

不完备 XML 信息系统的粗糙集模型

殷丽凤 邓 武

(大连交通大学软件学院 大连 116028)

摘 要 随着 XML 成为网络信息表示和交换的标准以及不确定数据的广泛存在, 不确定 XML 数据库管理技术成为了当今研究的热点。首先, 允许 XML 文档叶子节点的信息值为丢失的或遗漏的空值, 提出不完备的 XML 信息系统; 其次, 提出节点的相容关系、限制相容关系以及阈值相容关系等概念, 基于粗糙集理论分别定义这 3 种关系对应的粗糙集模型; 最后, 通过实例分析表明, 限制相容关系模型能克服相容关系模型分类粗糙的缺点, 阈值相容关系模型通过合理地设置阈值可达到更好的分类效果, 从而提高了对 XML 数据的预测、分类的精确度。

关键词 粗糙集, 不完备的 XML 信息系统, 相容关系, 限制相容关系, 阈值相容关系

中图法分类号 TP311.13 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.10.055

Rough Set Models for Incomplete XML Information System

YIN Li-feng DENG Wu

(Software Technology Institute, Dalian Jiaotong University, Dalian 116028, China)

Abstract The management technology of uncertain XML database becomes today's research focus with XML being the standards of information representation and data exchange on the Internet and uncertain data existing in various fields. Firstly, the leaf nodes' information value of XML document being lost or missing null values was allowed and the incomplete XML information system was proposed. Secondly, the definitions of node tolerance relation, limited tolerance relation and threshold tolerance relation were given, and three kinds of rough set models for incomplete XML information system were defined respectively based on rough set theory. Finally, the analysis of examples shows that limited tolerance relation can overcome the disadvantage of tolerance relation rough classification, and threshold tolerance relation can achieve better classification results through reasonable threshold setting, so as to enhance the prediction and the classification accuracy of XML data.

Keywords Rough set, Incomplete XML information system, Tolerance relation, Limited tolerance relation, Threshold value tolerance relation

粗糙集理论^[1]是描述和研究不精确、不确定和不完全知识的数学工具, 用于分析和处理各种不完备信息, 从中发现隐含的知识, 揭示潜在的规律。目前, 粗糙集理论^[2]基本成熟, 主要应用在与各个学科的相互结合上。由于粗糙集理论和关系数据库的表示形式都是二维表, 因此二者之间的结合已得到了很多学者的研究, 目前粗糙关系数据库理论^[3]已经很成熟, 但现实世界中由于各种原因存在大量不确定数据, 借助粗糙集理论研究不完备的关系数据处理成为了一个新的发展方向, 文献[4-8]分析了不确定数据的特点, 建立了扩展的粗糙关系模型, 为不确定关系数据库的发展奠定了基础。

由于不确定数据的普遍存在性, 研究表示和处理不确定 XML (eXtensible Markup Language) 数据也成为一个新的研究领域。不确定数据包含概率数据以及不完备数据, 很多学者提出了概率 XML 数据处理技术^[9-12], 笔者^[13-17]对不完备 XML 数据的规范化问题进行了研究。采用概率描述不确定 XML 数据存在很多问题, 如由于概率信息的存在, 可能世界实例的数量相对于数据规模为指数倍, XML 数据又为树型结

构, 从而导致查询种类、解决问题的难度都大大增加。为了弥补概率信息描述不确定 XML 数据时加大问题难度的缺陷, 本文借助粗糙集理论不需要先验知识的优势, 允许 XML 文档叶子节点的信息值为丢失的或遗漏的空值, 对不完备的 XML 信息系统的各种粗糙集模型进行研究, 为不完备 XML 信息系统的路径约简和查询问题等方面的研究奠定了基础。

1 基本概念

本文对文献[18]中的标签树信息系统进行扩展, 允许叶子节点的取值为遗漏型或丢失型空值。下面给出相应的定义。

定义 1 不完备的 XML 信息系统是一个由 n 棵树组成的集合, 记为 $T = \{T_1, T_2, \dots, T_n\}$ 或 T 是一个六元组 $(N, E, L, f, \delta, val)$, 其中:

(1) $T_i = (N_i, E_i, L_i)$ 为集合中的第 i 棵树。

(2) N_i 是 T_i 中所有节点的集合, $LN_i \subseteq N_i$ 是叶子节点集合, $BN_i \subseteq N_i$ 是分支节点集合, $r_i \in N_i$ 是唯一的根节点,

到稿日期: 2014-10-30 返修日期: 2015-02-10 本文受国家自然科学基金项目(51475065, U1433124), 辽宁省教育厅项目(L2014191)资助。

殷丽凤(1976-), 女, 博士, 副教授, 主要研究方向为不确定 XML 数据库理论及应用, E-mail: yinlife0598@sina.com; 邓武(1976-), 男, 博士, 副教授, 主要研究方向为智能计算与信息处理、故障诊断、软件工程。

$$N_i = LN_i \cup BN_i \cup r_i, N = \bigcup_{i=1}^n N_i.$$

(3) $E_i \subseteq N_i \times N_i$ 是树 T_i 中所有边的集合。 $E = \bigcup_{i=1}^n E_i$ 。

(4) L_i 为第 i 棵树的标签集合。 $L = \bigcup_{i=1}^n L_i$ 。

(5) $f: N \rightarrow L$ 是一个映射函数, 用来决定每个节点的标签。

(6) δ 表示从节点 N 到 N 中一系列节点的部分映射, 满足对 $\forall v \in N, \delta(v) = [v_1, \dots, v_n]$ 且边 $v \times v_i \in E_i$, 其中 $i \in [1, n]$; $children(x) = \{y | y \in \delta(x)\}$ 表示节点 x 的所有孩子节点的集合。

(7) $\forall v \in LN_i, val_i: v \rightarrow dom(lab(v)), dom(lab(v))$ 表示节点 v 所在标签类型的值域, $val_i(v)$ 表示节点 v 的值, 此值可能是丢失的或遗漏的空值(用“*”表示)。

定义 2 对于 T_i 中的 2 个节点 $v', v'' \in N_i$, 若存在节点集合 $\{v_1, v_2, \dots, v_n\}$, 使得 $v_1 \in \delta(v'), v_2 \in \delta(v_1), \dots, v_n \in \delta(v_{n-1}), v'' \in \delta(v_n)$ 成立, 其中, $w_0 = lab(v'), w_1 = lab(v_1), w_2 = lab(v_2), \dots, w_n = lab(v_n), w_{n+1} = lab(v'')$ 。则称 $w = w_0/w_1/\dots/w_{n+1}$ 为一条从 w_0 到 w_{n+1} 的路径; 称 v', v_1, \dots, v_n, v'' 为通过路径 w 的一个路径节点集。用函数 $last(w)$ 表示通过路径 w 的路径节点集最后节点的集合。若 w_0 为根节点, w_{n+1} 为叶子节点, 则称 w 为全路径。

P 表示 T 的所有全路径集合, 不完备的 XML 信息系统看作所有全路径集合的并集构成的信息树的集合。

2 不完备 XML 信息系统的粗糙模型

2.1 相容关系

假设所有未知叶子节点的信息值均被认为是遗漏类型的空值(用“*”表示), 由于遗漏类型空值的存在, 需要对节点相等的定义进行扩展。

定义 3(节点相容) 设 $T = \{T_1, T_2, \dots, T_n\}$ 是一个不完备的 XML 信息系统, 任意子树 T_i 中的节点 u_i 与 T_j 中的节点 v_j 相容(简记为 $u_i \doteq_R v_j$)需要满足如下条件:

(1) $f(u_i) = f(v_j)$;

(2) 若 $children(u_i) \neq \emptyset, children(v_j) \neq \emptyset$ 成立, 则 $\forall m_i \in children(u_i), \exists n_j \in children(v_j)$ 满足 $m_i \doteq_R n_j$; 反之, 若 $children(u_i) \neq \emptyset, children(v_j) \neq \emptyset$ 成立, 则 $\forall n_j \in children(v_j), \exists m_i \in children(u_i)$ 满足 $n_j \doteq_R m_i$;

(3) 若 $children(u_i) = \emptyset, children(v_j) = \emptyset$ 成立, 则满足 $val(u_i) = val(v_j)$ 或 $val(u_i) = *$ 或 $val(v_j) = *$ 成立。

定义 4(节点相容关系) 设 $T = \{T_1, T_2, \dots, T_n\}$ 是一个不完备的 XML 信息系统, P 为 T 的全路径集合, 由节点集 $V \subseteq N$ 决定的相容关系为

$$C_V(u_i, v_j) = \{(u_i, v_j) | u_i \in V_i, \exists v_j \in V_j, u_i \doteq v_j, i, j \in [1, n]\}.$$

$$I_V^C(u_i) = \{v_j | v_j \in T \wedge C_V(u_i, v_j), i, j \in [1, n]\}$$

定义 5 在不完备的 XML 信息系统 $T = \{T_1, T_2, \dots, T_n\}$ 中, 子树集合 T 关于节点集 $V \subseteq N$ 的上、下近似和近似精度分别为:

$$T_V^C = \{u_i | I_V^C(u_i) \cap T \neq \emptyset\}, T_V^{\bar{C}} = \{u_i | I_V^C(u_i) \subseteq T\}, \rho_V^C(T) = \frac{|T_V^C|}{|T_V^{\bar{C}}|}.$$

相容关系的要求过于宽松, 下面提出限制相容关系。

2.2 限制相容关系

定义 6(节点限制相容) 设 $T = \{T_1, T_2, \dots, T_n\}$ 是一个不完备的 XML 信息系统, 任意子树 T_i 中的节点 u_i 与 T_j 中的节点 v_j 限制相容(简记为 $u_i \doteq_{RV} v_j$)需要满足如下条件:

(1) $f(u_i) = f(v_j)$;

(2) 若 $children(u_i) \neq \emptyset, children(v_j) \neq \emptyset$ 成立, 则 $\forall m_i \in children(u_i), \exists n_j \in children(v_j)$ 满足 $m_i \doteq_{RN} n_j$; 反之, 若 $children(u_i) \neq \emptyset, children(v_j) \neq \emptyset$ 成立, 则 $\forall n_j \in children(v_j), \exists m_i \in children(u_i)$ 满足 $n_j \doteq_{RM} m_i$;

(3) 若 $children(u_i) = \emptyset, children(v_j) = \emptyset$ 成立, 则满足 $(val(u_i) = val(v_j) = *) \vee ((val(u_i) \neq * \wedge val(v_j) \neq *) \wedge val(u_i) = val(v_j))$ 。

限制相容限定叶子节点的信息值都为*, 或者都不为*, 但值相等。显然限制相容关系满足自反性、对称性, 但不一定满足传递性。

定义 7(节点限制相容关系) 设 $T = \{T_1, T_2, \dots, T_n\}$ 是一个不完备的 XML 信息系统, 由节点集 $V \subseteq N$ 决定的限制相容关系为:

$$R_V(u_i, v_j) = \{(u_i, v_j) | \forall u_i \in V_i, \exists v_j \in V_j, u_i \doteq_{RV} v_j, \text{其中 } i \in [1, n], j \in [1, n]\}.$$

定义 8 在不完备的 XML 信息系统 $T = \{T_1, T_2, \dots, T_n\}$ 中, 子树集合 T 关于节点集 $V \subseteq N$ 的上、下近似和近似精度分别为:

$$T_V^R = \{u_i | I_V^R(u_i) \cap T \neq \emptyset\}, T_V^{\bar{R}} = \{u_i | I_V^R(u_i) \subseteq T\}, \rho_V^R(T) = \frac{|T_V^R|}{|T_V^{\bar{R}}|}.$$

例 1 假设 $T' = \{T_1, T_2, T_3, T_4\}$ 为符合全路径集合 $\{p_1, \dots, p_5\}$ 的不完备 XML 信息系统, 信息如图 1 所示。根据限制相容关系 T_1 与 T_2, T_3 与 T_4 都是可以区分的, 但事实上两者之间叶子节点没有任何取值明确不同, 根据现实情况判断这两者应该是不可区分的。

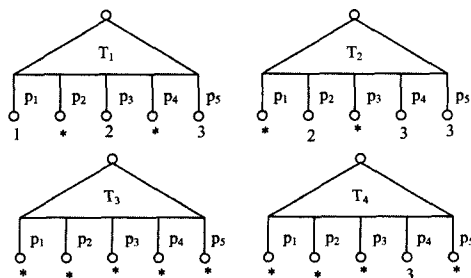


图 1 不完备的 XML 信息系统 T'

根据以上分析, 为弥补以上关系中分类过于粗糙和不合理的缺陷, 下面给出一种阈值相容关系。

2.3 阈值相容关系

定义 9(节点阈值相容) 设 $T = \{T_1, T_2, \dots, T_n\}$ 是一个不完备的 XML 信息系统, 任意子树 T_i 中的节点 u_i 与 T_j 中的节点 v_j 阈值相容(简记为 $u_i \doteq_a v_j$)需要满足如下条件:

(1) $f(u_i) = f(v_j)$;

(2) 若 $children(u_i) \neq \emptyset, children(v_j) \neq \emptyset$ 成立, 则 $\forall m_i \in children(u_i), \exists n_j \in children(v_j)$ 满足 $m_i \doteq_a n_j$; 反之, 若 $children(u_i) \neq \emptyset, children(v_j) \neq \emptyset$ 成立, 则 $\forall n_j \in children(v_j), \exists m_i \in children(u_i)$ 满足 $n_j \doteq_a m_i$;

(3) 若 $children(u_i) = \emptyset, children(v_j) = \emptyset$ 成立, 则满足

以下条件:

$(val(u_i) \neq * \wedge val(v_i) \neq * \wedge val(u_i) = val(v_i) \wedge \frac{|N_P(T_i) \cap N_P(T_j)|}{|P|} \geq \alpha) \vee val(u_i) = * \vee val(v_i) = *$, 其中, $0 \leq \alpha \leq 1, N_P(T_i) = \{p \in P \mid val_i(last(p)) \neq *\}$.

定义 10 (节点阈值相容关系) 设 $T = \{T_1, T_2, \dots, T_n\}$ 是一个不完备的 XML 信息系统, 由节点集 $V \subseteq N$ 决定的阈值相容关系为:

$Y_V^\alpha = \{(u_i, v_j) \mid \forall u_i \in V_i, \exists v_j \in V_j, u_i \doteq_\alpha v_j, \text{其中 } i \in [1, n], j \in [1, n]\}$.

$I_V^{\alpha}(u_i) = \{v_j \mid v_j \in T \wedge Y_V^\alpha(u_i, v_j), i \in [1, n], j \in [1, n]\}$.

由定义可知, 节点阈值相容关系弥补了节点限制相容关系分类不合理的缺陷. 同时, 通过引进阈值 $\alpha (0 \leq \alpha \leq 1)$, 可以很好地控制已知且相同路径信息在所有路径中所占比例, 进一步提高分类的准确度.

很显然, 阈值相容关系是自反的、对称的, 但不是传递的. 当 $\alpha = 0$ 时, 阈值相容关系就退化为相容关系.

定义 11 在不完备 XML 信息系统 $T = \{T_1, T_2, \dots, T_n\}$ 中, 子树集合 T 关于节点集 $V \subseteq N$ 的上、下近似和近似精度分别为:

$T_{Y_V^\alpha}^V = \{u_i \mid I_V^{\alpha}(u_i) \cap T \neq \emptyset\}, T_{Y_V^\alpha}^{\vee} = \{u_i \mid I_V^{\alpha}(u_i) \subseteq T\},$
 $\rho_{R(\alpha)}^V = \frac{|T_{Y_V^\alpha}^V|}{|T_{Y_V^\alpha}^{\vee}|}$

2.4 阈值相容关系的性质

定理 1 设 $T = \{T_1, T_2, \dots, T_n\}$ 为一个不完备 XML 信息系统, 由 $V \subseteq N$ 决定的阈值相容关系为 $Y_V^\alpha(u_i, v_j)$, 对于任意的 $T', T'' \subseteq T$ 有:

- (1) $T_{Y_V^\alpha}^V \subseteq T' \subseteq T_{Y_V^\alpha}^{\vee}$.
- (2) 若 $T' \subseteq T''$, 则 $T_{Y_V^\alpha}^V \subseteq T''_{Y_V^\alpha}$.
- (3) 若 $T' \subseteq T''$, $T_{Y_V^\alpha}^{\vee} \subseteq T''_{Y_V^\alpha}$.

证明: 直接验证即可.

定理 2 设 $T = \{T_1, T_2, \dots, T_n\}$ 为一个不完备的 XML 信息系统, 由 $V \subseteq N$ 决定的阈值相容关系为 $Y_V^\alpha(u_i, v_j)$, 若 $0 \leq \alpha \leq \beta \leq 1$, 则 $I_V^{\beta}(u_i) \subseteq I_V^{\alpha}(u_i)$, 其中 $i, j \in [1, n]$.

证明: 由定义验证即可.

定理 3 设 $T = \{T_1, T_2, \dots, T_n\}$ 为一个不完备的 XML 信息系统, 由 $V \subseteq N$ 决定的阈值相容关系为 $Y_V^\alpha(u_i, v_j)$, 若 $0 \leq \alpha \leq \beta \leq 1$, 则下列关系成立:

- (1) $T_{Y_V^\beta}^V \subseteq T_{Y_V^\alpha}^V$;
- (2) $T_{Y_V^\beta}^{\vee} \subseteq T_{Y_V^\alpha}^{\vee}$.

证明: (1) 对于任意 $n \in T_{Y_V^\beta}^V$, 则有 $I_V^{\beta}(n) \cap n \neq \emptyset$, 由定理 1 可得 $I_V^{\alpha}(n) \cap n \neq \emptyset$, 所以 $n \in T_{Y_V^\alpha}^V$, 从而结论成立.

(2) 对于任意 $n \in T_{Y_V^\alpha}^{\vee}$, 有 $I_V^{\alpha}(n) \subseteq V$, 由定理 1 可得 $I_V^{\beta}(n) \subseteq I_V^{\alpha}(n) \subseteq V$, 所以 $n \in T_{Y_V^\beta}^{\vee}$, 从而结论成立.

定理 4 设 $T = \{T_1, T_2, \dots, T_n\}$ 为一个不完备的 XML 信息系统, 对于任意节点集合 $V \subseteq N$ 及子树集合 $\bar{T} \subseteq T, \alpha \in [0, 1]$, 则下列关系成立:

- (1) $\bar{T}_V \subseteq \bar{T}_V^\alpha$;
- (2) $\bar{T}_{Y_V^\alpha}^{\vee} \subseteq \bar{T}_V^\alpha$.

证明: 略.

定理 5 设 $T = \{T_1, T_2, \dots, T_n\}$ 为一个不完备的 XML

信息系统, 对于任意节点集合 $V \subseteq N$ 及子树集合 $\bar{T} \subseteq T$, 若 $\alpha = 0$, 则下列关系成立:

- (1) $T_P^{\alpha} \subseteq T_P^0$;
- (2) $T_R^{\alpha} \subseteq T_R^0$.

证明: 略.

3 实例分析

例如, 不完备的 XML 信息系统 $T = \{T_1, T_2, \dots, T_{12}\}$ 的信息如图 2 所示, T 所对应的全路径集合为 $P = \{p_1, p_2, p_3, p_4, p_5\}$, 其中 p_1, p_2, p_3, p_4 为条件路径, p_5 为决策路径, $dom(last(p_i)) = \{0, 1, 2, 3\}, i \in [1, 4]$. “*” 表示未知值. $T/ind(p_5) = \{\psi, \phi\}$, 其中 $\phi = \{T_1, T_2, T_4, T_7, T_{10}, T_{12}\}, \psi = \{T_3, T_5, T_6, T_8, T_9, T_{11}\}$.

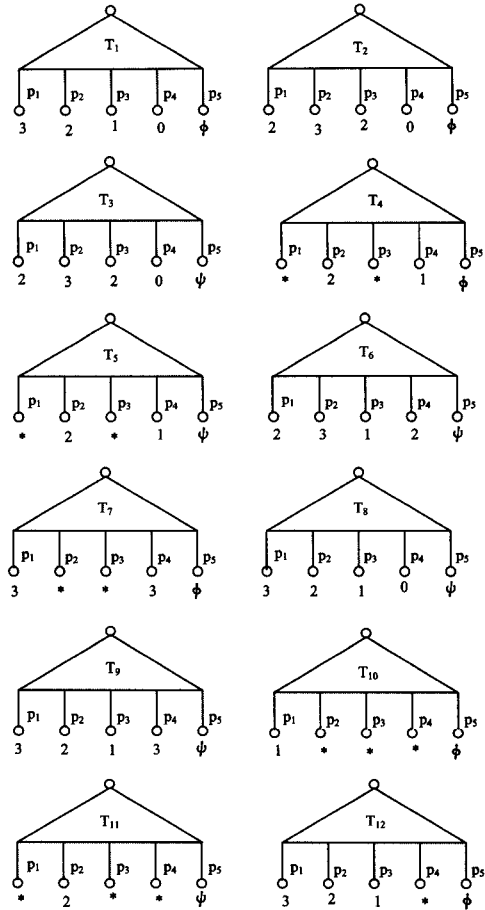


图 2 不完备的 XML 信息系统 T

(1) 根据相容关系分析图 2 得:

$\phi_P^C = \emptyset, \psi_P^C = \{T_6\}, \phi_C^P = \{T_1, T_2, T_3, T_4, T_5, T_7, T_8, T_9, T_{10}, T_{11}, T_{12}\}, \psi_C^P = T, \rho_C^P(\phi) = 0, \rho_C^P(\psi) = \frac{1}{12}$.

(2) 根据限制相容分析图 2 得:

$\phi_R^R = \{T_{10}\}, \psi_R^R = \{T_6, T_8\}, \phi_R^R = \{T_1, T_2, T_3, T_4, T_5, T_7, T_9, T_{10}, T_{11}, T_{12}\}, \psi_R^R = \{T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_9, T_{11}, T_{12}\}, \rho_R^R(\phi) = \frac{1}{10}, \rho_R^R(\psi) = \frac{2}{9}$.

容易验证定理 5 结论成立.

(3) 根据阈值相容分析图 2 得 (当 $\alpha = 0$):

$I_P^{(0)}(T_1) = \{T_1, T_{11}, T_{12}\}, I_P^{(0)}(T_2) = \{T_2, T_3\}, I_P^{(0)}(T_3) = \{T_2, T_3\}, I_P^{(0)}(T_4) = \{T_4, T_5, T_{10}, T_{11}, T_{12}\}, I_P^{(0)}$

$$(T_5) = \{T_4, T_5, T_{10}, T_{11}, T_{12}\}, I_P^{Y(0)}(T_6) = \{T_6\}, I_P^{Y(0)}(T_7) = \{T_7, T_8, T_9, T_{11}, T_{12}\}, I_P^{Y(0)}(T_8) = \{T_7, T_8, T_{10}\}, I_P^{Y(0)}(T_9) = \{T_7, T_9, T_{11}, T_{12}\}, I_P^{Y(0)}(T_{10}) = \{T_4, T_5, T_8, T_{10}, T_{11}\}, I_P^{Y(0)}(T_{11}) = \{T_1, T_4, T_5, T_7, T_9, T_{10}, T_{11}, T_{12}\}, I_P^{Y(0)}(T_{12}) = \{T_1, T_4, T_5, T_7, T_9, T_{11}, T_{12}\}$$

$$\phi_P^{Y(0)} = \emptyset, \psi_P^{Y(0)} = \{T_6\}, \phi_P^{Y(0)} = \{T_1, T_2, T_3, T_4, T_5, T_7, T_8, T_9, T_{10}, T_{11}, T_{12}\}, \rho_{Y(0)}^P(\phi) = 0, \rho_{Y(0)}^P(\psi) = \frac{1}{12}.$$

(4) 根据阈值相容分析图 2 得(当 $\alpha=0.7$):

$$I_P^{Y(0.7)}(T_1) = \{T_1, T_{12}\}, I_P^{Y(0.7)}(T_2) = \{T_2, T_3\}, I_P^{Y(0.7)}(T_3) = \{T_2, T_3\}, I_P^{Y(0.7)}(T_4) = \{T_4, T_{10}\}, I_P^{Y(0.7)}(T_5) = \{T_5, T_{10}\}, I_P^{Y(0.7)}(T_6) = \{T_6\}, I_P^{Y(0.7)}(T_7) = \{T_7, T_{11}\}, I_P^{Y(0.7)}(T_8) = \{T_7, T_8, T_{10}\}, I_P^{Y(0.7)}(T_9) = \{T_9, T_{12}\}, I_P^{Y(0.7)}(T_{10}) = \{T_4, T_5, T_8, T_{10}, T_{11}\}, I_P^{Y(0.7)}(T_{11}) = \{T_7, T_{10}, T_{11}\}, I_P^{Y(0.7)}(T_{12}) = \{T_7, T_9, T_{12}\}.$$

$$\phi_P^{Y(0.7)} = \{T_1, T_4\}, \psi_P^{Y(0.7)} = \{T_6\}, \phi_{Y(0.7)}^P = \{T_1, T_2, T_3, T_4, T_5, T_7, T_8, T_9, T_{10}, T_{11}, T_{12}\}, \psi_{Y(0.7)}^P = \{T_2, T_3, T_5, T_6, T_7, T_8, T_9, T_{10}, T_{11}, T_{12}\}, \rho_{Y(0.7)}^P(\phi) = \frac{2}{11}, \rho_{Y(0.7)}^P(\psi) = \frac{1}{10}.$$

容易验证上述定理 4 结论成立。上面的例子表明, 阈值相容关系克服了限制相容关系的缺点, 更加符合客观实际。比如个体 T_{10} 、 T_{11} 在限制相容关系下认为是可以分辨的。而在现实中, 两者在没有任何已知相同路径信息值的情况下, 显然是无法区分的。另外, 根据限制相容关系和相容关系, T_1 、 T_{11} 虽然仅有一个路径信息值相等但仍然归为一类, 这显然过于牵强。阈值相容关系通过合理地设置 α 来克服相容关系、限制相容关系的缺陷, 使分类更加合理。

结束语 目前, 基于粗糙集研究不完备 XML 信息系统的粗糙集模型, 在国内外还处于空白。本文借助粗糙集对确定的 XML 树信息系统进行了扩展, 允许叶子节点的信息值是丢失型或遗漏型空值。给出了不完备 XML 信息系统的相关定义, 借助相容关系、限制相容关系、阈值相容关系定义 3 种粗糙 XML 模型, 通过实例分析得到阈值相容关系粗糙 XML 模型能够弥补前两者的缺陷, 在分类方面具有很大优越性。本文的理论成果为不完备 XML 信息系统的路径约简及查询问题的进一步研究奠定了基础。

参考文献

[1] Pawlak Z. Rough sets[J]. International Journal of Computer and Information Science, 1982, 11(5): 341-356

[2] 苗守谦, 李道国. 粗糙集理论、算法与应用[M]. 北京: 清华大学出版社, 2008: 1-218

Miao Shou-qian, Li Dao-guo. Rough sets theory algorithms and applications[M]. Beijing: Tsinghua University Press, 2008: 1-218

[3] 安秋生. 粗糙关系数据库[M]. 北京: 电子工业出版社, 2009: 1-117

An Qiu-sheng. Rough relational database [M]. Beijing: Electronic Industry Press, 2009: 1-117

[4] Kryszkiewicz M. Rough set approach to incomplete information systems[J]. Information Sciences, 1998, 112(4): 39-49

[5] 王国胤. 粗糙集理论在不完备信息系统中的扩充[J]. 计算机研究与发展, 2002, 39(10): 1238-1243

Wang Guo-yin. Extension of rough set under incomplete information systems[J]. Journal of Computer Research and Develop-

ment, 2002, 39(10): 1238-1243

[6] Yang Xiao-ping. An improved model of rough sets on incomplete information systems [C] // The International Conference on Management of E-Commerce and E-Government, 2009: 193-196

[7] 刘后胜, 熊范纶, 李绍稳. 基于限制容差关系的不完备信息系统粗集模型拓展[J]. 生物数学学报, 2008, 23(3): 534-538

Liu Hou-sheng, Xiong Fan-lun, Li Shao-wen. The extension of rough set in incomplete information system based on limited tolerance relational model[J]. Journal of Biomathematics, 2008, 23(3): 534-538

[8] Hung E, Getoor L, Subrahmanian V S. Probabilistic interval XML[J]. ACM Transactions on Computational Logic, 2007, 8(4): 24

[9] Kimelfed B, Sagiv Y. Modeling and querying probabilistic XML data[C]//Special Interest Group on Management of Data Conference. 2008: 701-714

[10] Abiteboul S, Chan T-H, Huber, Kharlamov E. Aggregate queries for discrete and continuous probabilistic XML[C]//The 13th International Conference of Database Theory. Lausanne, Switzerland, 2010: 50-61

[11] 王建卫, 郝忠孝. 概率 XML 文件树结点概率的查询算法[J]. 计算机研究与发展, 2012, 49(4): 785-794

Wang Jian-wei, Hao Zhong-xiao. Node Probability Query Algorithm in Probabilistic XML Document Tree[J]. Journal of Computer Research and Development, 2012, 49(4): 785-794

[13] 殷丽凤, 郝忠孝. XML 强函数依赖的推理规则[J]. 计算机科学, 2008, 35(9): 165-167

Yin Li-feng, Hao Zhong-xiao. Inference Rules of XML Strong Functional Dependency[J]. Computer Science, 2008, 35(9): 165-167

[14] 殷丽凤, 郝忠孝. XML 强闭包依赖的研究[J]. 计算机科学, 2008, 35(11): 591-594

Yin Li-feng, Hao Zhong-xiao. Research on XML Strong Inclusion Dependency[J]. Computer Science, 2008, 35(11): 591-594

[15] 殷丽凤, 郝忠孝. 不完全信息环境下存在 XML 强多值依赖的 XML 文档的规范化研究[J]. 计算机研究与发展, 2009, 46(7): 1226-1233

Yin Li-feng, Hao Zhong-xiao. Normalization of XML Document with Strong MVD under Incomplete Informational Circumstances[J]. Journal of Computer Research and Development, 2009, 46(7): 1226-1233

[16] 殷丽凤, 郝忠孝. 存在 XML 强多值依赖的 XML Schema 规范化研究[J]. 计算机科学, 2010, 37(1): 192-196

Yin Li-feng, Hao Zhong-xiao. Research on Normalization of XML Schema with XML Strong MVD[J]. Computer Science, 2010, 37(1): 192-196

[17] 殷丽凤, 邱占芝. 判定粗糙 XML 函数依赖的一种算法[J]. 计算机科学, 2014, 41(5): 215-218

Yin Li-feng, Qiu Zhan-zhi. Determinant algorithm for rough XML functional dependency [J]. Computer Science, 2014, 41(5): 215-218

[18] 李雄飞, 孙涛, 郭建芳. 基于标签树的粗糙集模型 LTRS[J]. 通信学报, 2010, 31(6): 35-43

Li Xiong-fei, Sun Tao, Guo Jian-fang. Rough set model based on the labelled tree [J]. Journal on Communications, 2010, 31(6): 35-43