第10卷第4期
1999年4月

软 件 学 报
JOURNAL OF SOFTWARE

Vol. 10, No. 4
Apr. 1999

单继承记录类型的代数描述*

何自强

(北京航空航天大学计算机科学与工程系 北京 100083)

摘要 通过在子类型和超类型之间引进类型转换函数,推广了有序类别代数的定义.按照这一定义,一个类型可以不一定是它的超类型的子集.这使得记录模型成为有序类别代数.为了描述类型之间的单继承关系,引进了森林类别型构的概念.给出了等式的新意义,并证明了等式理论有初始模型.

关键词 有序类别代数,等式理论,单继承性,记录.

中图法分类号 TP311

数据类型的代数方法把数据类型的规范和实现分离开.一个规范是一个代数理论,而它的实现是该理论的一个模型.为了解释和描述子类型机制和继承性,Goguen 等人[1]把多类别代数推广为有序类别代数.他们认为,子类型是超类型的子集,子类型的元素即是超类型的元素.在面向对象语言中,常常将对象的状态表示为记录,而类型是记录的集合.Cardelli 等人[2,3]对记录类型进行了深入的研究.在记录类型中,子类型除了具有超类型的域之外,还可能有更多的域.子类型中的记录并不简单地就是超类型中的记录.为了使记录类型中的子类型成为子集,Bruce 等人[4]引进了自然数之间的部分等价关系,将记录类型中的元素定义为等价类.这样的解决办法不自然,并使问题复杂化.实际上,子类型对超类型的继承性只是要求在一个需要使用超类型元素的上下文中,可以使用子类型的元素.因此,我们需要在子类型和超类型之间引进一个类型转换函数,将子类型中的元素转换为超类型中的元素.这使得子类型不一定是超类型的子集,记录类型能够构成有序类别代数,可以使用有序类别代数理论研究记录类型,将有序类别型构的等式理论看做面向对象程序设计环境的界面,而将等式理论的模型看做该环境的实现.

若子类型是超类型的子集,并且转换函数是包含映射,则我们定义的有序类别代数就成为 Goguen 的有序类别代数.因此,我们推广了 Goguen 的有序类别代数概念.

本文只考虑类型之间的单继承关系.在单继承情况下,一个类型不能有两个直接超类型,类型之间继承关系的哈斯图是图论意义上的森林.因此,我们把表达单继承性的有序类别型构称为森林类别型构.为了保证等式理论有初始模型,我们引进了可觉察的森林类别型构的概念,使得对于每个类型,都存在表示其中元素的基项.

1 森林类别型构和代数

设$(S,\leqslant)$是有穷偏序集,$\leqslant_s$表示$\leqslant$的对称闭包,$\leqslant_s^*$表示$\leqslant_s$的自反传递闭包.设$s_1,s_2\in S$.若$s_1\leqslant_s s_2$,则称s_1和s_2可比较.若$s_1\leqslant_s^* s_2$,则称s_1和s_2连通.

定义 1. 设$(S,\leqslant)$是有穷偏序集.若对于S中任意两个不可比较的元素s_1和s_2,都不存在S中元素s_3,使得$s_3\leqslant s_1$且$s_3\leqslant s_2$,则称$(S,\leqslant)$是森林序集.

因为在类型的单继承关系中,一个类型没有多于一个直接超类型,所以森林序集表达了类型的单继承关系.

可以将集合S上的二元关系$\leqslant$和$\leqslant_s$拓广到符号串的集合S^*上:

$$s_1\ldots s_n\leqslant s'_1\ldots s'_m \text{ iff } n=m\wedge s_1\leqslant s'_1\wedge\ldots\wedge s_n\leqslant s'_n,$$

$$s_1\ldots s_n\leqslant_s s'_1\ldots s'_m \text{ iff } n=m\wedge s_1\leqslant_s s'_1\wedge\ldots\wedge s_n\leqslant_s s'_n.$$

---

* 作者何自强,1943年生,副教授,主要研究领域为数理逻辑,计算机语言的形式语义.

本文通讯联系人:何自强,北京 100083,北京航空航天大学计算机科学与工程系

本文1997-04-10收到原稿,1998-05-12收到修改稿

定义 2. 森林类别型构是一个三元序偶$(S,\leqslant,\Sigma)$,其中$(S,\leqslant)$是一个森林序集,Σ是一簇集合$\{\Sigma_{w,s}|w\in S^*,s\in S\}$,并且满足以下单调性条件:

$$对于任意\ w_1,w_2\in S^*,s_1,s_2\in S,若\ \Sigma_{w_1,s_1}\cap\Sigma_{w_2,s_2}\neq\varnothing\ 且\ w_1\leqslant w_2,则\ s_1\leqslant s_2.$$

S中的元素称为类别,$\Sigma_{w,s}$中的元素称为运算符.若$\sigma\in\Sigma_{w,s}$,则说运算符σ有秩w和值类别s,记为$\sigma:w\to s$,并将$\sigma:\varepsilon\to s$简记为$\sigma:\to s$,其中ε是空串.集合$\{w|w\in S^*$,有$s\in S$使$\sigma\in\Sigma_{w,s}\}$称为运算符σ的秩集合,记为$ar(\sigma)$.若运算符σ的秩集合有多于一个元素,则称σ是多态的.

定义 3. 如果森林类别型构$(S,\leqslant,\Sigma)$满足:

$$若\ w_1,w_2\in ar(\sigma)\ 且\ w_1\leqslant_s w_2,则\ w_1\leqslant w_2\ 或\ w_2\leqslant w_1,$$

就称$(S,\leqslant,\Sigma)$是正则的.

引理. 设$(S,\leqslant,\Sigma)$是正则的森林类别型构,σ是运算符,$w\in S^*$.如果集合$\{w'|w'\in ar(\sigma),w\leqslant w'\}$非空,则它有最小元,称为$\sigma$在$w$上的最小秩.

证明:只需证明该集合是一个链.若w是空串ε,则该集合是单元素集$\{\varepsilon\}$.设w是非空串$s_1\ldots s_n$.若$s'_1\ldots s'_n,s''_1\ldots s''_n\in ar(\sigma),s_1\ldots s_n\leqslant s'_1\ldots s'_n,s_1\ldots s_n\leqslant s''_1\ldots s''_n$,则对每个$i\leqslant n,s_i\leqslant s'_i,s_i\leqslant s''_i$,由$(S,\leqslant)$是森林序集知,$s'_i\leqslant_s s''_i$.因此,$s'_1\ldots s'_n\leqslant_s s''_1\ldots s''_n$.再由$(S,\leqslant,\Sigma)$是正则的得出,$s'_1\ldots s'_n\leqslant s''_1\ldots s''_n$或$s''_1\ldots s''_n\leqslant s'_1\ldots s'_n$.

对于集合簇$\{A_s|s\in S\}$,我们用$A_{s_1\ldots s_n}$表示集合$A_{s_1}\times\ldots\times A_{s_n}$.

定义 4. $(S,\leqslant,\Sigma)$代数A是满足下面3个条件的三元序偶:

$$(\{A_s|s\in S\},\{A_{s_1\leqslant s_2}|s_1,s_2\in S;s_1\leqslant s_2\},\{A_\sigma^{w,s}|w\in S^*,s\in S,\sigma\in\Sigma_{w,s}\}),$$

其中每个A_s是非空集合,称为类型.如果$s_1\leqslant s_2$,则称A_{s_1}为A_{s_2}的子类型,也称A_{s_2}是A_{s_1}的超类型.子类型A_{s_1}到超类型A_{s_2}的类型转换函数$A_{s_1\leqslant s_2}:A_{s_1}\to A_{s_2}$,$\Sigma_{w,s}$中的运算符$\sigma$表示运算$A_\sigma^{w,s}:A_w\to A_s$.

(1) 对每个$s\in S$,$A_{s\leqslant s}$是S上的恒同映射;

(2) 若$s_1\leqslant s_2,s_2\leqslant s_3$,则$A_{s_1\leqslant s_3}$是$A_{s_1\leqslant s_2}$和$A_{s_2\leqslant s_3}$的复合函数$A_{s_2\leqslant s_3}\circ A_{s_1\leqslant s_2}$;

(3) 若$\sigma\in\Sigma_{s_1\ldots s_n,s}\cap\Sigma_{s'_1\ldots s'_n,s'}$且$s'_1\ldots s'_n\leqslant s_1\ldots s_n$,则对于每个$(x_1,\ldots,x_n)\in A_{s'_1\ldots s'_n}$,

$$A_{s'\leqslant s}(A_\sigma^{s'_1\cdots s'_n,s'}(x_1,\ldots,x_n))=A_\sigma^{s_1\cdots s_n,s}(A_{s'_1\leqslant s_1}(x_1),\ldots,A_{s'_n\leqslant s_n}(x_n)).$$

定义 5. 设A是$(S,\leqslant,\Sigma)$代数,$s,s_1,s_2\in S,s_1\leqslant s,s_2\leqslant s,a_1\in A_{s_1},a_2\in A_{s_2}$.如果$A_{s_1\leqslant s}(a_1)=A_{s_2\leqslant s}(a_2)$,则称$a_1$和$a_2$在$A_s$上相等,记为$a_1=_{A_s}a_2$.

定义 6. 设A和B是两个$(S,\leqslant,\Sigma)$代数.若函数簇$h=\{h_s:A_s\to B_s|s\in S\}$满足:

(1) 对于任意$s_1,\ldots,s_n,s\in S,\sigma\in\Sigma_{s_1\ldots s_n,s}$,

$$h_s(A_\sigma^{s_1\cdots s_n,s}(x_1,\ldots,x_n))=B_\sigma^{s_1\cdots s_n,s}(h_{s_1}(x_1),\ldots,h_{s_n}(x_n)),对一切(x_1,\ldots,x_n)\in A_{s_1\ldots s_n};$$

(2) 若$s_1\leqslant s_2$,则$B_{s_1\leqslant s_2}(h_{s_1}(x))=h_{s_2}(A_{s_1\leqslant s_2}(x))$,对一切$x\in A_{s_1}$,

则称h为从A到B的同态映射.若同态映射h中的每个h_s都是双射,则称h为同构映射.若存在从A到B的同构映射,则称A和B同构.

2 等式理论和它的初始模型

我们约定今后涉及的森林类别型构都是正则的,每个变元有唯一的S中元素作为它的类别,$(S,\leqslant,\Sigma)$代数A中的赋值v对类别是s的变元赋予A_s中的元素作为值.

定义 7. 森林类别型构$(S,\leqslant,\Sigma)$的项t及其类别$so(t)$定义如下:

(1) 若x是类别为s的变元,则x是类别为s的项;

(2) 若$\sigma\in\Sigma_{s_1\ldots s_n,s},t_1,\ldots,t_n$是项,并且$s_1\ldots s_n$是$\sigma$的$so(t_1)\ldots so(t_n)$上的最小秩,则$\sigma(t_1,\ldots,t_n)$是类别为$s$的项.

定义 8. 项t在$(S,\leqslant,\Sigma)$代数A和A中赋值v下所指的元素$A(t)(v)$定义如下:

(1) 若t是变元x,则$A(t)(v)=v(x)$;

(2) 若t是$\sigma(t_1,\ldots,t_n)$,$s_1\ldots s_n$是σ的$so(t_1)\ldots so(t_n)$上的最小秩,则

$$A(t)(v)=A_{\sigma}^{s_1\cdots s_n,so(t)}(A_{so(t_1)\leqslant s_1}(A(t_1)(v)),\ldots,A_{so(t_n)\leqslant s_n}(A(t_n)(v))).$$

不出现变元的项称为基项.基项 t 的意义与赋值无关,因此,可将 $A(t)(v)$ 简记为 $A(t)$.

定义 9. 若 t_1 和 t_2 是 $(S,\leqslant,\Sigma)$ 的项,$s\in S$,$so(t_1)\leqslant s$,$so(t_2)\leqslant s$,则称 $t_1=_s t_2$ 是 $(S,\leqslant,\Sigma)$ 等式.若 Γ 是一个由 $(S,\leqslant,\Sigma)$ 等式构成的集合,则称 $((S,\leqslant,\Sigma),\Gamma)$ 为等式理论.

定义 10. 设 $t_1=_s t_2$ 是 $(S,\leqslant,\Sigma)$ 等式,若对于 $(S,\leqslant,\Sigma)$ 代数 A 中每个赋值 v,$A(t_1)(v)=_{A_s}A(t_2)(v)$,则称 $t_1=_s t_2$ 在 A 中有效,记为 $A\models t_1=_s t_2$.

定义 11. 如果 Γ 中的每个等式都在 $(S,\leqslant,\Sigma)$ 代数 A 中有效,则称 A 是等式理论 $((S,\leqslant,\Sigma),\Gamma)$ 的模型.如果从 $((S,\leqslant,\Sigma),\Gamma)$ 的模型 A 到它的每个模型有唯一的同态映射,则称 A 为 $((S,\leqslant,\Sigma),\Gamma)$ 的初始模型.

定义 12. 设 S' 是满足以下条件的 S 的最小子集:

(1) 若 $\Sigma_{\varepsilon,s}$ 是非空集,则 $s\in S'$;

(2) 若 $s_1\in S'$ 且 $s_1\leqslant s_2$,则 $s_2\in S'$;

(3) 若 $s_1,\ldots,s_n\in S'$ 且 $\Sigma_{s_1\ldots s_n,s}$ 是非空集,则 $s\in S'$.

如果 $S'=S$,则称森林类别型构 $(S,\leqslant,\Sigma)$ 为可觉察的.

若 $(S,\leqslant,\Sigma)$ 是可觉察的正规森林类别型构,则对每个 $s\in S$,存在基项 t 使 $so(t)\leqslant s$.因此,对每个类型都存在指称其中元素的基项,这保证了下面定义的典范代数中的每个类型都是非空的.我们约定,今后仅讨论可觉察的正规森林类别型构.

设 $((S,\leqslant,\Sigma),\Gamma)$ 是等式理论.对于每个 $s\in S$,我们定义集合 $\{t\mid t$ 是基项,$so(t)\leqslant s\}$ 上的等价关系 $\equiv_s$ 如下:

$t_1\equiv_s t_2$　当且仅当　对于 $((S,\leqslant,\Sigma),\Gamma)$ 的每个模型 A,$A(t_1)=_{A_s}A(t_2)$.

对于类别 $\leqslant s$ 的基项 t,用 $[t]_s$ 表示 t 所在的关于 $\equiv_s$ 的等价类.

定义 13. 等式理论 $((S,\leqslant,\Sigma),\Gamma)$ 的典范代数 C 定义如下:

(1) 对每个 $s\in S$,$C_s=\{[t]_s\mid t$ 是基项,$so(t)\leqslant s\}$;

(2) 若 $s_1\leqslant s_2$,则 $C_{s_1\leqslant s_2}([t]_{s_1})=[t]_{s_2}$;

(3) 若 $\sigma\in\Sigma_{s_1\ldots s_n,s}$,则 $C_{\sigma}^{s_1\cdots s_n,s}([t_1]_{s_1},\ldots,[t_n]_{s_n})=[\sigma(t_1,\ldots,t_n)]_s$.

定理 1. C 是 $((S,\leqslant,\Sigma),\Gamma)$ 的模型,并称 C 为 $((S,\leqslant,\Sigma),\Gamma)$ 的典范模型.

证明:任取 Γ 中等式 $t=_s u$ 和 C 中赋值 v,$C_{so(t)\leqslant s}(C(t)(v))=C_{so(u)\leqslant s}(C(u)(v))$.

定理 2. 等式理论 $((S,\leqslant,\Sigma),\Gamma)$ 的典范模型 C 是它的初始模型.

证明:从 C 到 $((S,\leqslant,\Sigma),\Gamma)$ 的模型 A 的唯一同态映射 h 为 $h_s([t]_s)=A_{so(t)\leqslant s}(A(t))$.

例 1:集合、多重集、栈、非空栈、数的代数规范可表示为下面的等式理论.

首先引进一个森林类别型构 $(S,\leqslant,\Sigma)$.令 $S=\{j,d,z,f,s\}$,这里 j,d,z,f,s 分别是集合、多重集、栈、非空栈、数的汉语拼音的第 1 个字母.$\leqslant$ 是 S 上二元关系 $\{\langle s,f\rangle,\langle f,z\rangle,\langle z,d\rangle,\langle d,j\rangle\}$ 的自反传递闭包.Σ 中的运算符如下:

$\cup: jj\rightarrow j$,　$\cup: dd\rightarrow d$,　$0:\rightarrow s$,　$\mathrm{succ}: s\rightarrow s$,

$\mathrm{push}: sz\rightarrow f$,　$\mathrm{pop}: f\rightarrow z$,　$\mathrm{top}: f\rightarrow s$,　$\mathrm{empty}:\rightarrow z$.

我们用加下标的 x,y,z 表示变元,其中下标是变元的类别.Γ 中的等式如下:

$\mathrm{top}(\mathrm{push}(x_s,y_z))=_s x_s$,　$\mathrm{pop}(\mathrm{push}(x_s,y_z))=_z y_z$,　$\mathrm{push}(x_s,\mathrm{empty})=_f x_s$,

$\mathrm{push}(x_s,\mathrm{push}(y_s,z_z))=_d \mathrm{push}(y_s,\mathrm{push}(x_s,z_z))$,　多重集不考虑元素出现的次序

$\mathrm{push}(x_s,\mathrm{push}(x_s,y_z))=_j \mathrm{push}(x_s,y_z)$,　集合不考虑元素的出现次数

$\mathrm{push}(x_s,y_z)=_d x_s\cup y_z$,　多重集中增加一元素使该元素的重数加 1

取等式理论 $((S,\leqslant,\Sigma),\Gamma)$ 的模型 A 如下:

$A_s=N$(自然数集);　$A_f=N^+$;　$A_z=N^*$;

$A_d=\{g\mid g:N\rightarrow N,\{n\mid g(n)\neq 0\}$ 有穷$\}$,其中 $g(n)$ 是 n 在多重集 g 中的重数;

$A_j=N$ 的所有有穷子集构成的集合.

$A_{s\leqslant f}(n)=n$(长度为 1 的串),对每个 $n\in N$;

$A_{f\leqslant x}(w)=w$,对每个 $w\in N^+$;

对每个 $w\in N^*$,$A_{x\leqslant d}(w)=g_w$,其中 $g_w(n)=n$ 在 w 中的出现次数,对每个 $n\in N$;

$A_{d\leqslant j}(g)=\{n|n\in N,g(n)\neq 0\}$,对每个 $g\in A_d$.

$A_0^{\varepsilon,s}=0$; $A_{\mathrm{empty}}^{\varepsilon,x}=\varepsilon$; $A_{\mathrm{succ}}^{s,s}(n)=n+1$,对每个 $n\in N$;

对每个 $n\in N$,$w\in N^*$,$A_{\mathrm{top}}^{f,s}(nw)=n$,$A_{\mathrm{push}}^{sx,f}(n,w)=nw$,$A_{\mathrm{pop}}^{f,x}(nw)=w$;

对任意 $g_1,g_2\in A_d$,$A_{\cup}^{dd,d}(g_1,g_2)=g$,其中 $g(n)=g_1(n)+g_2(n)$,对每个 $n\in N$;

$A_{\cup}^{jj,j}(B,C)=B\cup C$,对 N 的任意有穷子集 B 和 C.

A 与$((S,\leqslant,\Sigma),\Gamma)$的典范模型同构,是该代数规范的一个正确实现.

参考文献

1 Goguen J A, Meseguer J. Order-sorted algebra I: equational deduction for multiple inheritance, overloading, exceptions and partial operations. Theoretical Computer Science, 1992,105(2):217~273

2 Cardelli L, Mitchell J C. Operations on records. Mathematical Structures in Computer Science, 1991,1(1):3~48

3 Cardelli L. Extensible records in a pure calculus of subtyping. In: Gunter C A, Mitchell J C eds. Theoretical Aspects of Object-oriented Programming. Cambridge: MIT Press, 1994. 373~425

4 Bruce K B, Longo G. A modest model of records, inheritance, and bounded quantification. Information and Computation, 1990,87(1,2):196~240

Algebraic Description for Record Model with Single Inheritance

HE Zi-qiang

(*Department of Computer Science and Engineering* *Beijing University of Aeronautics and Astronautics* *Beijing* 100083)

Abstract Introducing the type transformation functions between subtypes and supertypes, the definition of order-sorted algebra is generalized. According to this definition, a type needn't be a subset of its supertype, and a record model may form an order sorted algebra. The concept of forest-order signature is introduced for describing single inheritance relation between types. A new meaning of equation is given. It is proved that an equational theory has its initial model.

Key words Order-sorted algebra, equational theory, single inheritance, record.