0000 & 0000 & 0010 & 1110 & 0000 & 1010 & 0000 & 1010 & 1100 \\ 0100 & 1011 & 1100 & 0000 & 0111 & 1010 & 0000 & 1010 & 0000 & 0110 \\ 1000 & 0000 & 0000 & 0010 & 1011 & 0000 & 1010 & 0000 & 1010 & 1110 \\ 1100 & 1010 & 0100 & 0000 & 1110 & 1010 & 0000 & 1010 & 0000 & 1110 \\ 0000 & 1001 & 1000 & 0000 & 0000 & 1100 & 0000 & 1000 & 0000 & 0000 \end{pmatrix}$$

其中 $D_{0.5}$ 中的元素是以 4 位 2 进制数 $u_1 u_2 u_3 u_4$ 的形式表示的, $u_i = 0$ 表示 a_i 不属于集合 $u_1 u_2 u_3 u_4, u_i = 1$ 表示 a_i 属于集合 $u_1 u_2 u_3 u_4 (i=1, 2, 3, 4)$ 。

根据 0.5 水平下分配辨识矩阵 $D_{0.5}$, 得到 0.5 水平下分

$< \beta$ 。从而 $\forall d \in D - \delta_\beta^{AT}(x)$ 都有 $d^A(x) = \max\{d(z) | z \in T_A(x)\} < \beta$, 进一步有 $d \notin \delta_\beta^A(x)$, 即 $d \in D - \delta_\beta^A(x)$, 所以 $D - \delta_\beta^{AT}(x) \subseteq D - \delta_\beta^A(x)$, 即 $\delta_\beta^A(x) \subseteq \delta_\beta^{AT}(x)$ 。另一方面, 由定理 2 知 $\delta_\beta^A(x) \supseteq \delta_\beta^{AT}(x)$ 。所以有 $\delta_\beta^A(x) = \delta_\beta^{AT}(x)$ 。充分性得证。

定义 6 设 (U, AT, D) 为不完备模糊目标信息系统, $\alpha, \beta \in [0, 1]$, 记,

(x) 。

步骤 4: 求出 $D_\alpha(D^\beta, D_\alpha^\beta)$ 。

步骤 5: 求出 $M_\alpha(M^\beta, M_\alpha^\beta)$ 的极小析取范式, 并输出 (U, AT, D) 的所有 α 水平下分配约简(β 水平上分配约简, (α, β) 水平分配约简)。

例 1 表 1 给出了一个不完备模糊目标信息系统。

表 1 一个不完备模糊目标信息系统

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
a_1	3	2	1	*	3	2	3	2	1	3
a_2	1	*	1	2	1	2	2	*	2	1
a_3	*	1	*	3	2	1	3	1	3	1
a_4	1	2	*	1	3	*	1	1	*	1
d_1	0.7	0.1	0.2	0.9	0.8	0.4	0.8	0.4	1.0	0.8
d_2	0.2	0.9	0.8	0.3	0.2	0.9	0.1	0.8	0.3	0.1
d_3	0.1	0.4	0.2	0.8	0.3	0.4	0.7	0.3	1.0	0.3

其中 $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}, AT = \{a_1, a_2, a_3, a_4\}, D = \{d_1, d_2, d_3\}$ 。

在这里只计算 0.5 水平下分配约简。

$T_{AT}(x_1) = \{x_1, x_{10}\}, T_{AT}(x_2) = \{x_2, x_6\}, \dots, T_{AT}(x_{10}) = \{x_1, x_{10}\}$ 。

各模糊目标的下近似如下:

$$(d_1)_{AT} = 0.7/x_1 + 0.1/x_2 + 0.2/x_3 + 0.8/x_4 + 0.8/x_5 + 0.1/x_6 + 0.8/x_7 + 0.4/x_8 + 0.9/x_9 + 0.7/x_{10},$$

$$(d_2)_{AT} = 0.1/x_1 + 0.9/x_2 + 0.8/x_3 + 0.1/x_4 + 0.2/x_5 + 0.8/x_6 + 0.1/x_7 + 0.8/x_8 + 0.3/x_9 + 0.1/x_{10},$$

$$(d_3)_{AT} = 0.1/x_1 + 0.4/x_2 + 0.2/x_3 + 0.7/x_4 + 0.3/x_5 + 0.3/x_6 + 0.7/x_7 + 0.3/x_8 + 0.8/x_9 + 0.1/x_{10}.$$

各个对象关于 AT 的 0.5 水平下分配集为:

$$\delta_{AT}^{0.5}(x_1) = \{d_1\}, \delta_{AT}^{0.5}(x_2) = \{d_2\}, \dots, \delta_{AT}^{0.5}(x_{10}) = \{d_1\}.$$

0.5 水平下分配辨识矩阵如下:

配辨识函数的极小析取范式为 $M_{0.5} = a_1 \wedge a_2 \wedge a_3$, 所以该信息系统有唯一的 0.5 水平下分配约简为 $\{a_1, a_2, a_3\}$ 。

结论 本文介绍了不完备模糊目标信息系统的概念, 并基于不完备信息系统中的相容关系, 提出了不完备模糊目标

信息系统的分配协调集以及分配约简的概念。当 (U, AT, D) 是不完备单模糊目标信息系统并且 $\alpha \geq \beta$ 时, (α, β) 分配协调集便是文[15]中提出的 (α, β) 协调集, 故此时的 (α, β) 分配约简为文[15]当中的 (α, β) 精度约简; 当 (U, AT, D) 是完备的符号值目标信息系统, 并且当 $\alpha > 0$ 时, α 水平下分配协调集便是文[18]当中提出的分配协调集, 从而 α 水平下分配约简为文[18]当中提出的分配约简, 所以本文提出的分配约简是不完备单模糊目标信息系统的 (α, β) 精度约简概念的推广, 同时也是完备的符号值目标信息系统的分配约简概念的推广。本文还给出了各种分配约简相应的辨识矩阵, 从而给出了相应的约简算法, 并通过实例表明了算法的有效性。

参考文献

- 1 Kryszkiewicz M. Rough set approach to incomplete information systems[J]. Information Sciences, 1998, 112: 39~49
- 2 Kryszkiewicz M. Rules in incomplete information systems[J]. Information Sciences, 1999, 113: 271~292
- 3 Leung Y, Li Deyu. Maximal consistent block technique for rule acquisition in incomplete information systems[J]. Information Sciences, 2003, 153: 85~106
- 4 Zhang Wen-Xiu, Mi Ju-Sheng. Incomplete Information System and Its Optimal Selections[J]. Computers and Mathematics with Applications, 2004, 48: 691~698
- 5 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
- 6 Zhang Mei, Wu Wei-Zhi. Knowledge Reductions in Information

- Systems with Fuzzy Decisions[J]. Journal of Engineering Mathematics, 2003, 20(2): 53~58
- 7 张梅, 朱朝晖. Fuzzy 目标信息系统的知识发现[J]. 模糊系统与数学, 2005, 19(1): 121~125
- 8 袁修久, 张文修. 模糊目标信息系统的属性约简[J]. 系统工程理论与实践, 2004, 5: 116~120
- 9 管涛, 冯博琴. 模糊目标信息系统上的知识约简方法[J]. 软件学报, 2004, 15(10): 1470~1478
- 10 袁修久, 何华灿. 优势关系下模糊目标信息系统约简的辨识矩阵[J]. 空军工程大学学报(自然科学版), 2006, 7(2): 81~84
- 11 陈德刚, 刘民, 吴澄, 李法朝. 模糊信息系统的代数结构及其约简[J]. 清华大学学报(自然科学版), 2003, 43(9): 1233~1235, 1264
- 12 管涛, 薛亮, 冯博琴. 模糊信息系统上的粗糙约简[J]. 西安交通大学学报, 2005, 39(6): 574~577, 632
- 13 Wei Dakuan, Zhou Xianzhong, Dongjun Xin, Zhiwei Chen. Variable Rough Set Model and Its Knowledge Reduction for Incomplete and Fuzzy Decision Information Systems[J]. International Journal of Information Technology, 2006, 3(2): 140~144
- 14 Wei Da-kuan, Zhou Xian-zhong. Rough Set Model in Incomplete and Fuzzy Decision Information System Based on Improved-Tolerance Relation[C]. In: 2005 IEEE International Conference on Granular Computing, Tsinghua University, China, July 2005. 278~283
- 15 魏大宽, 黄兵, 周献中. 不完备模糊目标信息系统粗集模型与知识约简[J]. 计算机工程, 2006, 32(8): 48~51
- 16 魏大宽. 基于相似关系的不完备模糊决策信息系统知识约简[J]. 湖南师范大学自然科学学报, 2006, 29(2): 18~23
- 17 魏大宽, 周献中, 黄兵. 不完备模糊决策信息系统的粗集模型与精度约简[J]. 计算机科学, 2006, 33(6): 182~185
- 18 张文修, 米据生, 吴伟志. 不协调目标信息系统的知识约简[J]. 计算机学报, 2003, 26(1): 1~7

(上接第 167 页)

的关键字集将不能很完善地表达用户的意图, 将会有太多具有意义的关键字被漏掉, 因此会产生大量的新语义关键字, 从而加重了语义发现模块的负担, 影响了系统的效率; 反之, 则会使语义发现模块失去存在的意义, 因此我们将这一问题称为“筛眼现象”。

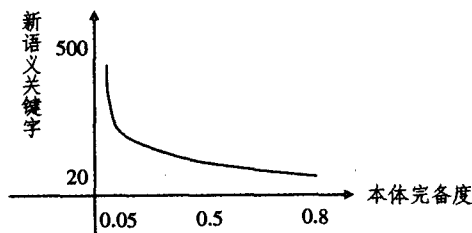


图 3 冗余互信息曲线图

其数学表达如下: 设过滤函数 F_i , 领域本体 Do , 互信息矩阵 Mo , 过滤后的互信息矩阵为 Mf 则有: $Mf = F_i(Do, Mo)$ 。互信息矩阵的过滤完全取决于其完备性, 令: W 为 Do 的完备性, 取值范围为 $[0, 1)$, R_i 为 Mf 相对于 Do 而言的冗余互信息量, R_{oi} 为 Mo 相对于 Do 而言的冗余互信息量, U_{fi} 为 Mf 针对 Do 而言的有用互信息量, U_{oi} 为 Mo 针对实际语义使用而言的有用互信息量, $U_{fi} \subseteq U_{oi}$ 。有以下两条定理:

定理 1 无论 W 取值如何, R_i 趋近于 0。

证明: 设过滤函数 F_i , 领域本体 Do , 互信息矩阵 Mo , 过滤后的互信息矩阵为 Mf 则有:

$$Mf = F_i(Do, Mo)$$

$\because Mo = \{U_{oi} \cup R_{oi} \text{ 且 } U_{oi} \cap R_{oi} = \phi\}$, U_{oi} 表示矩阵 Mo 当中的有用互信息量, R_{oi} 表示矩阵 Mo 当中的冗余互信息量。

\because 领域本体集合 Do 具有准确的语义表述能力。

$\therefore Mf \subset A \subseteq Do$, Mf 相对于 Do 的冗余度 $R_i \rightarrow 0$ (定理 1 得证)。

定理 2 $U_{fi} = W * U_{oi}$

证明: \because 令领域本体集合 Do 的冗余度为 W , 其取值范围为 $(0, 1)$

$$\therefore \text{当 } W \rightarrow 1 \text{ 时, } \lim \frac{A - Mf}{A} = 0, \text{ 即 } (U_{fi} = A - Mf) \rightarrow 1$$

$$\text{当 } W \rightarrow 0 \text{ 时, } \lim \frac{A - Mf}{A} = 1, \text{ 即 } (U_{fi} = A - Mf) \rightarrow 0$$

由上两式可得: U_{fi} 与 Do 的完备性 W 成正比, 也就是对互信息矩阵的过滤完全取决于 Do 的完备性和互信息矩阵当中的有用互信息量 U_{oi} 。

因此我们可得以下公式: $U_{fi} = W * U_{oi}$ (定理 2 得证)。

结束语 本文提出了一种基于互信息与本体相结合的共同检索模型。该模型将互信息与领域本体相结合, 利用互信息与本体在本质上的互补性, 从而在查准率、查全率和 F 量度上都优于文[2]中的 QSE_BMI 模型, 性能都有所改善。接下来的工作, 我们将主要集中在对本体净化与本体补充进行进一步的研究, 为实现本体的半自动学习和更新做准备。

参考文献

- 1 袁占亭, 张爱民, 张秋余. 基于概念的 Web 信息检索[J]. 计算机科学, 2004, 31: 13~16
- 2 夏磊, 周竹荣. 基于互信息的问句语义扩展研究[J]. 计算机工程与设计, 2008, 3
- 3 王晓琬, 关毅, 等. 计算机自然语言处理[M]. 清华大学出版社, ISBN: 7-302-10089-6, 2005. 18~21
- 4 刘炜, 李大玲, 等. 基于本体的元数据应用[M]. 上海图书馆
- 5 Chun-Xia Zhang, Cun-Gen Cao. Domain-Specific Formal Ontology of Archaeology and Its Application in Knowledge Acquisition and Analysis[J]. Computer Science, 2004(3): 290~301
- 6 Guarino N. Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration[C]. In: Paziienza MT, ed. Information Extraction: a Multidisciplinary Approach to an Emerging Information Technology, Springer Verlag, 2001. 139~170
- 7 徐立广, 金芝. 一个本体评议及本体构造工具的设计[J]. 计算机工程与应用, 2006, 25: 74~79