

带继承依赖的关系数据模型*

刘惟一

陆原

(云南大学计算机科学系 昆明 650091)

(云南大学物理系 昆明 650091)

摘要 面向对象技术为关系数据库设计提供了一些有用的方法. 本文给出了继承约束的概念, 并证明了继承依赖的推理规则是正确、完备的. 本文还得出继承依赖闭包、继承依赖的极小集合等一系列平行于函数依赖的结论.

关键词 继承, 关系数据库, 关系模式, 面向对象, 依赖.

面向对象技术受到普遍的关注^[1~3], 面向对象数据库也有许多成功的应用^[1,2,4,5], 但面向对象数据库还没有被广大关系数据库用户所接受. 将面向对象方法用于关系数据库设计是对关系数据库功能的扩充, 同时也丰富了面向对象的方法, 许多学者在这方面也做了不少工作.^[1,6]

类和类层次是面向对象方法中的重要概念, 很多文章涉及这方面的研究.^[3]和现有文章相比, 我们将类和子类间的关系看成是一种模式间的约束, 这就便于在关系数据库管理系统下实现.

首先在关系模型下, 我们定义了类模式和子类模式. 然后提出反映模式间继承关系的继承依赖概念, 给出了一套正确、完备的继承依赖推理规则. 利用这些规则, 我们得到继承依赖闭包、继承依赖极小集等一系列相近于函数依赖的结论.

1 基本概念

从面向对象的观点看: 具有相同特性和行为的对象构成类, 所有 1S-A 关系集构成类层次. 子类继承其超类的特性和行为, 并有自身的特性和行为.

本文为简便起见, 不考虑对象上的运算和属性间的函数依赖关系.

类模式 R 被看作是三元组 $\langle D, U \rangle$, 其中 U 是属性集, $D = \langle d_1, d_2, \dots, d_n \rangle$ 为标志属性序列. 每个标志属性由属性名 d_i 和属性值域 $dom(d_i)$ 描述, $dom(d_i)$ 可以是 0 或正整数集.

设 $R_1 = \langle D_1, U_1 \rangle, R_2 = \langle D_2, U_2 \rangle$ 是 2 个类模式. 设 r_1, r_2 分别为 R_1, R_2 上的 2 个关系实例. 设 $j_i = j_1, j_2, \dots, j_n, i'_r = i_1, i_2, \dots, i_n$ 为元组 $t \in r_1, t' \in r_2$ 的标志. 若对每个 $i_k (i_k \neq 0)$, 总对应于一个 j_k , 使 $i_k = j_k$, 就说 $t[D_1] \geq t'[D_2]$; 若存在 $i_i \neq j_i (i_i \neq 0, j_i \neq 0)$, 就说 $t[D_1] \neq t'[D_2]$.

* 作者刘惟一, 1950年生, 教授, 主要研究领域为数据库. 陆原, 女, 1956年生, 工程师, 主要研究领域为计算机应用.
本文通讯联系人: 刘惟一, 昆明 650091, 云南大学计算机科学系
本文 1995-08-14 收到修改稿

例 1: 设 $j_i = (j_1, j_2, \dots, j_5) = (01, 02, 03, 00, 00)$, $i_i = (i_1, i_2, \dots, i_5) = (01, 02, 00, 00, 00)$, 我们知道 $t[D_1] \geq t'[D_2]$. 若 $j_i = (03, 04, 01, 00, 00)$, 则 $t'[D_2] \neq t[D_1]$.

定义 1. 设 $R_1 = \langle D_1, U_1 \rangle, R_2 = \langle D_2, U_2 \rangle$ 是 2 个类模式, r_1, r_2 分别为 R_1, R_2 上的关系实例. 若对一切 $t' \in r_2$, 总存在 $t \in r_1$, 使得 $t'[D_2] \geq t[D_1]$ 且 $t[U_1 \cap U_2] = t'[U_1 \cap U_2]$, 就说 r_1, r_2 满足继承依赖 $R_1 \Rightarrow R_2$.

这里我们给出一个属性值约束: 若 $t[D_1] \leq t'[D_2]$, 则 $t[X] = t'[X]$, 其中 $X = U_1 \cap U_2$.

例 2: 设 r_1, r_2 是模式 R_1 和 R_2 上的 2 个实例, 如下:

r_1	d_1	d_2	d_3	B	C	D		r_2	d_1	d_2	d_3	A	B	C
	01	01	00	b	c	d			01	01	01	a	b	c
	02	02	00	b	b	c			02	02	01	b	b	b
	02	03	00	a	b	d			02	03	01	d	a	b
									02	03	02	c	a	b

由定义 1 可知 r_1, r_2 满足 $R_1 \Rightarrow R_2$.

这里我们仅考虑单继承. 关系实例对继承依赖的满足性是容易检验的.

根据语义模型^[1,7,8], 我们可以给出一个类模式集 O 和 O 上的继承依赖集 H . 为了保持类模式的不变性, 我们要求 O 中的类模式是不可分的.

为了讨论继承依赖关系, 我们定义以下类模式上的操作.

C-联接 (即类模式与其子模式的联接):

设 $R_1 = \langle D_1, U_1 \rangle, R_2 = \langle D_2, U_2 \rangle$ 是 O 中 2 个类模式且 $R_1 \Rightarrow R_2 \in H$. 设 r_1, r_2 为满足 $R_1 \Rightarrow R_2$ 的 2 个实例. r_1, r_2 间的 C 联接记为 $r_1 * * r_2$.

$$r = r_1 * * r_2 = \{t \mid u \in r_1, v \in r_2, (t[U_1 \cap U_2] = u[U_1 \cap U_2] = v[U_1 \cap U_2]) \wedge t[U_1 - U_2] = u[U_1 - U_2] \wedge t[U_2 - U_1] = v[U_2 - U_1] \wedge t[D] = v[D_2]\}$$

这里我们记 $R = \langle D_2, U_1 \cup U_2 \rangle = R_1 * * R_2$, $r_1 * * r_2$ 即为模式 $R_1 * * R_2$ 上的实例.

C-投影 (即类模式投影):

C-投影是 C-联接的逆运算. 设 R_1, R_2, \dots, R_n 是 O 中的类模式, $R = R_1 * * R_2 * * \dots * * R_n$. r 上的 C-投影只是在 R_i 或 $R_i * * R_j (1 \leq i, j \leq n)$ 上进行投影, 记为 $\prod_{R_i}(r)$ 或 $\prod_{R_i, R_j}(r)$, 这里 r 为 R 上的实例.

我们称 R_i 或 $R_i * * R_j (1 \leq i, j \leq n)$ 为 $R = R_1 * * \dots * * R_n$ 的 C-投影模式.

$$r' = \prod_{R_i}(r) = \{t' \mid t \in r, (t'[U_i] = t[U_i] \wedge t'[D'] = t[D_i])\}$$

记 $D' = \langle d'_1, \dots, d'_n \rangle, D = \langle d_1, \dots, d_n \rangle, D_i = \langle d_1, d_2, \dots, d_n \rangle, t'[D'] = t[D_i]$ 意味着

$$t'[d'_k] = \begin{cases} t[d_k] & \text{dom}(d'_k) \text{ 为整数集} \\ 0 & \text{dom}(d'_k) \text{ 为 } 0 \end{cases}$$

例 3: 设 r_1, r_2 是满足 $R_1 \Rightarrow R_2$ 的 2 个实例, r 为 r_1 和 r_2 的 C-联接, r' 是 r 在 R_1 上的 C-投影, 如下:

r_1	d_1	d_2	d_3	A	B		r_2	d_1	d_2	d_3	B	C
	01	02	00	a	b			01	02	01	b	c
	02	02	00	b	c			02	02	02	c	c
	02	03	00	a	c			02	02	03	c	a

$r=r_1 ** r_2$	d_1	d_2	d_3	A	B	C	$r' = \prod_{R_1}(r)$	d_1	d_2	d_3	A	B
	01	02	01	a	b	c		01	02	00	a	b
	02	02	02	b	c	c		02	02	00	b	c
	02	02	03	b	c	a						

设 O 是一个类模式集合. O 上的关系实例集记为 \mathcal{D} . 由 \mathcal{D} 经过 C -操作产生的实例集称为 \mathcal{D} 的扩张, 记为 \mathcal{D}^E .

例 4: 设 $O = \{R_1, R_2\}$, $R_1 = \langle D_1, U_1 \rangle$, $R_2 = \langle D_2, U_2 \rangle$, $U_1 = AB, U_2 = BC$. 表 2 中 r_1, r_2 是 \mathcal{D} 中的 2 个实例. 由 \mathcal{D}^E 定义可知, r, r' 是 \mathcal{D}^E 中的实例.

2 继承依赖公理

定义 2. 设 H 是 O 上的继承依赖集, $S \Rightarrow T$ 是一个继承依赖, $S, T \in O$. 若 \mathcal{D}^E 中任一个满足 H 的实例集也满足 $S \Rightarrow T$, 就称 G 逻辑蕴含 $S \Rightarrow T$.

关于继承依赖我们有下面的推理规则:

L0 设 R 为 O 中的模式, r 为 R 上的实例, 则 r, r 满足 $R \Rightarrow R$.

L1 设 R_1, R_2 是 O 中的 2 个类模式. 设 r_1, r_2 为满足 $R_1 \Rightarrow R_2$ 的 2 个实例. 记 $r = r_1 ** r_2$. r_1, r 和 r_2, r 分别满足 $R_1 \Rightarrow R, R_2 \Rightarrow R$, 其中 $R = R_1 ** R_2 = \langle D_2, U_1 \cup U_2 \rangle$.

L2 设 R_1, R_2, R_3 是 O 中的类模式, r_1, r_2, r_3 分别为 R_1, R_2 和 R_3 上的实例. 若 r_1, r_2 满足 $R_1 \Rightarrow R_2, r_2, r_3$ 满足 $R_2 \Rightarrow R_3$, 则 r_1, r_3 满足 $R_1 \Rightarrow R_3$.

L3 设 R_1, R_2, R_3 为 O 中的类模式, r_1, r_2, r_3 分别为 R_1, R_2, R_3 上的实例. 若 r_1, r_2 满足 $R_1 \Rightarrow R_2, r_2, r_3$ 满足 $R_2 \Rightarrow R_3$, 则 $r_1 ** r_2, r_3$ 满足 $R_1 ** R_2 \Rightarrow R_3$.

定理 1. 规则 L0~L3 是正确的.

证明: L0: 显然.

L1: 由 C -联接定义直接可得结论.

L2: 因为 r_1, r_2 满足 $R_1 \Rightarrow R_2$, 可知 $(\forall t' \in r_2)(\exists t \in r_1)(t'[D_2] \geq t[D_1])$. 由于 r_2, r_3 满足 $R_2 \Rightarrow R_3$, 可知 $(\forall t'' \in r_3)(\exists t' \in r_2)(t''[D_3] \geq t'[D_2])$. 所以 $(\forall t'' \in r_3)(\exists t \in r_1)(t''[D_3] \geq t[D_1])$. r_1, r_3 满足 $R_1 \Rightarrow R_3$.

L3: 因为 r_2, r_3 满足 $R_2 \Rightarrow R_3$, 可知 $(\forall t'' \in r_3)(\exists t'' \in r_2)(t''[D_3] \geq t''[D_2])$. 由于 $R_1 \Rightarrow R_2$, 由 C -联接定义可知 $(\forall t'' \in r_2)(\exists t \in r)(t[D] = t''[D_2])$, 其中 $r = r_1 ** r_2$. 所以 $(\forall t'' \in r_3)(\exists t \in r)(t''[D] \geq t[D])$. r, r_3 满足 $R \Rightarrow R_3$, 这里 $R = R_1 ** R_2$.

定义 3. 设 O 是类模式集, H 为 O 上的继承依赖集. H 的闭包, 记为 H^+ , 是由 H 经 L0~L3 导出的所有继承依赖的集合.

定义 4. 设 O 是类模式集, H 是 O 上的继承依赖集. 类模式 A 关于 H 的闭包, 记为 $(A_H)^+$, 是这样的类模式 R_i 的集合, 使得 $R_i \Rightarrow A$ 能够由 H 经 L0~L3 导出. 设 $(A_H)^+ = \{R_1 \langle D_1, U_1 \rangle, \dots, R_n \langle D_n, U_n \rangle, A \langle D_A, U_A \rangle\}$, 记 $R = \langle D_A, U_1 \cup U_2 \cup \dots \cup U_n \cup U_A \rangle$ 为 $(A_H)^+$.

算法 1. 计算 $(A_H)^+$

输入: A 和 H

输出: $(A_H)^+$

$s_1: i := 0; A_{(i)} := A;$

$s_2:$ 求类模式 $B;$

$$B = \{X | (\exists v)(\exists w)(v \Rightarrow w \in H \wedge w \subseteq_p(A_{(i)}) \wedge X \in_p V)\}.$$

// $W \subseteq_p A_{(i)}$ 表示 W 为 $A_{(i)}$ 的 C -投影模式. 设 $V = R_1 * * \dots * * R_k$. $X \in_p V$ 表示 X 为 V 的某个 C -投影模式 $R_i (1 \leq i \leq k)$. //

$s_3: A_{(i+1)} := A_{(i)} * * B;$

$s_4:$ 若 $A_{(i+1)} \neq A_{(i)}$, 则 $i := i + 1$, 转到 $s_2;$

$s_5:$ 输出 $A_{(i)}$.

定义 5. 设 O 是类模式集, H 是 O 上的继承依赖集. 我们称模式 $G \in O$ 是关于 H 的叶模式, 如果 O 中不存在模式 M 使 $G \Rightarrow M \in H$.

例 5: 设 $O = \{A, B, C, D, E, F, G, T\}$, $H = \{A \Rightarrow B, A \Rightarrow C, C \Rightarrow D, C \Rightarrow E, C \Rightarrow G, B \Rightarrow F, E \Rightarrow T\}$. 如图 1 所示, D, T, F, G 是叶模式.

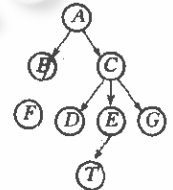


图 1

定理 2. 规则 $L0 \sim L3$ 是完备的.

设 O 是类模式集, H 是 O 上的继承依赖集. 存在一组 \mathcal{Q}^E 中的具体实例 INS , H 中所有依赖都为 INS 所满足, 而不能由 $L0 \sim L3$ 推出的某个 $E \Rightarrow F$ (即 $E \Rightarrow F \notin H^+$), 则不为 INS 所满足.

证明: 设 S_1, S_2, \dots, S_m 为 O 中关于 H 的叶模式. 我们给出一组类模式 $S_1^*, S_2^*, \dots, S_m^*$. 例如, 例 5 中 $D^* = A * * C * * D, T^* = A * * C * * E * * T, G^* = A * * C * * G, F^* = A * * B * * F$.

我们构造 S_i^* 上的实例 $r_i (i = 1, \dots, m)$ 如下:

r_1							
d_1	d_2	...	d_i	d_{i+1}	...	d_n	U_{r_1}
1	1	...	1	0	...	0	$a_1 a_1 \dots a_1$
r_2							
d_1	d_2	...	d_i	d_{i+1}	...	d_n	U_{r_2}
2	2	...	2	0	...	0	$a_2 a_2 \dots a_2$
.....							
r_m							
d_1	d_2	...	d_i	d_{i+1}	...	d_n	U_{r_m}
m	m	...	m	0	...	0	$a_m a_m \dots a_m$

设 $A \Rightarrow B$ 是 H 中任一继承依赖. 因为 $A \Rightarrow B \in H, B$ 一定是某个 S_i^* 的 C -投影模式 ($1 \leq i \leq m$). 由算法 1 可知 A 也是 S_i^* 的 C -投影模式. 由表 3 可知: $(\forall t' \in \prod_B(r_i)) (\exists t \in \prod_A(r_i)) (t[D_A] = t'[D_B] = i i \dots i 0 \dots 0 \wedge t[U_A \cap U_B] = t'[U_A \cap U_B] = a_i \dots a_i, A \Rightarrow B$ 被 INS 所满足.

设 $E \Rightarrow F \notin H^+$. 假定 E 是某个 S_j^* 的 C -投影模式, F 为 S_i^* 的 C -投影模式. 由于 $E \not\Rightarrow F, E$ 不是 S_i^* 的 C -投影模式, 所以不存在 $t \in \prod_E(r_j), t' \in \prod_F(r_i)$, 使 $t[D_E] \leq t'[D_F]$. $E \Rightarrow F$ 不为 INS 所满足.

3 继承依赖的极小覆盖

定义 6. 设 H 和 G 是继承依赖集. 若 $H^+ = G^+$, 称 H 和 G 等价.

定义 7. 设 H 是继承依赖集, 我们说 H 是极小的, 如果:

- (1) H 中每个依赖的左部是单模式;
- (2) 对 H 中任一依赖 $X \Rightarrow A$ 来说, $H - \{X \Rightarrow A\}$ 与 H 不等价;
- (3) 对 H 中任一依赖 $X \Rightarrow A$ 来说, $H - \{X \Rightarrow A\} \cup \{X \Rightarrow B\}$ 与 H 不等价, 其中 B 是 A 的 C -投影模式.

定理 3. 每个继承依赖 H 与它的极小集 H_{min} 等价.

证明. 我们分以下 3 步构造 H 的极小集.

- (1) 逐一查 H 中每个继承依赖 $X \Rightarrow Y$, 若 $X = A_1 * * A_2 * * \dots * * A_k$, 则用 $\{A_i \Rightarrow Y \mid i = 1 \dots k\}$ 取代 $X \Rightarrow Y$.
- (2) 逐一取 H 中每个继承依赖 $X \Rightarrow Y$, 令 $G = H - \{X \Rightarrow Y\}$, 若 $X \in (Y_G)^+$, 则从 H 中删去 $X \Rightarrow Y$.
- (3) 逐一取 H 中每个继承依赖 $X \Rightarrow Y$, 假定 $Y = B_1 * * \dots * * B_m$, 对每个 B_i 来说, 若 $X \in (B_i * * \dots * * B_{i-1} * * B_{i+1} * * \dots * * B_m)_H^+$, 则用 $X \Rightarrow (B_1 * * \dots * * B_{i-1} * * B_{i+1} * * \dots * * B_m)$ 代替 $X \Rightarrow Y$.

最后剩下的 H 就是极小继承依赖集. 因为我们对 H 施行的每步“改造”都保证了前后 2 个依赖集等价, 所以 H 和最后得到的 H_{min} 等价.

4 数据库模式设计举例

让我们来考察下面的汽车情况表.

表 1

车轮直径(mm)	引擎功率(h·p)	载重(吨)	座位数(个)	冷藏柜容积(m³)
900	99	5		
900	99	5		30
900	99	5		24
900	99		50	
650	80		50	
650	80	3		

表 1 中存在大量空值, 为此我们将其变为表 2.

表 2

卡车				
车轮直径(mm)	引擎功率(h·p)	载重(吨)		
900	99	5		
650	80	3		

冷藏车				
车轮直径(mm)	引擎功率(h·p)	载重(吨)	冷藏柜容积(m³)	
900	99	5	30	
900	99	5	24	

客车

车轮直径(mm)	引擎功率(h·p)	座位数(个)
900	99	50
650	80	30

表 2 存在大量冗余数据,我们运用继承技术来减少冗余.根据语义模型^[1,7,8],给出下面的类模式:

$O = \{ \text{卡车}(T), \text{客车}(B), \text{冷藏车}(C), \text{汽车}(A) \}, H = \{ A \Rightarrow B, A \Rightarrow T, T \Rightarrow C \}$.显然 H 是极小继承依赖集.图 2 描述了这个继承结构.由表 1 和图 2 我们得到图 3.

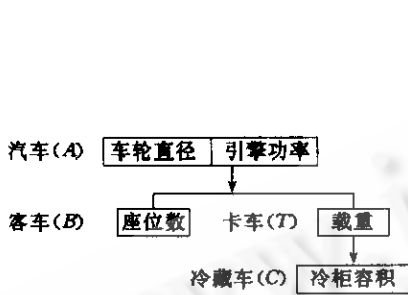


图2

A	d_1	d_2	d_3	d_4	车轮直径(mm)	引擎功率(p. h)
	1	1	0	0	900	99
	1	2	0	0	650	80

B	d_1	d_2	d_3	d_4	座位数(个)
	1	1	1	0	50
	1	2	2	0	30

T	d_1	d_2	d_3	d_4	载重(吨)
	1	1	3	0	5
	1	2	4	0	3

C	d_1	d_2	d_3	d_4	载重(吨)
	1	1	3	1	30
	1	1	3	2	24

图3

继承性约束为语义一致性提供了保证.例如,通过求 $(C_H)^*$,我们可以得到冷藏车的全部信息.在这个例子中,各个模式都是 3 范式.一般来说,因为标志属性是码,类模式至少是 2 范式.^[9,10]

5 结 论

本文给出了继承依赖的概念,在关系数据库管理系统下,这种约束很容易实现.将继承方法用于关系数据库设计可以节省大量存储空间.本文将属性间的函数依赖推广到模式间的继承依赖,发展了依赖理论.本文仅讨论了单继承,关于多重继承和关系模式的聚集语义描述将在以后的文章中给予讨论.

参 考 文 献

- 1 Davies P B. Entity models to object models; object-oriented analysis and database design. Information and Software Technology, 1992,34(4):255~262.
- 2 Khoshatian S. Insight into object-oriented databases. Information and Software Technology, 1990,32(4):274~289.
- 3 Sciore E. Object specialigation. ACM Trans. on Information Systems, 1989,7(2):103~112.
- 4 Fishman D H, Beech D, Cate H P et al. Iris, an object-oriented database management system. ACM Trans. on Office Information Systems, 1987,5(1):46~69.
- 5 Ngugen G T, Rieu D. Schema exclusion in object-oriented database systems. Data and Knowledge Engineering, 1989,4:43~67.
- 6 Blaha M R, Premerlani W J, Baugh J E Rum. Relational database design using on object-oriented methodology. communications of the ACM, 1988,31(4):414~427.
- 7 Hull R, King R. Semantic database modeling; survey, applications, and research issues. ACM Computing Sur-

veys, 1987, 19(3):201~260.

8 Peckham J, Marganski F. Semantic data models. *ACM Computing Surveys*, 1988, 20(3):153~189.

9 Ullman J D. *Principle of database systems*. Maryland: Computer Science Press, 1982.

10 Maier D. *The theory of relational database*. Maryland: Computer Science Press, 1983.

A RELATIONAL DATA MODEL WITH INHERITANCE DEPENDENCIES

Liu Weiyi

(*Department of Computer Science Yunnan University Kunming 650091*)

Lu Yuan

(*Department of Physics Yunnan University Kunming 650091*)

Abstract Object-oriented concepts provide some useful methods for designing a relational database. In this paper, the concept of inheritance constraints is given. The authors show that the inference rules for inheritance dependencies are sound and complete. Furthermore, a series of results about the closure of inheritance dependencies, the minimal set of inheritance dependence is obtained, which are analogous to functional dependencies.

Key words Inheritance, relational database, relational scheme, object-oriented, dependency.