

属性文法 语义分析 拓广属性文法



28-33

李中华 施丽华 李玉茜

TP31

(华东师范大学计算机科学系, 上海200062)

摘要

The standard attribute grammar cannot work efficiently for the incremental semantic evaluation of language-based editors. An extended version of attribute grammar, that is, a new description mechanism with "broadcast" and "broadcast net" introduced into the attribute grammar, is presented in this paper, so as to establish the theoretical basis for the efficient remote message passing in the semantic tree, and efficient detection and processing of symbol confliction. At last, the implementation of broadcast and broadcast net in the incremental semantic evaluation is discussed.

一、引言

属性文法被普遍用作作为基于语言的环境中增量语义分析的基础^[1-4]。传统属性文法中,属性是局部的,即语义等式仅能引用相同产生式实例中属性的值。对于以属性文法为基础的增量语义分析,属性的局部性无法有效地传递语义树中符号定义点和引用点之间的语义信息。换言之,为了把符号的定义信息传送到引用点上,定义信息必须逐结点地经过许多中介属性;一个符号的定义和删

除可能导致对几乎整棵语义树的重新检查。

此外,为进行定义信息的传递,往往引进符号表这类聚合属性,不同结点上的聚合属性的信息存在很大冗余,极大浪费了空间,因此属性文法不能很好地作为增量语义分析的基础。

为克服传统属性文法的这些缺点,本文提出一种拓广属性文法,在传统属性文法中引进新的描述机制:广播和广播网,使符号的定义点和引用点能够有效地进行语义信息

虑时空问题,否则知识处理就失去其意义了。当然,目前各种非经典逻辑(以及现代逻辑)的研究还处在一个起步阶段,许多问题还需要人们去深入探讨,相信在这些领域里将会取得丰硕的成果。

参考文献

1. Richard Frost, "Introduction to knowledge base systems". William Collins Sons & Co. Ltd, 1986
2. Kowalski, R. A. "Logic for problem solving", Northholland, 1979
3. 石纯一等编著, "数理逻辑与集合论", 清华大学出版社, 1989
4. 王克宏等, "关于知识库研究中的若干问题", 全国第二届计算与软件理论学术会议, 1988, 成都
5. 王克宏等编著, "知识工程与知识处理系统", 清华大学出版社, (即将出版)

的交流,而且,该描述机制也为有效地进行符号冲突的检测和处理提供了理论基础。

本文的讨论中,语义树是一个基本的概念。它是属性化的推导树,即推导树的各结点上包含与其标记的语法符号相联系的属性实例的一个赋值序列。设定null是每个属性的定义域之外的特殊值,可以把它赋予一个属性实例。引进广播、广播网之后,语义树的一些结点上还包含相联系的广播、广播网的数据结构。

我们用Tree和Node分别标记语义树及其结点类型,而 $n:Node$, $n.Production$ 表示结点 n 上所使用的产生式。

二、广播

广播设置在语义树的结点上,管理以所在结点为根的子树内信息的交流。信息发送者发送二元组(关键字,值),而接收者具有接收参数,广播要使接收者接受相匹配的发送信息。

广播的定义形式为 $B: (K, V, PTS, MF, cf)$ 。其中 B 是广播名,五元组是广播规格:

K : 信息的关键字类型。

V : 信息值域。

PTS : 接收参数类型的集合。

MF : 匹配函数集合。

$\forall PT \in PTS, \exists mf: K \times PT \rightarrow$

$Boolean \in MF$

若 $mf(key, par) = true$ 则称信息关键字 key 与接收参数 par 相匹配。

$cf: K \times K \rightarrow Boolean$ 称为冲突函数,它满足:

$\forall mf \in MF, k_1, k_2 \in K, pt \in PT,$

$(mf(k_1, pt) \wedge mf(k_2, pt)) \implies (k_1 = k_2)$

$\vee cf(k_1, k_2)$

广播上定义了一组通信原语:

•发送原语 $send[B](key, sa)$: 向广播 B 申请发送信息(key, sa 的值)称 sa 为发送属性。

•接收原语 $receive[B](mf, par, ra)$ 、

$preceive[B](mf, par, ra)$: 通过 B 接收关键字与 par 相匹配的信息,将其值赋予接收属性 ra 。 $preceive$ 要求信息发送点按前序遍历次序超前于接收点。

•测试原语 $test[B](key, ta)$: 根据 cf 测试是否已有关键字与 key 冲突的发送信息,结果赋予测试属性 ta 。

称安装在结点 n 上的广播 B 为 B 的实例,记为 $n.B$ 。同一广播在语义树中可有多个实例,若结点 m 上使用通信原语 com ,则记该使用为 (m, com) ,假定 com 请求的广播名为 B ,则从 m 到语义树根结点路径上至少存在一名为 B 的广播实例,实际响应 com 的是离 m 最近的实例 $n.B$ 。此时称 (m, com) 是 $n.B$ 的通信实例。视 com 的类型分别称之为发送、接收或测试实例。每个发送实例可把信息发送至多个接收实例,而每个接收实例只能接受一个发送实例的信息。

广播实例 $n.B$ 的定义:

•发送环境 $Senv(n.B)$ = 其发送实例的集合。

•接收环境 $Renv(n.B)$ = 其接收实例的集合。

•测试操作集 $Tops(n.B)$ = 其测试实例的集合。

•匹配关系 $match(n.B) \subseteq Senv(n.B) \times Renv(n.B)$:

$match(n.B) = \{((m_1, send[B](key, sa)), (m_2, receive[B](mf, par, ra))) \mid mf(key, par)\} \cup \{((m_1, send[B](key, sa)), (m_2, preceive[B](mf, par, ra))) \mid mf(key, par) \wedge pred(m_1, m_2)\}$

这里 $pred(m_1, m_2)$ 表示按前序遍历 m_1 超前 m_2 。

•发送实例 s 的接收集 $Rset(s) = \{r \mid (s, r) \in match(n.B)\}$ 。

•接收实例 r 的发送集 $Sset(r) = \{s \mid (s, r) \in match(n.B)\}$ 。

发送环境无冲突时,对 $r \in Renv(n.B)$, $|Sset(r)| \leq 1$; 当 $|Sset(r)| > 1$ 时, r 接收 $Sset$

(r)中某一个发送实例的信息, $|Sset(r)|=0$ 时, r 称广播实例 $n.B$ 的失配(接收)点。

广播的功能是使如下广播谓词成立:

$$(1) \forall r(ra) \in Renv(n.B) ((Sset(r(ra)) \neq \emptyset \Rightarrow \exists s(sa) \in Sset(r(ra)) \wedge (ra.value = sa.value) \wedge (Sset(r(ra)) = \emptyset \Rightarrow ra.value = null))$$

其中 $r(ra)/s(sa)$ 指接收/发送属性为 $ra/$ sa 的接收/发送实例。

$$(2) \forall (m, test[B](key, ta)) \in Tops(n.B) (ta.value = true \iff \exists (m_1, send[B](key_1, sa)) \in Senv(n.B) cf(key, key_1))$$

例如, 对C语言中标号, 为实现标号定义处与引用处语义信息的交流, 可定义简单的标号广播 $LB_i(K, V, PTS, MF, cf)$, 它安装在每个函数所对应的语义子树的根结点上。其中, $K=String$ (字符串), $V=Addr$ (代码地址), $PTS=\{String\}$, $MF=\{mf\}$, $\forall key, par \in String, mf(key, par) \equiv (key=par)$, $\forall key_1, key_2 \in String$ $cf(key_1, key_2) \equiv (key_1=key_2)$ 。

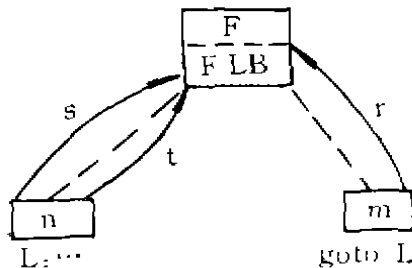


图1

图1中 $s=send[LB](L, codeaddr)$, $t=test[LB](L, defined)$, $r=receive[LB](mf, L, gotoaddr)$ 。标号定义处通过测试原语 t 测试 L 是否重复定义; 通过发送原语 s 发送标号 L 及其代码地址; 标号使用处通过接收原语 r 得到相应的代码地址。

三、广播网

由上节讨论可知, 广播的管理范围是局部的。广播中没有提供刻画名作用域的手段, 因此对Pascal这类嵌套过程语言, 它不能支持语义信息从名定义点到引用点的有效传

递。为此本节引进广播网的概念。

广播网(简称网)的定义形如 $BN_i(P_0, K, V, PTS, MF, cf)$, 其中 BN 是网名, 六元组为网规格。 P_0 是属性文法中产生式集合的子集。其余符号同广播中定义。

网上提供的通信原语与广播相同。网与整棵语义树相联, 一旦定义了网 BN , 则在语义树的每个结点上均可对该网使用网通信原语。网的通信实例、发送环境 $NSenv(BN)$ 、接收环境 $NRenv(BN)$ 、测试操作集 $NTops(BN)$ 的定义与广播实例中相应的定义类同。

为定义网上的匹配关系 $match(BN) \subseteq NSenv(BN) \times NRenv(BN)$, 引进如下函数及关系(以下所有讨论均针对网 BN , 故参数 BN 从略):

- (1) $bound: Node \rightarrow Node \cup \{null\}$
 $bound(n) = \text{if isroot}(n) \text{ null}$
 $\text{else if parent}(n).production \in P_0 \text{ parent}(n)$
 $\text{else bound}(\text{parent}(n))$
- (2) $boundset: Node \rightarrow 2^{NSenv}$
 $boundset(n) = \{(m, send(key, sa)) \mid bound(m) = n\}$
- (3) $weakuse \subseteq NRenv \times Node$
 $(m, receive(mf, par, ra)) \text{ weakuse } n \iff \text{isancestor}(n, m) \text{ and } (\exists (m_1, send(key, sa)) \in boundset(n) \text{ mf}(key, par))$
 $(m, preceive(mf, par, ra)) \text{ weakuse } n \iff \text{isancestor}(n, m) \text{ and } (\exists (m_1, send(key, sa)) \in boundset(n) (mf(key, par) \wedge pred(m_1, m)))$
- (4) $notuseuntil \subseteq NRenv \times Node$
 $ri(m) = (m, rcom(mf, par, ra))$
 $\text{其中 } rcom \text{ 为接收原语, 则}$
 $ri(m) \text{ notuseuntil } n \iff \text{isancestor}(n, m) \wedge \forall n' ((\text{isancestor}(n', m) \wedge \text{isancestor}(n, n')) \Rightarrow ((ri(m), n') \notin \text{weakuse}))$

由以上定义, BN 上的 $match$ 关系定义为:

$$\begin{aligned} \text{match} = & \{ ((m_1, \text{send}(\text{key}, \text{sa})), \\ & (m_2, \text{receive}(\text{mf}, \text{par}, \text{ra})) \mid \text{mf}(\text{key}, \\ & \text{par}) \wedge \exists n \\ & (\text{bound}(m_1) = n \wedge ((m_2, \text{receive} \\ & (\text{mf}, \text{par}, \text{ra})) \text{notuseuntil } n)) \} \\ & \cup \{ ((m_1, \text{send}(\text{key}, \text{sa})), (m_2, \text{prece-} \\ & \text{ive}(\text{mf}, \text{par}, \text{ra})) \mid \text{mf}(\text{key}, \text{par}) \wedge \text{pre-} \\ & \text{ced}(m_1, m_2) \\ & \wedge \exists n(\text{bound}(m_1) = n \wedge ((m_2, \text{preceive} \\ & (\text{mf}, \text{par}, \text{ra})) \text{notuseuntil } n)) \} \end{aligned}$$

网的发送实例接收集NRset, 接收实例发送集NSset定义与广播实例的相同。对 $n: \text{Node}$, 若 $n.\text{production} \in P_0$, 则称 n 为网BN的网点。对通信实例 $sr = (m, \text{com})$, 若 $\text{bound}(m) = n$, 则称 n 为 sr 的束缚结点。

网的功能是使如下定义的网谓词成立:

(1) $\forall r (ra) \in \text{NRenv}(\text{NSset}(r(ra)))$
 $\neq \phi \implies \exists s(sa) \in \text{NSset}(r(ra))$
 $ra.\text{value} = sa.\text{value}$
 and $(\text{NSset}(r(ra)) = \phi \implies ra.\text{value} = \text{null})$

(2) $\forall (m, \text{test}(\text{key}, \text{ta})) \in \text{NTops}$
 $(\text{ta}.\text{value} = \text{true} \iff \exists (m_1, \text{send}(\text{key}_1, \text{sa}_1)) \in \text{boundset}(\text{bound}(m)) \text{cf}(\text{key}, \text{key}_1))$

由上述形式描述不难看出, 广播网实际是当前程序设计语言中名作用域规则的一种抽象的描述机制, 具有广泛的适用性。文[6]中应用广播网给出了Pascal语言符号定义和引用的语义规则的描述, 演示了该机制的能力。由于广播网取代了传统属性文法中需大量冗余设置的符号表这类聚合属性, 从而为构造有效的增量语义分析算法提供了理论基础^[6, 8]。

四、拓广属性文法

拓广属性文法是六元组 $\langle G, D, A, BA, BS, NS \rangle$, 其中:

$G = (V_t, V_n, S, P)$ 是上下文无关文法。

$D: (C, F)$ 是一代数结构, C 是属性域的有穷集合, F 是函数集合, $f \in F$ 是以属

性域的笛卡尔积为定义域的函数。

A : 属性符号集合。

$BA \subseteq A$: 基本属性符号集合, 基本属性满足: (1) 定义它的语义等式不以基本属性为自变量。(2) 基本属性不能作为接收属性和测试属性。

BS : 广播集合。

NS : 广播网集合。

对 $\bar{x} \in V_n \cup V_t$, 有一组属性与之相联, 对 $x \in V_n$, 可能有一组广播与之相联, 对 $p \in P$, 除有一组语义等式外, 必要时有一组对广播及网的通信原语。仅基本属性可作为发送原语及测试原语中的信息关键字。

五、广播及广播网的实现

1. 通信关系的定义及表示

广播的功能是使接收实例 r 接受来自 $Sset(r)$ 中某发送实例 s 的信息, 此时称 s 和 r 实际联结, 或称 s 和 r 有通信关系。广播实例 $n.B$ 上的通信关系 $\text{com}(n.B)$ 满足以下谓词:

(1) $\forall r \in \text{Renv}(n.B) (\text{Sset}(r) = \phi \wedge (\exists s \in \text{Sset}(r), (s, r) \in \text{com}(n.B)))$

(2) $\forall s_1, s_2 \in \text{Sset}(r), (s_1, r) \in \text{com}(n.B) \wedge (s_2, r) \in \text{com}(n.B) \implies (s_1 = s_2)$

随着程序的编辑, 某些发送实例可能被删除, 相应地与其通信的接收实例的接收属性的值要随之调整, 在新的发送集不空时, 接受其中某一发送实例的信息, 否则置其值为 null , 新的发送实例可能出现, 从而原发送集为空的接收实例可能接收其中某发送信息。为动态有效地调整发送实例与接收实例的通信关系, 对每个广播实例建立信息引用表, 表示发送、接收实例的通信关系, 每个表项含一发送实例及与其通信的接收实例的集合; 建立等待表表示发送集为空的接收实例的集合。

广播网上通信关系的定义与广播实例类同, 但其建立与调整远比广播复杂。删除发送实例 s 、原与 s 通信的接收实例 r , 可能 (1) 成为失配接收点或 (2) 转与发送实例 s' 通

信: s' 或与 s 束缚到同一网点, 或者束缚到 s 的某一祖先网点。插入发送实例 s , 可能导致 (1) 某失配点与 s 通信或 (2) 与束缚到 s 的某祖先结点的发送实例通信的接收实例 r , 改与 s 通信。

为了有效地调整网 BN 上通信关系 $\text{com}(\text{BN})$, 对网点 n 定义 $\text{subcom}(\text{BN}, n) = \{ (s(m), r) \mid (s(m), r) \in \text{com}(\text{BN}) \wedge \text{bound}(m) = n \}$ 其中 $s(m)$ 指结点 m 上任一发送实例。这样, 所有网点上的 subcom 形成了 com 的一个划分。因此, 可在各网点上建立信息引用表表示 subcom , 所有这些信息引用表的集合构成网通信关系 com 的表示。为了在语义树修改过程中对信息引用表作有效的调整, 在每一网点上另建立辅助数据实体: 未定义引用表 $\text{udt}(n)$ 。对 $(r, m) \in \text{udt}(n)$, 则 r 是束缚到 n 的接收实例, 且无束缚到 n 的发送实例与之匹配; m 是与 r 通信的发送实例的束缚网点, 若 r 为失配点, 则 $m = \text{null}$ 。

广播网的各网点上建立了信息引用表和未定义引用表, 类似于建立了一个“广播实例”, 不同网点的“广播实例”间相互联系。

2. 通信关系的调整

程序编辑过程可视为语义子树的一个替换序列 (图2)。替换过程必须保持语义树

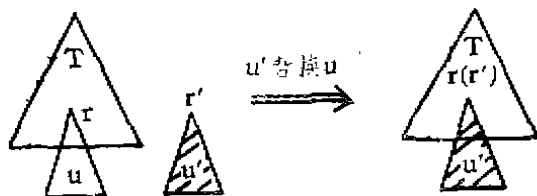


图2

的一致性^{[2][6]}, 为此须对语义树中一些属性实例重新赋值。有效的算法应避免属性实例的重复赋值与无效赋值, 故赋值应按一定的优先关系进行。通信关系是确定该优先关系的重要依据, 从而替换语义子树时应该同时调整通信关系, 包括:

(1) 子树 u 被删除之前, 先删除 u 与 $T-u$ 之

间的通信关系, 即从 $T-u$ 的所有广播实例、网点的信息引用表、等待表或未定义引用表中删除 u 的所有非局部发送实例及接收实例。

(2) 子树的替换可能导致语义树中基本属性实例的重新赋值, 由于基本属性可作为发送、接收实例的参数, 故首先对它们重新赋值, 并对通信关系作相应的处理。

(3) 子树 u' 插入后, 应建立 u' 与 $T-u$ 的通信关系, 即把 u' 所有非局部发送、接收实例插入至 $T-u$ 相应的广播实例及网点的数据集中。

六、结束语

我们的工作受到 Reps^{[1][2]}、Johnson^[3] 的影响, 但基于拓广属性文法可以更有效地实现增量语义分析^{[5][6]}。Reps 虽然引进“向上远程引用”及“表”的概念, 但定义信息向引用点的传递仍然不直接, 在说明区信息仍需逐点传递。此外, 表的定义中未考虑名活动域。Johnson 虽然引进了“非局部属性文法”, 但未给出有效的实现, 且未讨论名冲突、名活动域等问题。

参考文献

- [1] Reps, T., et al, Remote Attribute Updating for Language-Based Editors, 13th POPL Conference (1986)
- [2] Reps, T., et al, Incremental context-dependent analysis for language-based editors, ACM Trans. Prog. Lang. Syst. 5, 3 (July 1983), 449-477
- [3] Johnson, G.F., et al, A meta-language and system for nonlocal incremental attribute evaluation in language-based editors, 12th POPL Conference (1985)
- [4] Hoover, R., Dynamically bypassing copy rule chains in attribute grammars, 13th POPL Conference (1986)
- [5] Li, Z.H., et al, The Extended Attribute Grammar and Its Applicati-

OBMS IDKE 模型 数据库 面向对象

计算机科学1992 Vol.19 No.2

(7)

编者按语

自八十年代以来,非传统应用领域对数据库需求的增长和数据库研究的深入导致了面向对象数据库(简称OODB)的产生。OODB在近几年发展中所显示的强大生命力使其有可能成为下一代数据库系统的主流。

下面刊登的三篇论文,阐述了一个OODBMS——OBMS/IDKE的数据模型、无缝的面向对象编程语言、动态模式管理以及对象存贮/存取管理中的关键技术问题。从此可以看到,OBMS/IDKE吸收了现有OODB系统和OO模型的特点,从对象语义出发对OO模型的语义关联进行了考察,形成了自己的对象模型语义体系,以及形成了从对象的物理存贮、缓冲区管理、概念模型到语言界面一整套具有特色的对象库系统。

OBMS/IDKE: 模型及其 无缝的面向对象编程语言*)

33-43

唐元昌 王珊

TP317.13

(中国人民大学数据与知识工程研究所, 北京100872)

摘

要

With the development of database application, it is urgent to break through Relation Model and RDBMS's limitation in structure modelling and semantics expressing. In the paper, we have briefly summarized various approaches to the aim, and introduced our object base management system OBMS/IDKE, especially its object-oriented data model PUC which is based on complete semantics inheritance, and its seamless object-oriented programming language ODPL.

一、关系模型以来的发展方向 (后关系模型)

在CAD/CAM、OIS、VLSI、CASE等计

算机应用领域,传统的关系数据库缺乏灵活、丰富的建模能力,已不能满足要求。为了解决这一问题,人们提出了许多新的并发展了一些原有的数据模型。这些尝试沿着如下几

*) 国家自然科学基金资助项目

on in Incremental Semantic Evaluation, 2th Proceeding of ICYCS

(6) 施丽华, 拓广属性文法及在增量语义分析中的应用, 硕士论文, 华东师大, 91.1

(7) 面向Pascal的句法制导的程序设计环境的设计与实现, 技术报告, 华东师大计算机科学系, 90.9