

不完备决策表中基于对象矩阵属性约简算法

王 炜 徐章艳 李晓瑜

(广西师范大学计算机科学与信息工程学院 桂林 541004)

摘 要 基于差别矩阵的属性约简是粗糙集属性约简中最常用的方法。对通常给出的以存储条件属性为基础差别矩阵进行比较后,给出一种对象矩阵的定义。对象矩阵从相容类内对象的决策值与条件属性的关系出发,存储的是对象集。给出对象矩阵的属性约简定义,证明了属性约简与基于正区域的属性约简的等价性。给出一个启发式的属性约简算法,其时间复杂度为 $\max(O(|C|^2|U_{pos}| |U|), O(|C| |U|^2))$, 空间复杂度为 $O(|C| |U|^2)$; 通过实例说明方法的可行性。

关键词 差别矩阵, 属性约简, 对象矩阵, 正区域, 约简算法

中图分类号 TP18 **文献标识码** A

Attribute Reduction Algorithm Based on Object Matrix in Incomplete Decision Table

WANG Wei XU Zhang-yan LI Xiao-yu

(College of Computer Science and Information Technology, Guangxi Normal University, Guilin 541004, China)

Abstract Attribute reduction based on discernibility matrix is the most common methods in rough set attribute reduction. Compared with discernibility matrix based on storage condition attribute, object matrix was defined in this paper. This object matrix uses the relationship between the decision value of the objects in tolerance class and condition attribute, and what it stores is object set. Then the definition of attribute reduction based on object matrix was given. It is proved that the attribute reduction acquired from this new method is equivalent to that based on positive region in incomplete information system. A heuristic algorithm for attribute reduction was presented which time complexity is $\max(O(|C|^2|U_{pos}| |U|), O(|C| |U|^2))$, and space complexity is $O(|C| |U|^2)$. An example was used to illustrate the feasibility of this new algorithm.

Keywords Discernibility matrix, Attribute reduction, Object matrix, Positive region, Reduction algorithm

1 引言

粗糙集理论^[1,2]是一种处理不精确、不完全与不确定性知识的数学工具。属性约简是其重要内容之一。目前,最常用的属性约简方法是基于差别矩阵^[3,4]的方法以及在此基础上的启发式算法。虽然差别矩阵方法存在占用存储空间大,并且对于大型数据集,其可辨别函数会过于复杂、难以化简等缺陷,但是这种方法因其简洁直观而得到很多研究学者的关注。

经典的粗糙集理论以等价关系为基础,面向的是完备的信息系统。然而在现实生活中,属性值不完备的情况广泛存在。对于这样的不完备信息系统, Kryszkiewicz 利用相容关系,给出了差别矩阵的属性约简方法^[5]。文献[6]提出了相容矩阵的概念,通过矩阵计算,给出了属性约简的启发式算法。文献[7-9]给出了不完备信息系统的几种差别矩阵的二进制形式,并进行相应的属性约简。

本文在分析了不完备决策表中相容类内对象的决策值与

属性约简中元素的关系基础上,给出一种对象矩阵及其对应的属性约简的定义,证明了此属性约简与基于正区域的属性约简是一致的,并给出一个启发式的属性约简算法。最后,通过实例说明本文方法的可行性。

2 相关概念

粗糙集理论的主要思想是把客观世界描述成一个四元组,通过一对上、下近似集进行知识的描述与刻画。

四元组 $S = (U, A, V, f)$ (一般简写为 $S = (U, A)$) 是一个信息系统,其中 $A = C \cup D$ 。这里 U 表示对象集合,一般称为论域; $A = C \cup D$ 表示属性的非空有限集, C, D 分别表示条件属性集和决策属性集,且 $C \cap D = \emptyset$; $V = \bigcup V_a$, V_a 表示属性 $a \in A$ 的值域; f 表示 $U \times A \rightarrow V$ 的一个信息函数,它为每个对象在每个属性上赋予一个信息值,即 $\forall a \in A, x \in U$ 有 $f(x, a) \in V_a$ 。如果 $D = \emptyset$,则称信息系统为信息表,否则称为决策表。若存在一个 $x \in U, a \in C$ 有 $f(x, a)$ 未知(记作 $f(x, a) = *$),则称信息系统为不完备的,否则称为完备的。

到稿日期:2011-05-09 返修日期:2011-07-16 本文受国家自然科学基金(60963008),广西自然科学基金(2011GXNSFA018163),广西研究生教育创新计划项目资助。

王 炜(1985—),女,硕士生,主要研究方向为粗糙集理论、数据挖掘,E-mail: wangweihdne@126.com;徐章艳(1972—),男,博士,教授,主要研究方向为粗糙集理论、数据挖掘;李晓瑜(1987—),男,硕士生,主要研究方向为数据挖掘。

定义 1 在不完备信息系统 $S=(U,A)$ 中, 设 $P \subseteq A$, 则由 P 决定的相容关系 T 为

$$T(P) = \{(x,y) \in U \times U \mid \forall a \in P, f(x,a) = f(y,a) \vee f(y,a) = * \vee f(y,a) = *\}$$

显然, 相容关系 T 满足自反性、对称性。令 $T_P(x) = \{y \mid y \in U, (x,y) \in T(P)\}$, 则 $T_P(x)$ 表示在属性集合 P 中与对象 x 满足相容关系 T 的个体对象集, 即对象 x 在属性集合 P 中的相容类。

定义 2 在不完备决策表 $S=(U,A)$ 中, 设 $P \subseteq C, Q \subseteq D$, 则 P 相对于 Q 的正区域定义为

$$POS_P(Q) = \bigcup_{y \in U/Q} \{x \mid x \in U, T_P(x) \subseteq y\}$$

如果取 $P=C, Q=D$, 则 $POS_C(D) = \bigcup \{x \mid x \in U, T_P(x) \subseteq D_i\}$ 。其中, $D_i \in U/D$ 表示决策属性集 D 对 U 的划分。一般简记为 $U_{pos} = POS_C(D), U_{neg} = U - POS_C(D)$ 。

定义 3 给定不完备决策表 $S=(U,A), \forall b \in B \subseteq C$, 若 $POS_B(D) = POS_{B-\{b\}}(D)$, 则称 b 为 B 中相对于 D 是不必要的, 否则称 b 为 B 中相对于 D 是必要的; 对 $B \subseteq C$, 如果 B 中任意属性相对于 D 都是必要的, 则称 B 相对于 D 是独立的。

定义 4 给定不完备决策表 $S=(U,A)$, 对 $B \subseteq C$, 如果 $POS_B(D) = POS_C(D)$ 且 B 相对于 D 是独立的, 则称 B 是 C 相对于 D 的属性约简。

定义 5 所有属性约简的交集称为决策表的核, 记为 $Core(C)$ 。

显然, 核中的属性相对于 D 都是必要的。

3 对象矩阵及其约简

对于不完备决策表的差别矩阵, 文献[5,6]给出的差别矩阵将相冲突的对象也进行考虑。事实上, 这些差别元素是没有必要的, 不仅浪费空间, 而且会降低获得属性约简的效率。

定义 6 在不完备决策表 $S=(U,A)$ 中, 定义对象矩阵 $M=(m_{ik})_{U_{pos} \times C}$, 其中 $m_{ik} = \{x_j \mid x_i \in U_{pos}, x_j \in U, (x_i, x_j) \in T(a_k) \wedge D(x_i) \neq D(x_j)\}$ 。

由定义可看出, m_{ik} 中的元素个数越少, 则说明在条件属性 a_k 下, 正区域中的对象 x_i 确定的相容类内对象的不一致性越小。

定义 7 对不完备决策表 $S=(U,A), B \subseteq C$, 定义对象 x_i 在条件属性集合 B 下的不一致类 $F_B(x_i) = \{x_j \mid (x_i, x_j) \in T(B), \text{且 } D(x_i) \neq D(x_j)\}$ 。

$F_B(x_i)$ 表示的是在集合 B 下与 x_i 满足相容关系且决策值不同的对象集合。显然, $F_B(x_i) \subseteq T_B(x_i)$ 。

性质 1 不完备决策表 $S=(U,A), M=(m_{ik})_{U_{pos} \times C}$ 是其对象矩阵, $B \subseteq C$, 则 $F_B(x_i) = \bigcap_{a_k \in B} m_{ik}$ 。

性质 2 不完备决策表 $S=(U,A), M=(m_{ik})_{U_{pos} \times C}$ 是其对象矩阵, $B \subseteq C, \forall a_k \in C-B$ 有 $F_{B \cup \{a_k\}}(x_i) = F_B(x_i) \cap m_{ik}$ 。

定义 8 不完备决策表 $S=(U,A)$ 中, $M=(m_{ik})_{U_{pos} \times C}$ 是其对象矩阵, 称 $DCore(C)$ 为对象矩阵的核, $DCore(C) = \{a_k \mid \exists x_j \notin m_{ik} \text{ 且 } x_j \text{ 均出现在第 } i \text{ 行的其他差别元素中}\}$ 。

定理 1 $\forall a_k \in Core(C)$, 则至少存在一个 $x_i \in U_{pos}$, 一个 $x_j \in U$ 有 $D(x_i) \neq D(x_j) \wedge f(x_i, a_k) \neq f(x_j, a_k)$ 。

证明: $\forall a_k \in Core(C)$, 根据定义 5 可知 $\forall B$ 是 C 相对于 D 的属性约简, 一定有 $a_k \in B$ 。由于 B 是属性约简, 则 $\forall a_k \in B$ 有 $POS_{B-\{a_k\}}(D) \neq POS_B(D)$ 并且 $POS_B(D) = POS_C(D)$ 。

因此至少存在一个 $x_i \in POS_C(D)$, 即 $x_i \in U_{pos}$, 且 $x_i \notin POS_{B-\{a_k\}}(D)$ 。所以必有 $x_j \in U$, 使得 $x_j \in T_{B-\{a_k\}}(x_i) \wedge D(x_i) \neq D(x_j)$ 。又因为 $x_i \in POS_B(D)$, 所以 $x_j \notin T_B(x_i)$ 。故有 $D(x_i) \neq D(x_j) \wedge f(x_i, a_k) \neq f(x_j, a_k)$ 。

定理 2 在不完备决策表 $S=(U,A)$ 中, $DCore(C) = Core(C)$ 。

证明: 根据定理 1 知, $\forall a_k \in Core(C)$, 则至少存在一个 $x_i \in U_{pos}$, 一个 $x_j \in U$, 有 $D(x_i) \neq D(x_j) \wedge f(x_i, a_k) \neq f(x_j, a_k)$ 。那么在对象矩阵 $M=(m_{ik})_{U_{pos} \times C}$ 中就有 $x_j \notin m_{ik}$ 且 $x_j \in m_{ip} (p \neq k)$ 。即 $a_k \in DCore(C)$, 故 $Core(C) \subseteq DCore(C)$ 。 $\forall a_k \in DCore(C)$, 根据定义 8, 有 $x_i \in U_{pos}, x_j \notin m_{ik}$ 且 $x_j \in m_{ip} (p \neq k)$, 也即是只有条件属性 a_k 有能力区分 x_i 与 x_j , 所以 a_k 存在于所有的属性约简中, 故 $a_k \in Core(C)$, $DCore(C) \subseteq Core(C)$ 。综上所述, 知命题成立。

定义 9 不完备决策表 $S=(U,A), B \subseteq C$, 对象矩阵 $M=(m_{ik})_{U_{pos} \times C}$, 若 B 满足 (1) $\forall x_i \in U_{pos}$, 有 $F_B(x_i) = \emptyset$; (2) $\forall b \in B$, 有 $B-\{b\}$ 均不满足 (1), 则称 B 是 C 关于 D 的基于对象矩阵的属性约简。

定理 3 不完备决策表 $S=(U,A), B \subseteq C$, 若 B 是基于对象矩阵的属性约简, 则 B 是基于正区域的属性约简。

证明: 假设 B 不是基于正区域的属性约简, 则有 $POS_B(D) \neq POS_C(D)$, 则至少存在 $x_i \in U_{pos}$ 使得 $T_C(x_i)$ 在决策属性 D 上是一致的, 且 $T_B(x_i)$ 在 D 上是不一致的。由 $T_B(x_i)$ 在 D 上是不一致的, 则一定存在 $x_j \in T_B(x_i)$ 使得 $D(x_i) \neq D(x_j)$, 根据定义 6、定义 7 和性质 1 可知 $x_j \in F_B(x_i)$, 即 $F_B(x_i) \neq \emptyset$, 那么与 B 是基于对象矩阵的属性约简矛盾。故假设不成立, 即命题成立。

定理 4 不完备决策表 $S=(U,A), B \subseteq C$, 若 B 是基于正区域的属性约简, 则 B 是基于对象矩阵的属性约简。

证明: 假设 B 不是基于对象矩阵的属性约简, 分两种情况: (1) 至少存在一个 $x_i \in U_{pos}$, 有 $F_B(x_i) \neq \emptyset$; (2) B 不满足基于对象矩阵属性约简定义的第 2 个条件。对于第一种情况, 一定存在 $x_j \in T_B(x_i)$ 使得 $D(x_i) \neq D(x_j)$, 从而使得 $x_i \notin POS_B(D)$ 。这与 $POS_B(D) = POS_C(D)$ 相矛盾, 故 B 不是基于正区域的属性约简。而对于第二种情况, 一定存在 $b \in B$, 使得 $F_{B-\{b\}}(x_i) = \emptyset$, 因而有 $POS_{B-\{b\}}(D) = POS_C(D)$, 故与 B 是基于正区域的属性约简相矛盾。所以, 命题成立。

4 对象矩阵的启发式属性约简算法

利用对象矩阵求不完备决策表的属性约简时, 首先需要得到不完备决策表对应的对象矩阵 $M=(m_{ik})_{U_{pos} \times C}$ 及核, 然后通过向核中添加条件属性, 直到找到属性约简为止。 $G(x_j)$ 表示对象 x_j 在一行中出现的次数, $P(x_j)$ 表示一行中包含 x_j 的元素所对应的条件属性。

算法 1 求不完备决策表的对象矩阵及其核

输入: 不完备决策表 $S=(U,A)$

输出: 决策矩阵 $M_{ik}=(m_{ik})_{U_{pos} \times C}$ 与核 $DCore(C)$

Step1 计算 $U_{pos} = \{x_1', x_2', \dots, x_s'\}$;

Step2 对 $\forall x_j \in U$ 初始化 $G(x_j)=0, P(x_j)=\emptyset$; 令 $DCore(C)=\emptyset$;

Step3 for($i=1; i \leq s; i++$) {

for($k=1; k \leq |C|; k++$) {

for($j=1; j \leq |U|; j++$) {

```

if((f(xi', ak)=f(xj, ak)) || f(xi', ak)= * || f(xj,
ak)= *) && D(xi') ≠ D(xj))
mik = mik ∪ {xi'};
G(xj) ++;
P(xj) = P(xj) ∪ {ak};
}
for(j=1; j ≤ |U|; j++) {
if(G(xj) = (|C| - 1))
DCore(C) = DCore(C) ∪ (C - P(xj));
for(j=1; j ≤ |U|; j++) {
G(xj) = 0;
P(xj) = ∅;
}
}

```

在算法 1 中, Step1 求得不完备决策表的正区域, 一般时间复杂度为 $O(|C||U|^2)$; Step3 是得到对象矩阵并得到核, 时间复杂度为 $O(|U_{pos}|(|C||U| + |U| + |U|))$ 。故算法 1 总的复杂度为 $O(|C||U|^2)$, 空间复杂度为 $O(|C||U|^2)$ 。

对于对象矩阵 $M=(m_{ik})_{U_{pos} \times C}$ 中的差别元素 m_{ik} , 其包含的元素越少, 所对应的条件属性对 x_i 的分类能力就越强。进一步 $m_{ik} = \emptyset$, 则说明条件属性 a_k 能将对象 x_i 与 U 中其他需要区分的对象都区分开。基于这种启发思想, 给出一个基于对象矩阵的不完备决策表属性约简算法。

算法 2 基于对象矩阵的启发式属性约简算法

输入: 不完备决策表 $S=(U, A), C=\{a_1, a_2, \dots, a_m\}$

输出: 属性约简 R

- Step1 利用算法 1 求不完备决策表对应的对象矩阵 $M=(m_{ik})_{U_{pos} \times C}$ 及其核 $DCore(C)$, 并初始化 $R=\emptyset, F_R(x_i)=U$;
- Step2 统计对象矩阵中每个元素的基, 记作 $num(m_{ik})$;
- Step3 $R=R \cup DCore(C)$; 如果 R 不为空, 对 $\forall x_i \in U_{pos}$ 计算 $F_R(x_i)$, 并判断其是否全为空; 若满足, 输出 R , 算法结束。否则删除 $F_R(x_i)$ 为空的 x_i 对应的行与核元素对应的列以及所删除元素对应的基, 从 U_{pos} 中删除该行所对应的对象;
- Step4 对对象矩阵中剩余的每行进行更新: $m_{ik} = m_{ik} \cap F_R(x_i)$, 并计算每一个此时 m_{ik} 的基 $N_num(m_{ik})$;
- Step5 矩阵中若有空元素, 则找到空元素对应的条件属性 a_k (若这样的属性有多个, 则选择基 $num(m_{ik})$ 值小的); 否则找出 $N_num(m_{ik})$ 值最小且 $F_R(x_i) \subsetneq m_{ik}$ 的元素对应的条件属性 a_k (若这样的属性有多个则任选一个);
- Step6 对 $\forall x_i \in U_{pos}$ 运用性质 2 计算 $F_{R \cup \{a_k\}}(x_i)$, 将 a_k 并入集合 R 。若有 $F_R(x_i) = \emptyset$, 则删除对象矩阵中 x_i 对应的行与 a_k 对应的列以及所删除元素对应的基; 从 U_{pos} 中删除该行所对应的对象。判断 U_{pos} 是否为空, 若空, 则算法结束; 否则, 转 Step4;

算法 2 中, Step1 要调用算法 1, 得到对象矩阵及核, 故时间复杂度为 $O(|C||U|^2)$, 空间复杂度为 $O(|C||U|^2)$; Step2 要搜索对象矩阵中的每个元素, 由于矩阵中元素的基最大为 $|U|$, 因此时间复杂度为 $O(|C||U_{pos}||U|)$; Step3 最坏的时间复杂度为 $O(|C||U_{pos}||U|)$; Step4 到 Step6 的时间复杂度为 $O(|C|^2|U_{pos}||U|)$ 。Step2 到 Step6 的空间复杂度为 $O(|C||U_{pos}||U|)$ 。

故算法 2 总时间复杂度为 $\max(O(|C|^2|U_{pos}||U|), O(|C||U|^2))$, 空间复杂度为 $O(|C||U|^2)$ 。

下面通过如表 1 所列的决策表 1, 对上述算法进行实例说明。

表 1 决策表 1

	a_1	a_2	a_3	a_4	a_5	d
x_1	0	*	1	0	0	1
x_2	1	1	0	1	2	0
x_3	0	3	0	1	0	0
x_4	1	*	1	*	1	1
x_5	0	*	1	*	1	1
x_6	1	2	0	1	2	1
x_7	1	1	*	0	*	0
x_8	0	0	*	*	1	0

首先求得正区域 $U_{pos} = \{x_1, x_2, x_3, x_6\} = \{x_1', x_2', x_3', x_4'\}$ 。建立的对象矩阵如表 2 所列。

表 2 对象矩阵

	a_1	a_2	a_3	a_4	a_5
x_1'	$\{x_3, x_8\}$	$\{x_2, x_3, x_7, x_8\}$	$\{x_7, x_8\}$	$\{x_7, x_8\}$	$\{x_3, x_7\}$
x_2'	$\{x_4, x_6\}$	$\{x_1, x_4, x_5\}$	$\{x_6\}$	$\{x_4, x_5, x_6\}$	$\{x_6\}$
x_3'	$\{x_1, x_5\}$	$\{x_1, x_4, x_5\}$	$\{x_6\}$	$\{x_4, x_5, x_6\}$	$\{x_1\}$
x_4'	$\{x_2, x_7\}$	$\emptyset$	$\{x_2, x_3, x_7, x_8\}$	$\{x_2, x_3, x_8\}$	$\{x_2, x_7\}$

第一行: $G(x_7)=4, P(x_7)=\{a_2, a_3, a_4, a_5\}, DCore(C)=\emptyset \cup (C - \{a_2, a_3, a_4, a_5\}) = \emptyset \cup \{a_1\} = \{a_1\}, G(x_8)=4, P(x_8)=\{a_1, a_2, a_3, a_4\}, DCore(C)=\{a_1\} \cup (C - \{a_1, a_2, a_3, a_4\}) = \{a_1\} \cup \{a_5\} = \{a_1, a_5\}$; 第二行: $G(x_6)=4, P(x_6)=\{a_1, a_3, a_4, a_5\}, DCore(C)=\{a_1, a_5\} \cup \{a_2\} = \{a_1, a_2, a_5\}$; 第三行: 无 $G(x_i)=4$; 第四行: $G(x_2)=4, P(x_2)=\{a_1, a_3, a_4, a_5\}, DCore(C)=\{a_1, a_2, a_5\} \cup \{a_2\} = \{a_1, a_2, a_5\}$ 。

令 $R=R \cup DCore(C)=\{a_1, a_2, a_5\}, F_R(x_1')=\{x_3, x_8\} \cap \{x_2, x_3, x_7, x_8\} \cap \{x_3, x_7\} = \{x_3\}, F_R(x_2')=\emptyset$, 删除第二行。 $F_R(x_3')=\{x_1\}, F_R(x_4')=\emptyset$, 删除第四行。此时矩阵只剩下第 1, 3 行, $U_{pos}=\{x_1, x_3\}=\{x_1', x_3'\}$ 更新后如表 3 所列。

表 3 更新后的矩阵

	a_1	a_2	a_3	a_4	a_5
x_1'			$\emptyset$	$\emptyset$	
x_3'			$\emptyset$	$\emptyset$	

此时矩阵中有多个空集, $num(m_{33})=1$ 最小, 选择 m_{33} 对应的条件属性 a_3 , 计算 $F_{R \cup \{a_3\}}(x_1')=\{x_3\} \cap \emptyset = \emptyset$, 删除第一行; $F_{R \cup \{a_3\}}(x_3')=\{x_1\} \cap \emptyset = \emptyset$, 删除第三行, $R=R \cup \{a_3\} = \{a_1, a_2, a_3, a_5\}$, 此时 U_{pos} 为空, 算法结束。 $R=\{a_1, a_2, a_3, a_5\}$ 为属性约简。

下面选取 UCI 数据库中的 5 个数据集进行实验, 数据集的具体信息见表 4。计算机硬件配置为 CPU Core2 E4500, 内存 1GB, 开发平台为 Visual Studio 2005, 并用 C# 编程语言实现本文的算法。

表 4 数据集信息

数据集	样本大小	条件属性	类别	完备
Hepatitis	155	19	2	否
Heart	270	13	2	否
Ionosphere	351	34	2	是
Credit	690	15	2	否
Soybean	307	35	19	否

求核与属性约简的结果如表 5 所列。其中核属性和属性约简对应的值是数据集中条件属性的序号, 如数据集 Hepatitis 的核属性是条件属性集中的第一个条件属性。运行时间是实验运行 10 次的平均时间。

表5 求核与属性约简结果

数据集	核属性	属性约简	运行时间(s)
Hepatitis	1	1,16,6,3,5	1.6991
Heart	—	8,5,4	4.8546
Ionosphere	—	18,24,4	8.1760
Credit	2	2,3,8,6	12.7890
Soybean	17,7,16,12	17,7,16,12,1,22, 6,15,8,35,4,9	11.8171

结束语 在不完备决策表的属性约简中,通常定义的差别矩阵是以存储条件属性集为基础的。而本文从不一致对象的角度,给出了一种对象矩阵及其对应的属性约简的定义。证明了该属性约简与基于正区域的属性约简是一致的,并给出一个启发式的属性约简算法。

相对于不完备信息系统中的相容关系,差异关系是从对象的个性出发。针对不完备决策表的属性约简,研究对象矩阵的约简与基于差异关系的约简的关系,是我们下一步的工作。

参考文献

[1] Pawlak Z. Rough Sets [J]. International Journal of Computer

and Information Science, 1982, 11(5): 341-356

- [2] Pawlak Z. Rough Sets [M]. London, UK: Kluwer Academic Publishers, 1991
- [3] Hu X H, Cerccone N. Learning in relational databases: a rough set approach[J]. Computational Intelligence, 1995, 11(2): 323-337
- [4] 叶东毅, 陈昭炯. 一个新的差别矩阵及其求核方法[J]. 电子学报, 2002, 30(7): 1086-1088
- [5] Kryszkiewicz M. Rough set approach to incomplete information systems[J]. Information Sciences, 1998, 112(1): 39-49
- [6] Huang Bing, He Xin, Zhou Xian-zhong. Rough Computational Methods Based on Tolerance Matrix [J]. Acta Automatica Sinica, 2004, 30(3): 364-370
- [7] 冯朝一, 梁家荣, 黄柳萍, 等. 基于集合覆盖的不完备信息系统属性约简方法[J]. 计算机应用, 2006, 26(11): 2664-2666
- [8] 赵亚鹏, 丁以中. 一种不完备信息系统的属性约简算法研究[J]. 计算机应用研究, 2008, 17(7): 46-48
- [9] 周海岩. 采用布尔矩阵不完备信息系统的属性约简[J]. 计算机工程与应用, 2010, 46(1): 119-121

(上接第176页)

足 $c \geq a, d \geq b$; p_1 是子实体子类化程度^[7], 即不相交子类化与重叠子类化; p_2 是子实体参与程度, 即全参与和部分参与, min, max 是整型常量。

$$\mathcal{C}[\text{Entity Type } E_1 \text{ Dynamic Extension } E_2 \text{ with } T_i] = \Lambda_{i=1,2} \text{inSch}(E_i, \text{SCH}) \wedge \exists ro_1, ro_2 \in \text{Role}(\mathcal{T}[T_i + 1] \rightarrow ro_2; E_2 \neq \mathcal{T}[T_i + 1] \rightarrow ro_1; E_1) \wedge (\mathcal{T}[T_i + 1] \rightarrow ro_2; E_2 \in \mathcal{T}[T_i] \rightarrow ro_1; E_1) \quad (34)$$

$$\mathcal{C}[\text{Entity Type } E_1 \text{ Dynamic Evolution } E_2 \text{ with } T_i] = \Lambda_{i=1,2} \text{inSch}(E_i, \text{SCH}) \wedge \exists ro_1, ro_2 \in \text{Role}((\mathcal{T}[T_i + 1] \rightarrow ro_2; E_2 \neq \mathcal{T}[T_i + 1] \rightarrow ro_1; E_1) \wedge (\mathcal{T}[T_i + 1] \rightarrow ro_2; E_2 \notin \mathcal{T}_i[T_i + 1] \rightarrow ro_1; E_1)) \quad (35)$$

用户自定义变迁规则^[10]约束源实体到目标实体的演变, 根据演变前、后实体角色是否相容可分为动态扩展和动态进化两类, 前者角色相容, 如图1所示的职工晋升为经理后其角色仍属于职工, 用带箭头的虚线表示并标以 DEX; 后者角色不相容, 从部门经理晋升为高级主管, 其角色不再相容, 在图1中标以 DEV。动态进化不能作用于父实体与子实体, 两个不相交实体间的动态扩展即为动态进化。

结束语 基于形式语言理论和形式语义学的指称语义方法, 同时参考 TimeER 模型的文本表示法, 建立一种形式化时态数据模型 $\mathcal{TDM}$ 及其形式语言模型 $\mathcal{L}(\mathcal{TDM})$ 。在 $\mathcal{L}(\mathcal{TDM})$ 框架内综合分析时态完整性约束常见的 7 种类型: 实体完整性约束、弱实体完整性约束、参照完整性约束、固有完整性约束、用户定义完整性约束、参与完整性约束和变迁约束, 并定义了各类时态完整性约束的形式语义规则, 深入分析了时态数据模型内在的时态语义联系, 为时态数据模型的研究提供了一个便利、高效的形式化理论框架。 $\mathcal{L}(\mathcal{TDM})$ 模型各语言成份的独立性与极小性, 由 $\mathcal{L}(\mathcal{TDM})$ 构成复杂形式系统的完备性与协调性等元理论性质的分析与论证是下一步研究工作的主要

内容, 局部性约束和聚集约束等特殊时态完整性约束类型的语义分析也是下一步工作的研究内容。

参考文献

- [1] 郝忠孝. 时态数据库设计理论[M]. 北京: 科学出版社, 2009: 5-38
- [2] 汤庸, 叶小平, 汤娜. 时态信息处理技术及应用[M]. 北京: 清华大学出版社, 2010: 16-85
- [3] Gregersen H, Jensen C S. Temporal entity-relationship models—a survey[J]. IEEE Transactions on Knowledge and Data Engineering, 1999, 11(3): 464-497
- [4] 汤庸, 汤娜, 毛承洁, 等. 时态知识/数据模型研究及应用[J]. 中山大学学报: 自然科学版, 2004, 43(6): 62-66
- [5] Gregersen H. The formal semantics of the TimeER model[C]// Proceedings of the Third Asia-Pacific Conference on Conceptual Modelling. Hobart, Australia: Australia Computer Society, 2006: 35-44
- [6] Gregersen H, Jensen C S. Conceptual modeling of time-varying information[R]. TR-35. Aalborg, Denmark: TimeCenter Publication, 1998
- [7] Combi C, Degani S, Jensen C S. Capturing temporal constraints in temporal ER models[J]. Lecture Notes in Computer Science, 2008, 5231: 397-411
- [8] McBrien P. Temporal constraints in non-temporal data modeling languages[J]. Lecture Notes in Computer Science, 2008, 5231: 412-425
- [9] Hoang Q, Nguyen T V. Extraction of TimeER model from a relational database[J]. Lecture Notes in Computer Science, 2011, 6591: 57-66
- [10] Artale A, Franconi E. Foundations of temporal conceptual data models[J]. Lecture Notes in Computer Science, 2009, 5600: 10-35