Compiler Construction

Lecture 11: Syntactic Analysis VI ($LR(0)$ Parsing)

Thomas Noll

Lehrstuhl für Informatik 2
(Software Modeling and Verification)

RWTH Aachen University

noll@cs.rwth-aachen.de

http://www-i2.informatik.rwth-aachen.de/i2/cc10/

Winter semester 2010/11

Outline

Nondeterministic Bottom-Up Automaton I

Definition (Nondeterministic bottom-up parsing automaton)

Let $G = \langle N, \Sigma, P, S \rangle \in CFG_\Sigma$. The nondeterministic bottom-up parsing automaton of G, $\mathrm{NBA}(G)$, is defined by the following components.

- Input alphabet: Σ
- Pushdown alphabet: X
- Output alphabet: $[p]$
- Configurations: $\Sigma^* \times X^* \times [p]^*$ (top of pushdown to the right)
- Transitions for $w \in \Sigma^*$, $\alpha \in X^*$, and $z \in [p]^*$:
  shifting steps: $(aw, \alpha, z) \vdash (w, \alpha a, z)$ if $a \in \Sigma$
  reduction steps: $(w, \alpha\beta, z) \vdash (w, \alpha A, zi)$ if $\pi_i = A \to \beta$
- Initial configuration for $w \in \Sigma^*$: $(w, \varepsilon, \varepsilon)$
- Final configurations: $\{\varepsilon\} \times \{S\} \times [p]^*$

Nondeterminisn in NBA(G)

Remark: NBA(G) is generally nondeterministic

- Shift or reduce? Example:

$$(bw, \alpha a, z) \vdash \begin{cases}(w, \alpha ab, z) \\ (bw, \alpha A, zi)\end{cases} \text{ if } \pi_i = A \to a$$

- If reduce: which "handle" β? Example:

$$(w, \alpha ab, z) \vdash \begin{cases}(w, \alpha A, zi) \\ (w, \alpha aB, zj)\end{cases} \text{ if } \pi_i = A \to ab \text{ and } \pi_j = B \to b$$

- If reduce β: which left-hand side A? Example:

$$(w, \alpha a, z) \vdash \begin{cases}(w, \alpha A, zi) \\ (w, \alpha B, zj)\end{cases} \text{ if } \pi_i = A \to a \text{ and } \pi_j = B \to a$$

- When to terminate parsing? Example:

$$\underbrace{(\varepsilon, S, z)}_{\text{final}} \vdash (\varepsilon, A, zi) \text{ if } \pi_i = A \to S$$

Outline

$LR(0)$ Grammars

The case $k = 0$ is relevant (in contrast to $LL(0)$): here the decision is just based on the contents of the pushdown, without any lookahead.

Corollary 11.1 ($LR(0)$ grammar)

$G \in CFG_\Sigma$ has the $LR(0)$ property if for all rightmost derivations of the form

$$S \begin{cases} \Rightarrow_r^* \alpha A w \Rightarrow_r \alpha\beta w \\ \Rightarrow_r^* \gamma B x \Rightarrow_r \alpha\beta y \end{cases}$$

it follows that $\alpha = \gamma$, $A = B$, and $x = y$.

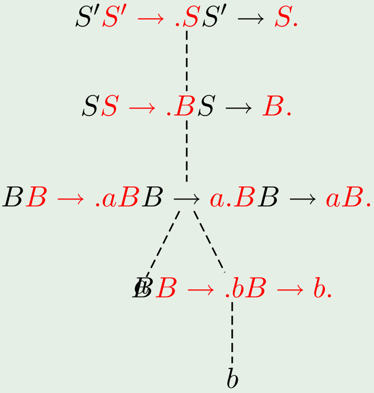
Goal: derive a finite information from the pushdown which suffices to resolve the nondeterminism (similar to abstraction of right context in LL parsing by fo-sets)

Example 11.2

$$
\begin{array}{llll}
G: & S' & \rightarrow S & (1) \\
 & S & \rightarrow B \mid C & (2,3) \\
 & B & \rightarrow aB \mid b & (4,5) \\
 & C & \rightarrow aC \mid c & (6,7)
\end{array}
$$

NBA(G):

$$
\begin{array}{ll}
 & (ab, \varepsilon, \varepsilon) \\
\vdash & (b, a, \varepsilon) \\
\vdash & (\varepsilon, ab, \varepsilon) \\
\vdash & (\varepsilon, aB, 5) \\
\vdash & (\varepsilon, B, 54) \\
\vdash & (\varepsilon, S, 542) \\
\vdash & (\varepsilon, S', 5421)
\end{array}
$$



$LR(0)$ Items and Sets II

Definition 11.3 ($LR(0)$ items and sets)

Let $G = \langle N, \Sigma, P, S \rangle \in CFG_\Sigma$ be start separated by $S' \to S$ and $S' \Rightarrow_r^* \alpha A w \Rightarrow_r \alpha\beta_1\beta_2 w$ (i.e., $A \to \beta_1\beta_2 \in P$).

- $[A \to \beta_1 \cdot \beta_2]$ is called an $LR(0)$ item for $\alpha\beta_1$.
- Given $\gamma \in X^*$, $LR(0)(\gamma)$ denotes the set of all $LR(0)$ items for γ, called the $LR(0)$ set (or: $LR(0)$ information) of γ.
- $LR(0)(G) := \{LR(0)(\gamma) \mid \gamma \in X^*\}$.

Corollary 11.4

1. *For every $\gamma \in X^*$, $LR(0)(\gamma)$ is finite.*
2. *$LR(0)(G)$ is finite.*
3. *The item $[A \to \beta\cdot] \in LR(0)(\gamma)$ indicates the possible reduction $(w, \alpha\beta, z) \vdash (w, \alpha A, zi)$ where $\pi_i = A \to \beta$ and $\gamma = \alpha\beta$.*
4. *The item $[A \to \beta_1 \cdot Y\beta_2] \in LR(0)(\gamma)$ indicates an incomplete handle β_1 (to be completed by shift operations or ε-reductions).*

$LR(0)$ Conflicts

Definition 11.5 ($LR(0)$ conflicts)

Let $G = \langle N, \Sigma, P, S \rangle \in CFG_\Sigma$ and $I \in LR(0)(G)$.

- I has a shift/reduce conflict if there exist $A \to \alpha_1 a \alpha_2, B \to \beta \in P$ such that
$$[A \to \alpha_1 \cdot a\alpha_2], [B \to \beta \cdot] \in I.$$
- I has a reduce/reduce conflict if there exist $A \to \alpha, B \to \beta \in P$ with $A \neq B$ or $\alpha \neq \beta$ such that
$$[A \to \alpha \cdot], [B \to \beta \cdot] \in I.$$

Lemma 11.6

$G \in LR(0)$ iff no $I \in LR(0)(G)$ contains conflicting items.

Proof.

omitted □

Computing $LR(0)$ Sets I

Theorem 11.7 (Computing $LR(0)$ sets)

Let $G = \langle N, \Sigma, P, S \rangle \in CFG_\Sigma$ be start separated by $S' \to S$ and reduced.

1. *$LR(0)(\varepsilon)$ is the least set such that*
   - *$[S' \to \cdot S] \in LR(0)(\varepsilon)$ and*
   - *if $[A \to \cdot B\gamma] \in LR(0)(\varepsilon)$ and $B \to \beta \in P$, then $[B \to \cdot\beta] \in LR(0)(\varepsilon)$.*
2. *$LR(0)(\alpha Y)$ $(\alpha \in X^*, Y \in X)$ is the least set such that*
   - *if $[A \to \gamma_1 \cdot Y\gamma_2] \in LR(0)(\alpha)$, then $[A \to \gamma_1 Y \cdot \gamma_2] \in LR(0)(\alpha Y)$ and*
   - *if $[A \to \gamma_1 \cdot B\gamma_2] \in LR(0)(\alpha Y)$ and $B \to \beta \in P$, then $[B \to \cdot\beta] \in LR(0)(\alpha Y)$.*

Computing $LR(0)$ Sets II

Example 11.8 (cf. Example 11.2)

$$
\begin{aligned}
G:\quad & S' \to S \\
& S \to B \mid C \\
& B \to aB \mid b \\
& C \to aC \mid c
\end{aligned}
$$

$[S' \to \cdot S] \in$

$LR(0)(\varepsilon)$ $\quad [A \to \cdot B\gamma] \in LR(0)(\varepsilon), B \to \beta \in P \implies [B \to \cdot\beta] \in LR(0)(\varepsilon)$ $\quad [A \to \gamma_1 \cdot Y\gamma_2] \in LR(0)(\alpha) \implies [A \to \gamma_1 Y \cdot \gamma_2] \in LR(0)(\alpha Y)$

$I_0 := LR(0)(\varepsilon):\quad [S' \to \cdot S] \quad [S \to \cdot B] \quad [S \to \cdot C] \quad [B \to \cdot aB] \quad [B \to \cdot b] \quad [C \to \cdot aC] \quad [C \to \cdot c]$

$I_1 := LR(0)(S):\quad [S' \to S\cdot]$

$I_2 := LR(0)(B):\quad [S \to B\cdot]$

$I_3 := LR(0)(C):\quad [S \to C\cdot]$

$I_4 := LR(0)(a):\quad [B \to a \cdot B] \quad [C \to a \cdot C] \quad [B \to \cdot aB] \quad [B \to \cdot b] \quad [C \to \cdot aC] \quad [C \to \cdot c]$

$I_5 := LR(0)(b):\quad [B \to b\cdot]$

$I_6 := LR(0)(c):\quad [C \to c\cdot]$

$I_7 := LR(0)(aB):\quad [B \to aB\cdot]$

$I_8 := LR(0)(aC):\quad [C \to aC\cdot]$

$(LR(0)(aa) = LR(0)(a) = I_4,\ LR(0)(ab) = LR(0)(b) = I_5,$
$LR(0)(ac) = LR(0)(c) = I_6,\ I_9 := LR(0)(\gamma) = \emptyset$ in all remaining cases)

Outline

The goto Function I

Observation: if $G \in LR(0)$, then $LR(0)(\gamma)$ yields deterministic shift/reduce decision for NBA(G) in a configuration with pushdown γ
$\Longrightarrow$ new pushdown alphabet: $LR(0)(G)$ in place of X

Moreover $LR(0)(\gamma Y)$ is determined by $LR(0)(\gamma)$ and Y but independent from γ in the following sense:

$$LR(0)(\gamma) = LR(0)(\gamma') \implies LR(0)(\gamma Y) = LR(0)(\gamma' Y)$$

Definition 11.9 ($LR(0)$ goto function)

The function $\text{goto} : LR(0)(G) \times X \to LR(0)(G)$ is determined by

$$\text{goto}(I, Y) = I' \quad \text{iff} \quad \begin{array}{l} \text{there exists } \gamma \in X^* \text{ such that} \\ I = LR(0)(\gamma) \text{ and } I' = LR(0)(\gamma Y). \end{array}$$

Example 11.10 (cf. Example 11.8)

$I_0 := LR(0)(\varepsilon)$: $[S' \to \cdot S]$
$[S \to \cdot B]$ $[S \to \cdot C]$
$[B \to \cdot aB]$ $[B \to \cdot b]$
$[C \to \cdot aC]$ $[C \to \cdot c]$

$I_1 := LR(0)(S)$: $[S' \to S\cdot]$

$I_2 := LR(0)(B)$: $[S \to B\cdot]$

$I_3 := LR(0)(C)$: $[S \to C\cdot]$

$I_4 := LR(0)(a)$: $[B \to a \cdot B]$ $[C \to a \cdot C]$
$[B \to \cdot aB]$ $[B \to \cdot b]$
$[C \to \cdot aC]$ $[C \to \cdot c]$

$I_5 := LR(0)(b)$: $[B \to b\cdot]$

$I_6 := LR(0)(c)$: $[C \to c\cdot]$

$I_7 := LR(0)(aB)$: $[B \to aB\cdot]$

$I_8 := LR(0)(aC)$: $[C \to aC\cdot]$

$I_9 := \emptyset$

goto	S	B	C	a	b	c
I_0	I_1	I_2	I_3	I_4	I_5	I_6
I_1						
I_2						
I_3						
I_4		I_7	I_8	I_4	I_5	I_6
I_5						
I_6						
I_7						
I_8						
I_9						

(empty $= I_9$)

Example 11.10 (continued)

Representation of goto funtion as finite automaton: