第 6 卷第 4 期
1995 年 4 月

软 件 学 报
JOURNAL OF SOFTWARE

Vol. 6, No. 4
Apr. 1995

基于复杂对象的知识库语言*

施伯乐 周傲英 郭德培 叶道兵

(复旦大学计算机科学系,上海 200433)

摘要 近年来,复杂对象数据模型及其逻辑数据语言的研究引起了人们的广泛的重视. 本文提出了一种复杂对象模型 COM,描述了该模型上的说明性查询语言 CO-DATALOG 的语法结构. 在语义方面,定义了空间、基、解释、满足、模型等概念;利用 COM 实例的格性质证明了 CO-DATALOG 的模型相交定理、最小模型存在定理以及最小模型的不动点性质,从而重构了其语义理论框架.

关 键 词 复杂对象,逻辑数据语言,模型论语义.

近 10 年来,在 CAD、CASE 和 CIMS 等应用领域提出了对数据库技术的新要求. 传统的关系模型不能有效地满足这些要求.这促使人们研究更通用的数据模型,复杂对象数据模型就是其中一种. 迄今已建成的复杂对象数据库系统大多是以扩充的关系模型和可高效实现的扩充的关系代数为基础的[1,2].

随着复杂对象数据模型研究的开展,新的数据库查询语言的研究也自然地被提到议事日程上来. 早期研究复杂对象查询语言的两种主要方法是代数方法和演算方法. 但由于这两种方法得到的查询语言不便于使用、难以表达递归. 人们渐渐转向其它方法,其中最具吸引力的是基于逻辑数据语言(也称基于规则)的方法.

逻辑数据语言是知识库研究的核心内容之一. 以前的研究主要以平关系数据库为背景,提出的语言,如 DATALOG,提供了描述关系数据库表示和查询的较自然的方法. 这类语言是说明性的,有完善的数学基础,在系统实现和查询优化技术方面也取得了一定的进展.

复杂对象的基于规则的查询语言的研究是近几年才开展起来的. 文献[3]中提出的复杂对象演算体现了这方面研究的本质思想. 文献[4,5]反映了这方面研究的最新进展,但他们的研究都存在诸如不直观(不自然)、不易于系统实现等问题. 如果把嵌套关系模型理解为属性值可以是一个关系,即从谓词逻辑的观点来看,谓词可作为参量出现在谓词内部的参量位置上,那么,文献[4]中的模型都是符合这个理解的. 本文提出的模型中,属性值只是纯粹意义上的元组或集合,而不是关系,即对应的谓词演算语言中,谓词的参量可以是高阶项但不

* 本文 1993-05-21 收到,1993-09-06 定稿

本课题得到国家 863 高技术计划和国家自然科学基金的资助. 作者施伯乐,1936 年生,教授,主要研究领域为数据库,知识库,面向对象数据库. 周傲英,1965 年生,副教授,主要研究领域为数据库,知识库,面向对象数据库. 郭德培,1969 年生,博士生,主要研究领域为知识库,人工智能. 叶道兵,1969 年生,助教,主要研究领域为知识库,人工智能.

本文通讯联系人:施伯乐,上海 200433,复旦大学计算机科学系

能是谓词，不会产生高阶逻辑的问题，故而与嵌套关系模型有所区别.

本文第1节描述COM数据模型，并定义了COM实例之间的信息包含关系，证明了COM实例空间的格性质.第2节定义该模型上的说明性查询语言CO－DATALOG，描述了其语法结构，利用COM实例的性质证明了其模型相交定理、最小模型存在定理以及最小模型的不动点性质，从而重构了CO－DATALOG语义理论框架.

1 数据模型

1.1 COM数据模型

定义 1.1.1. 属性模式

设 $\mathscr{A}$ 是一非空属性名集合，$A\in\mathscr{A}$（A 称为属性名），则

①$A()$，$A^*()$是属性模式，A 称为基本属性；

②$A(L)$，$A^*(L)$是属性模式（其中 $L=l_1,\cdots,l_k$，$l_i(1\leqslant i\leqslant k)$是属性模式）.

若 $A(L)$是属性模式，则 A 称为单值属性；若 $A^*(L)$是属性模式，则 A 称为多值属性.

定义 1.1.2. 关系模式、数据库模式

设 $\mathscr{R}$ 是一非空关系名集合，$R\in\mathscr{R}$，则

$R(A_1(L_1),\cdots,A_m(L_m),B_1^*(L'_1),\cdots B_n^*(L'_n))$（其中 $A_i(L_i)(1\leqslant i\leqslant m)$，$B_j^*(L'_j)$，$(1\leqslant j\leqslant n)$为属性模式）是一关系模式，记为 $R(A_1,\cdots,A_m,B_1,\cdots,B_n)$.

数据库模式是由关系模式构成的一个元组 $D=(R_1,\cdots,R_k)$.

一个关系模式中的所有属性可划分为两类，$A_1,\cdots,A_m,B_1,\cdots,B_n$ 为关系属性，其它为嵌套属性.

定义 1.1.3. 属性实例

设 Ω 是一非空的原子值的集合，则

①若 A 为单值基本属性，$a\in\Omega$，则 $A:a$ 为 A 的属性实例；

②若 A 为多值基本属性，对所有 $i(1\leqslant i\leqslant n)$，$a_i\in\Omega$，则 $A:\{a_1,\cdots,a_n\}$为 A 的属性实例；

③若 A 的属性模式为 $A(l_1,\cdots,l_k)$，则

$A:[r_1,\cdots,r_k]$是 A 的属性实例如果 $\forall\ i(1\leqslant i\leqslant k)$ r_i 是 l_i 的属性实例.

④若 A 的属性模式为 $A^*(l_1,\cdots,l_k)$则

$A:\{[r_{11},\cdots,r_{1k}],\cdots,[r_{m1},\cdots,r_{mk}]\}$是 A 的属性实例如果 $\forall\ i(1\leqslant i\leqslant m)$，$\forall\ j(1\leqslant j\leqslant k)$，$r_{ij}$是 l_j 的属性实例.

定义 1.1.4. 关系实例、数据库实例

设有关系模式 $R(A_1,\cdots,A_n)$则

$r=\{[r_{11},\cdots,r_{1n}],\cdots,[r_{m1},\cdots,r_{mn}]\}$（其中 $m\geqslant 0$，$\forall\ i(1\leqslant i\leqslant m)$，$\forall\ j(1\leqslant j\leqslant n)$，$r_{ij}$是 A_j 的关系实例）是 R 的一个关系实例.

数据库模式 $D=(R_1,\cdots,R_k)$的实例是一个元组 $d=(r_1,\cdots,r_k)$，其中 $r_i(1\leqslant i\leqslant k)$，是关系模式 R_i 的实例.

1.2 COM实例及其性质

定义 1.2.1. COM实例

设 $A(L)$ 或 $A^*(L)$ 为一属性模式，则属性 A 的 COM 实例记为 inst(A)，递归定义如下：

①若 $L=\Phi$

(i)A 为单值基本属性，则 $\text{inst}(A)=\text{dom}(A)=\Omega$；

(ii)A 为多值基本属性，则 $\text{inst}(A)=\text{dom}(A)=2^{\Omega}$.

②若 $L=A_1(L_1),\cdots,A_m(L_m),B_1^*(L'_1),\cdots,B_n^*(L'_n)$

令 $I=\text{inst}(A_1)\times\cdots\times\text{inst}(A_m)\times\text{inst}(B_1)\times\cdots\times\text{inst}(B_n)$

(i)A 为单值属性，则 $\text{inst}(A)=I$；

(ii)A 为多值属性，则 $\text{inst}(A)\subset 2^I$，且对任一 $S\in\text{inst}(A)$，其中的任意两个元组

$[A_1,a_{i1},\cdots,A_m:a_{im},B_1:b_{i1},\cdots,B_n:b_{in}]$ 和 $[A_1:a_{j1},\cdots,A_m:a_{jm},B_1:b_{j1},\cdots,B_n:b_{jn}]$ $(i\neq j)$

满足 $\bigwedge_{q=1}^{m}(a_{iq}=a_{jq})\Rightarrow\bigwedge_{k=1}^{n}(b_{ik}=b_{jk})$.

类似地，我们可定义关系模式 R 及数据库模式 D 的 COM 实例 inst(R) 及 inst(D).

定义 1.2.2. 信息包含

设 r_1、r_2 是关系模式 $R=(A_1,\cdots,A_m,B_1^*,\cdots,B_n^*)$ 的 COM 实例中的两个元素. 则 r_1 信息包含在 r_2 中 $(r_1\preceq r_2)$，如果 $\forall\ t\in r_1$，$\exists\ t'\in r_2$，使得 $t[A_1,\cdots,A_m]=t'[A_1,\cdots,A_m]$ 且 $t[B_j]\preceq t'[B_j](1\leqslant j\leqslant n)$.

对数据库模式 $D=(R_1,\cdots,R_k)$，数据库实例 $d=(r_1,\cdots,r_k)$，$d'=(r'_1,\cdots,r'_k)$ 间的信息包含关系定义为：$d\preceq d'$ 当且仅当 $\forall\ i$，$r_i\preceq r'_i$.

定义 1.2.3. 嵌套深度

属性 A 的嵌套深度，记为 NH(A) 定义为：

①A 为单值属性，则 $\text{NH}(A)=0$；

②若 A 的属性模式为 $A^*(A_1,\cdots,A_m,B_1^*,\cdots B_n^*)$，则

$$\text{NH}(A)=\max\{\text{NH}(B_1),\cdots,\text{NH}(B_n)\}+1.$$

对一个 COM 关系而言，其嵌套深度定义为其关系属性的最大深度加一.

定理 1.2.1. 对一个关系 R，信息包含关系 $(\preceq)$ 是其 COM 实例 inst(R) 上的偏序关系.

证明：设 R 的关系模式是 $R=(A_1,\cdots,A_m,B_1^*,\cdots,B_n^*)$，关于 R 的嵌套深度 NH(R). 利用归纳法证明：

归纳基础：当 $\text{NH}(R)=1$ 时，R 中没有多值属性，$r_1\preceq r_2$ 即是 $r_1\subseteq r_2$，因此，$\preceq$ 是 inst(R) 上的偏序关系.

归纳假设：假设当 $\text{NH}(R)\leqslant n(n\geqslant 1)$ 时，$\preceq$ 是 inst(R) 上的偏序关系.

下面证明 $\text{NH}(R)=n+1$ 时，这一性质仍成立.

归纳步骤：

自反性：根据定义可知：$r\preceq r$.

反对称性：设 $r1,r2\in\text{inst}(R)$ 且 $r_1\preceq r_2$，$r_2\preceq r_1$，由定义：$\forall\ t\in r_1$，$\exists\ t'\in r_2$，又 $\exists\ t''\in r_1$ 使得 $t[A_1,\cdots,A_m]=t'[A_1,\cdots,A_m]=t''[A_1,\cdots,A_m]$，且 $t[B_j]\preceq t'[B_j]\preceq t''[B_j]$ $(1\leqslant j\leqslant n)$.

由 inst(R) 定义及 $r_1,r_2\in\text{inst}(R)$，可知 $t=t''$

又因 $\text{NH}(B_j)<\text{NH}(R)=n+1$，

根据归纳假设，由 $t[B_j]\preceq t'[B_j]$ 和 $t'[B_j]\preceq t[B_j]$，可得 $t[B_j]=t'[B_j]$ 因此 $t=t'$.

由以上证明可知：$\forall\ t\in r_1$，$\exists\ t'\in r_2$，使得 $t=t'$，即 $r_1\preceq r_2$，

同样可证 $\forall\ t\in r_2$，$\exists\ t\in r_1$，使得 $t=t'$，即 $r_2\subseteq r_1$，于是，$r_1=r_2$，反对称性得证.

传递性：设 $r_1,r_2,r_3\in \mathrm{inst}(R)$ 且 $r_1\preceq r_2$，$r_2\preceq r_3$.

由定义可得：$\forall\ t\in r_1$，$\exists\ t'\in r_2$，$\exists\ t''\in r_3$，使得

$t[A_1,\cdots A_m]=t'[A_1,\cdots,A_m]=t''[A_1,\cdots,A_m]$ 且 $t[B_j]\preceq t'[B_j]\preceq t''[B_j]$ $(1\leqslant j\leqslant n)$.

根据归纳假设，因 $\mathrm{NH}(B_j)<\mathrm{NH}(R)=n+1$，由 $t[B_j]\preceq t'[B_j]\preceq t''[B_j]$ 得 $t[B_j]\preceq t''[B_j]$.

由此可得 $r_1\preceq r_3$，传递性得证.

综上可得，当 $\mathrm{NH}(R)=n+1$ 时，$\preceq$ 也是 $\mathrm{inst}(R)$ 上的偏序关系.

根据归纳证明，可得 $\preceq$ 是 $\mathrm{inst}(R)$ 上的偏序关系.

定义 1.2.4. 设关系 $R=(A_1,\cdots,A_m,B_1^*,\cdots,B_n^*)$，$r_1,r_2\in \mathrm{inst}(R)$，则 r_1 和 r_2 的合并记为 $r_1\tilde{\cup} r_2$，r_1 和 r_2 的交记为 $r_1\tilde{\cap} r_2$，分别定义如下：(其中 $1\leqslant i\leqslant m$，$1\leqslant j\leqslant n$)

$$r_1\tilde{\cup} r_2=\{t\,|\,\exists\ t_1\in r_1,\exists\ t_2\in r_2:(\forall\ i\forall\ j:t[A_i]=t_1[A_i]=t_2[A_i]\wedge t[B_j]=t_1[B_j]\tilde{\cup} t_2[B_j])\}\cup\{t\,|\,t\in r_1\wedge(\forall\ t'\in r_2:(\exists\ i:t[A_i]\neq t'[A_i]))\}\cup\{t\,|\,t\in r_2\wedge(\forall\ t'\in r_1:(\exists\ i:t[A_i]\neq t'[A_i]))\}$$

$$r_1\tilde{\cap} r_2=\{t\,|\,(\exists\ t_1\in r_1,\exists\ t_2\in r_2:(\forall\ i\forall\ j:t[A_i]=t_1[A_i]=t_2[A_i]\wedge t[B_j]=t_1[B_j]\tilde{\cap} t_2[B_j])\}$$

定理 1.2.2. 设 r_1、r_2 是关系 $R=(A_1,\cdots,A_m,B_1^*,\cdots,B_n^*)$ 的两个实例，则 $r_1\tilde{\cup} r_2$，$r_1\tilde{\cap} r_2$ 分别是 r_1 和 r_2 的最小上界和最大下界，即

$$\mathrm{GLB}(r_1,r_2)=r_1\tilde{\cap} r_2\quad \mathrm{LUB}(r_1,r_2)=r_1\tilde{\cup} r_2$$

证明：先证 $\mathrm{LUB}(r_1,r_2)=r_1\tilde{\cup} r_2$.

设 $r=r_1\tilde{\cup} r_2$，则只需证明：对任意的 $r'\in \mathrm{inst}(R)$，如果 $r_1\preceq r'$，$r_2\preceq r'$ 那么 $r\preceq r'$.

对嵌套深度施行归纳：

归纳基础：当 $\mathrm{NH}(R)=1$ 时，退化为集合论中运算和比较，即 $r=r_1\cup r_2$，$r_1\preceq r'$，$r_2\preceq r'$，由以上可得 $r\preceq r'$.

归纳假设：设 $\mathrm{NH}(R)=n$ 时，要证明的命题成立.

归纳步骤：当 $\mathrm{NH}(R)=n+1$ 时，若 $t\in r=r_1\tilde{\cup} r_2$，则 t 有 3 种来源

①$t\in r_1\wedge(\forall\ t'\in r_2:(\exists\ i:t[A_i]\neq t'[A_i]))$

②$t\in r_2\wedge(\forall\ t'\in r_1:(\exists\ i:t[A_i]\neq t'[A_i]))$

③$t=(t[A_1],\cdots,t[A_m],t_1[B_1]\tilde{\cup} t_2[B_1],\cdots,t_1[B_n]\tilde{\cup} t_2[B_n])$

(其中 $t_1\in r_1$，$t_2\in r_2$ 且 $t_1[A_1,\cdots,A_m]=t_2[A_1,\cdots,A_m]$)

对于来源①②，t 分别属于 r_1 和 r_2，因为 $r_1\preceq r'$，$r_2\preceq r'$，所以 $\exists\ t'\in r'$，

$$t[A_1,\cdots,A_m]=t'[A_1,\cdots,A_m]\wedge\forall\ j(1\leqslant j\leqslant n):t[B_j]\preceq t'[B_j]$$

对于来源③，根据归纳假设可知：

$\forall\ j$，$1\leqslant j\leqslant n$，$t[B_j]=t_1[B_j]\tilde{\cup} t_2[B_j]=\mathrm{LUB}(t_1[B_j],t_2[B_j])$ $\exists\ t_1\in r_1$，$\exists\ t_2\in r_2$ 使得 $t_1[A_1,\cdots,A_m]=t_2[A_1,\cdots,A_m]$.

由 $r_1\preceq r'$，$r_2\preceq r'$，可知 $\exists\ t'\in r'$，$t'=(t[A_1,\cdots,A_m],t'[B_1,\cdots,B_m])$. 且 $\forall\ j$，$1\leqslant j\leqslant n$，有 $t_1[B_j]\preceq t'[B_j]$，$t_2[B_j]\preceq t'[B_j]$. 而由 $t_1[B_j]\tilde{\cup} t_2[B_j]=\mathrm{LUB}(t_1[B_j],t_2[B_j])$，可得 $t[B_j]$

$=t_1[B_j]\widetilde{\cup}t_2[B_j]\leq t'[B_j]$ 从而可证结论成立.

归纳结论：综上所述 $\mathrm{LUB}(r_1,r_2)=r_1\widetilde{\cup}r_2$.

同理可证 $\mathrm{GLB}(r_1,r_2)=r_1\widetilde{\cap}r_2$.

定理 1.2.3. 对于关系 R，偏序集(inst(R)，$\leq$)构成一个完全格，它所诱导的代数系统是(inst(R)，$\widetilde{\cup}$，$\widetilde{\cap}$).

证明：由定理 1.2.1 知 $\leq$ 是 inst(R)上的偏序关系，由定理 1.2.2 知对任意的 $r_1,r_2\in$ inst(R)，它们组成的集合都有最大下界和最小上界. 因而可知，inst(R)的每个子集都有最大下界和最小上界. 由此，定理得证.

推论. 设 $D=(R_1,\cdots,R_k)$是一个数据库模式，$\leq$是 inst(D)上的关系，则(inst(D)，$\leq$)构成一个完全格.

证明：因为 inst(R_1)，…，inst(R_k)在$\leq$关系下分别构成完全格，由格的性质可知完全格的叉乘也是完全格，故(inst(D)，$\leq$)构成完全格.

2　COM 的查询语言 CO－DATALOG

2.1　CO－DATALOG 语法

定义 2.1.1. 项、基项

设 R 是一个关系，$\mathscr{A}$ 是属性名集合，$\mathscr{V}$ 是变量集合，则 CO－DATALOG 项可递归定义如下：

①如果 $A\in\mathscr{A}$，$X\in\mathscr{V}$，则

(i)A 为单值属性，则 $A:X$ 是项；

(ii)A 为多值属性，则 $A:\{X\}$是项.

② 如果 $A\in\mathscr{A}$，且 A 为基本属性，

(i)A 为单值属性，则 $A:a(a\in\Omega)$是项；

(ii)A 为多值属性，则 $A:\{a_1,\cdots,a_k\}$(其中 $a_i\in\Omega$，$1\leqslant i\leqslant k$)是项.

③如果 $A,A_1,\cdots,A_n\in\mathscr{A}$，$A_1,\cdots,A_n$ 是 A 的属性，L 是 $A_1,\cdots,A_n$ 的属性模式的表列.

(i)A 的模式为 $A(L)$，则 $A:[r_1,\cdots,r_n]$(其中对$\forall\ i,1\leqslant i\leqslant n,r_i$ 为属性 A_i 的项)是项；

(ii)A 的模式为 $A^*(L)$，则 $A:\{[r_{11},\cdots,r_{1n}],\cdots,[r_{m1},\cdots,r_{mn}]\}$ (其中对$\forall\ i,1\leqslant i\leqslant m$，$\forall\ j,1\leqslant j\leqslant n,r_{ij}$是属性 A_j 的项)是项.

没有变量出现的项称为基项.

定义 2.1.2. 原子、事实

设 R 的关系模式为 $R(A_1,\cdots,A_m,B_1^*,\cdots,B_n^*)$，则 $R(r_1,\cdots,r_m,r'_1,\cdots,r'_n)$ (其中$\forall\ i$，$1\leqslant i\leqslant m$，$\forall\ j,1\leqslant j\leqslant n,r_i$ 为 A_i 项，r'_j 为 B_j 的项)是一个原子. 关系名 R 也称为谓词名. 没有变量出现的原子称为基原子，也称为事实.

定义 2.1.3. 规则、程序

设 $D=(R_1,\cdots,R_k)$是一个数据库模式，则

① CO－DATALOG 规则定义为 $H:-B_1,\cdots,B_n$ 其中 $H,B_i(1\leqslant i\leqslant n)$是关系 R_j $(1\leqslant j\leqslant k)$的原子.

各原子中都不出现变量的规则称为基规则;无体($n=0$)基规则表示事实.

②CO－DATALOG 程序是一个有穷的 CO－DATALOG 规则集合.

从以上定义可以看出,CO－DATALOG 是 DATALOG 在 COM 模型下的推广.为保证有实际意义,我们对 CO－DATALOG 规则(程序)作一些限制.

定义 2.1.4. 规则的实例化

设 $D=(R_1,\cdots,R_k)$是一个 COM 数据库模式,$\rho=H:-B_1,\cdots,B_n(n\geqslant 0)$是一个 CO－DATALOG规则.

①ρ 的实例化 Θ 是从 ρ 中出现的所有变量到所有实例值的一个映射,用 $H\Theta,B_i\Theta$ 表示对应的原子中变量被 Θ 所规定的实例值替换后的结果(基原子),称为实例化原子 $\rho\Theta=H\Theta:-B_1\Theta,\cdots,B_n\Theta$,称为 ρ 的实例化规则.

②对于 ρ 的一个实例化 Θ 而言,如果每个实例化的原子 $H\Theta,B_1\Theta,\cdots,B_n\Theta$ 都是数据库 D 的 COM 实例中对应的关系实例的一个元组,则称 Θ 是有效的.即存在一个 $d=(r_1,\cdots,r_k)\in \mathrm{inst}(D)$使得 $H\Theta$ 是某个关系 R_j 的实例 r_j 的一个元组,并且$\forall\ i,1\leqslant i\leqslant m,B_i\Theta$ 是某个关系 R_j 的实例 r_j 的一个元组.

定义 2.1.5. 合理规则、合理程序

①规则 ρ 是合理的,如果至少存在一个对于它的有效实例化.

②程序是合理的,如果 P 中每个规则 ρ 都是合理的.

2.2 CO－DATALOG 语义

定义 2.2.1. CO－DATALOG 空间,基及解释

设 $D=(R_1,\cdots,R_k)$是个 COM 数据库模式,P 是 D 上一个 CO－DATALOG 程序,则

①P 的 CO－DATALOG 空间 U_P 是由 $R_1,\cdots,R_k$ 各个层次上出现的属性的所有属性实例组成的集合;

②P 的 CO－DATALOG 基 B_P 即是 $\mathrm{inst}(D)$;

③P 的解释 I_P 是 D 的一个 COM 实例 $d=(r_1,\cdots,r_k)\in \mathrm{inst}(D)$其中 $r_i(1\leqslant i\leqslant k)$称为 R_i 的解释,记为 I_{Ri}.

定义 2.2.2. 满足

设 $I_P=(I_{R1},\cdots,I_{Rk})$为一解释,满足$\models$定义为

①对于 $R_j(1\leqslant j\leqslant k)$的基原子 F,$I_P\models F$ 如果 $F\preceq I_{Rj}$,

②对于规则 $\rho=H:-B_1,\cdots,B_n$,$I_P\models\rho$ 当且仅当对 ρ 的每一个实例化 Θ,如果 $I_P\models B_1\Theta,\cdots,I_P\models B_n\Theta$ 则 $I_P\models H\Theta$.

对于程序 P,$I_P\models P$,当且仅当对 P 中任一规则 ρ 都有 $I_P\models\rho$.

注意,此处所定义的概念与 DATALOG 中的概念有一重要区别,即实例都限制为 COM 实例集合,实例化也指有效的实例化,正因为这一限制才可以为语法上含有高阶项的 CO－DATALOG 构造一阶语义.

定义 2.2.3. CO－DATALOG 程序模型、最小模型

程序 P 的模型是使 P 得以满足的解释.

若 M 和 N 都为 P 的模型,如果 $M\preceq N$,则称 M 小于 N.

M 为 P 的最小模型当且仅当对 P 的任何模型 M',都有 M 小于 M',即 $M\preceq M'$.

定理 2.2.1. 模型相交性质、最小模型存在性

设 P 是 $D=(R_1,\cdots,R_k)$ 上的一个 CO-DATALOG 程序,MOD(P)是 P 的所有模型

①若 M_1、$M_2\in$ MOD(P),则 $M_1 \tilde{\sqcap}^D M_2\in$ MOD(P).

②程序 P 有最小模型 MM,且 $MM=\tilde{\sqcap}^D_{M_i\in MOD(P)} M_i$.

证明:①设 M_1,$M_2\in$ MOD(P)需证$\forall\ \rho\in P$,$M_1\tilde{\sqcap}^D$, $M_2\models\rho$.

再假设 $\rho=H$:$-B_1,\cdots,B_n$ 是 P 中的任一规则,Θ 是 ρ 的任一实例化替换.

要证 $M_1\tilde{\sqcap}^D M_2\models\rho$,只需证:

如果 $M_1\tilde{\sqcap}^D M_2\models B_1\Theta,\cdots,M_1\tilde{\sqcap}^D M_2\models B_n\Theta$,则 $M_1\tilde{\sqcap}^D M_2\models H\Theta$.

因为 $M_1\tilde{\sqcap}^D M_2\models B_i\Theta$ 等价于 $M_1\models B_i\Theta,M_2\models B_i\Theta,(1\leqslant i\leqslant n)$.

因而只需证:

如果 $M_1\models B_1\Theta,\cdots,M_1\models B_n\Theta,M_2\models B_1\Theta,\cdots,M_2\models B_n\Theta$,则 $M_1\tilde{\sqcap}^D M_2\models H\Theta$.

因为 M_1,$M_2\in$ MOD(P),根据模型的定义:

由 $M_1\models B_1\Theta,\cdots,M_1\models B_n\Theta$,可知 $M_1\models H\Theta$

由 $M_2\models B_1\Theta,\cdots,M_2\models B_n\Theta$,可知 $M_2\models H\Theta$

因而只需证:如果 $M_1\models H\Theta,\cdots,M_2\models H\Theta$,则 $M_1\tilde{\sqcap}^D M_2\models H\Theta$

不失一般性,假定 H 的谓词是 R_j,那么 $H\Theta$ 是关系 R_j 上的一个实例.设模型 M_1、M_2 在 R_j 上的解释分别为 r_{1j}和 r_{2j},则由 $M_1\models H\Theta,M_2\models H\Theta$ 可知 $H\Theta\leq r_{1j},H\Theta\leq r_{2j}$,由此可得 $H\Theta\leq r_{1j}\tilde{\sqcap} r_{2j}$,根据定义 $M_1\tilde{\sqcap}^D M_2$ 在 R_j 上的解释为 $r_{1j}\tilde{\sqcap} r_{2j}$,所以可得 $M_1\tilde{\sqcap}^D M_2\models H\Theta$.

综上所述:对任意的 $\rho\in P$,如果 $M_1\models\rho,M_2\models\rho$,则 $M_1\tilde{\sqcap}^D M_2\models\rho$.

因此,如果 M_1,$M_2\in$ MOD(P),则 $M_1\tilde{\sqcap}^D M_2\in$ MOD(P).

② 由①可知,MOD(P)在 $\tilde{\sqcap}^D$ 运算下是封闭的.而在数据库实例组成的完全格(inst(D),$\leq^D$)中,格的最大值总是 P 的模型,因为它是 inst(D)中所有元素的上界,可以使任何实例化规则得以满足.因此(MOD(P),$\leq^P$)也组成一个完全格,其诱导的代数系统为(MOD(P),$\tilde{\sqcup}^D$,$\tilde{\sqcap}^D$),所有的模型在$\leq^D$ 关系下组成格,最小模型 MM 就是该格中的最小元素,故 $MM=\tilde{\sqcap}^D_{M_i\in MOD(P)} M_i$.

以上定理奠定了 CO-DATALOG 的模型论语义的基础.

定义 2.2.4. 映射 T_p

设 P 是数据库 $D=(R_1,\cdots,R_k)$上的一个 CO-DATALOG 程序,I 为 P 的一个解释,则映射 T_P:inst(D)→inst(D)定义如下:

设 $I=(r_1,\cdots,r_k)$,$I'=(r'_1,\cdots,r'_k)$,$T_P(I)=I\cup I'$,其中$\forall\ j,1\leqslant j\leqslant k,r'_j=\tilde{\bigcup}\{h\mid h:-b_1,\cdots,b_n$ 是一实例化规则,$I\models b_i(1\leqslant i\leqslant n)\}$.

映射 T_p 是一阶逻辑程序中"直接结果算子"的推广,可用类似于[6]中的方法证明以下定理.

定理 2.2.2. 最小模型的不动点性质

设 P 是数据库 $D=(R_1,\cdots,R_k)$的一个 CO-DATALOG 程序.其最小模型 $MM=lfp(T_P)$,即 $MM=T\uparrow\omega(T_P(\Phi_0))$,其中 $\Phi_0=(\Phi_{R1},\cdots,\Phi_{Rk})$是 inst($D$)中的最小(空)实例,$\Phi_{Ri}$是inst($R_i$)的最小(空)实例.

证明:①在关系$\preceq^D$之下,inst(D)构成格. 由定义 2.2.4 显见 T_P 是单调及连续映射.

②解释 I 是 P 的模型当且仅当 I 使 P 中每一实例化规则 $h:-b_1,\cdots,b_n$. 满足,即对于 $1\leqslant i\leqslant n$,$b_i\preceq I$ 意味着 $I\models h$ 当且仅当 $T_P(I)\preceq I$.

③根据模型相交性质,可知 $MM=GLB(I|I\models P)$. 即 $MM=GLB\{I|T_P(I)\preceq I\}$,则 $MM=lfp(T_p)$.

3 结 论

本文中,我们提出了一种复杂对象模型 COM,它与其它复杂对象模型相比具有更直观、更易于系统实现等优点,可作为面向对象数据库系统和知识库系统的基础数据模型. 我们定义了 COM 实例,从而把讨论范围局限于一类有用的实例. 然后,我们证明了(inst(R),$\preceq$)构成一个完全格. 由这一重要性质出发,我们描述了 COM 模型上的说明性逻辑数据语言 CO-DATALOG 及其语义,并证明 CO-DATALOG 程序具有唯一最小模型,并且说明最小模型的不动点性质,从而在 COM 模型下重构了一阶逻辑数据语言的语义理论框架.

参考文献

1 Dadam F *et al*. A DBMS prototype to support extended NF² relations: a integrated view on flat tables and hierarchies. 1986 ACM SIGMOD, 1986. 356—366.

2 Deshpande A, Van Gucht D. An implementation for nested relational databases. Proc. of 1988 VLDB, 1988.

3 Bancilhon F, Khoshafian S. A calculus for complex objects. Journal of Computer and System Sciences, 1989, **38** (3):326—340.

4 Chen Q, Kambayashi Y. Nested relation based database knowledge representation. 1991 ACM SIGMOD, 1991. 328—337.

5 Beeri C, Kornatzky Y. The many faces of query monotonicity. Advances in Database Technology—EDBT'90 (LNCS 466), 1990. 120—135.

6 Lloyd J. Foundations of logic programming, 2nd edition. New York: Springier—Verlag, 1987.

COMPLEX OBJECT BASED KNOWLEDGE—BASE LANGUAGE

Shi Baile Zhou Aoying Guo Depei Ye Daobing

(*Department of Computer Science, Fudan University, Shanghai* 200433)

Abstract The study on complex object data model and its logic data language has received a lot of attention in recent years. In this paper, a complex object data model COM is proposed, and the syntax of its declarative query language, called CO—DATALOG, is presented. In the respect of semantics, the concepts such as universe, base, interpretation, satisfaction, and model are defined, then based on the lattice property of COM instances the model intersection theorem, the theorem on existence of the least model, and the fixpoint characteristic of the least model are proved. Therefore the semantic theoretical framework is reconstructed.

Key words Complex object, logic data language, model—theoretical semantics.