プッシュダウンリストの先読みを利用したＬＬ(２)文法の構文解析

3Ｃ－6

川島浩[†]　竹内淑子[‡]　吉田敬一[†]
静岡大学大学院理工学研究科[†]
浜松職業能力開発短期大学校[‡]

1. 目的

LL(2)文法に対する解析表の作成はAho-Ullman[5]らにより、固有の強LL(2)文法に変換され、その文法から解析表を作成する方法が提案された。しかし、このアルゴリズムにより生成された生成規則の数はかなりの大きさになり、それに伴って解析表も膨大な大きさになるという欠点がある。

また、解析表を縮小した準LL(2)文法の解析アルゴリズムの提案があるが[1]、この場合はLL(2)文法から準LL(2)文法への書き換えを必要とする。[6]

本研究では文法への書き換えをすることなく、準LL(2)文法の解析表を改良したものと、プッシュダウンリストの先読みを組み合わせることにより、LL(2)文法の新しい構文解析法を提案する。

2. 基本的定義とその記法

[定義１]文脈自由文法Gを$G=(N, \Sigma, P, S)$とする。ここで、N, Σはそれぞれ文法Gの非終端記号の集合ならびに終端記号の集合、Pは生成規則の集合であり、Sは出発記号である。

[記法１]Nの要素を$A, B, C, \ldots$、Σの要素を$a, b, c, \ldots$、$N \cup \Sigma$の要素をX, Y, Z、また、Σ^*の要素を$s, t, u, \ldots$で表し、$(N \cup \Sigma)^*$の要素を$\alpha, \beta, \gamma, \ldots$で表す。特に$\epsilon$, ϕはそれぞれ空列、空集合を表す。ϵ, ϕ以外のこれらの記号は添字等をつけて用いることもある。

[記法２]解析アルゴリズムの説明に使用する必要な事柄について述べる。

・プッシュダウンリストをRで表し、トップからi番目の記号を$R(i)$で表す。初期値は$R = S\$\$$

・入力列の終わりに$$を付加したものをMで表し、i番目の記号を$M(i)$で表す。

[定義２]集合$FIRST_k(\alpha)$は以下で定義される。

$$FIRST_k(\alpha) = \{u \mid (\alpha \overset{*}{\Rightarrow} u\beta, |u| = k) \text{ または } (\alpha \overset{*}{\Rightarrow} u, |u| < k)\}$$

ただし、$\alpha, \beta \in (N \cup \Sigma)^*$, $u \in \Sigma^*$, $|u|$はuの長さを表す。また、$\overset{*}{\Rightarrow}$は、生成規則を０回以上使用する最左導出である。

[定義３]集合PF(partial-Follow)は以下で定義される。

文脈自由文法$G=(N, \Sigma, P, S)$において

$S \overset{*}{\Rightarrow} u_i A X \xi_i$なる導出に対し

$$PF_k(A, X) = \cup_i FIRST_k(X \xi_i)$$

[定義４]L_1, L_2をΣ^*の部分集合とするとき、演算子$\oplus_k$は以下で定義される。

$L_1 \oplus_k L_2 = \{w \mid$ある$x \in L_1$, $y \in L_2$に対して、
$|xy| \leq k$ならば$w = xy$, $|xy| > k$ならば$w = u$、
ただし$xy = uv$, $|u| = k\}$

[定義５]文脈自由文法$G=(N, \Sigma, P, S)$において、$A \to \alpha$, $A \to \beta$ $(\alpha \neq \beta)$とするとき

$$S \overset{*}{\Rightarrow} uA\gamma$$

に対して

$$(FIRST_k(\alpha) \oplus_k FIRST_k(\gamma)) \cap (FIRST_k(\beta) \oplus_k FIRST_k(\gamma)) = \phi$$

が成り立つとき、文法GはLL(k)文法という。

[定義６]文脈自由文法$G=(N, \Sigma, P, S)$において、$A \to \alpha$, $A \to \beta$ $(\alpha \neq \beta)$とするとき

$$S \overset{*}{\Rightarrow} uAX\xi$$

に対して

$$(FIRST_k(\alpha) \oplus_k PF_k(A, X)) \cap (FIRST_k(\beta) \oplus_k PF_k(A, X)) = \phi$$

が成り立つとき、文法Gは準LL(k)文法という。

[定義７]生成規則の番号をpで表すとき、[ ]pまたは、[X]pをτ型生成規則番号という。ここでpは以下の導出を満足するものとする。また、$\underset{p}{\Rightarrow}$は生成規則番号$p$を用いた導出を表す。

$$S' \overset{*}{\Rightarrow} uA\gamma \underset{p}{\Rightarrow} u\beta\gamma$$

ただし、Xは$\beta \overset{*}{\Rightarrow} w$で、$|w| = 1$または$|w| = 0$のとき、$\gamma = \alpha X \delta$ $(X \overset{*}{\nRightarrow} \epsilon,\ \alpha \overset{*}{\Rightarrow} \epsilon)$である。$|w| \geq 2$のとき、[ ]$p$を表すものとする。

[定義8]解析表Ｔは行がN,列が$\Sigma^2 \cup (\Sigma \times \{\$\})$の要素によって名付けられたマトリクスとする。解析表TのA行、ab列の要素を$T(A, ab)$で表し、これらの各要素にはτ型生成規則番号の集合または、NULL(空)が記入されている。

3. ＬＬ(２)文法の解析表

導出$S \overset{*}{\Rightarrow} uA\gamma \overset{*}{\Rightarrow} uab\xi$に対応する表の要素との関係を以下に示す。

(i)準ＬＬ(２)文法に対応する表要素の関係

(１) $S \overset{*}{\Rightarrow} uA\gamma \underset{p}{\Rightarrow} u\beta\gamma \overset{*}{\Rightarrow} uab\xi\gamma$
$\Leftrightarrow T(A, ab) = [\ \ ]p$

A Parsing Method for the LL(2) Grammar Using a Lookahead Action on Pushdown List.

[†]Hiroshi Kawashima
[‡]Yoshiko Takeuchi
[†]Keiichi Yoshida
[†]Graduate School of Science and Engineering, Shizuoka University
[‡]Hamamatsu polytechnic college

(2) $S\overset{*}{\Rightarrow} uAX\delta \underset{p}{\Rightarrow} u\beta X\delta \overset{*}{\Rightarrow} uaX\delta$
$\overset{*}{\Rightarrow} uab\xi$
$\Leftrightarrow T(A,\ ab)\ = [X]p$

(3) $S\overset{*}{\Rightarrow} uAX\delta \underset{p}{\Rightarrow} u\beta X\delta \overset{*}{\Rightarrow} uX\delta$
$\overset{*}{\Rightarrow} uab\xi$
$\Leftrightarrow T(A,\ ab) = [X]p$

(ii)提案するＬＬ(２)文法に対応する表要素の関係

(i)に対する作成法を以下のように改良する。この改良はＰＬ表の初期化「step3」[1]、ＥＦ表の初期化「step4」[1]、$A\overset{*}{\Rightarrow}\varepsilon$の処理をする「step10」[1]を改良することで(ii)の表作成が可能になる。(ＬＬ(２)文法に対する解析表作成アルゴリズム、及びその証明は省略)

(1) $S\overset{*}{\Rightarrow} uA\gamma \underset{p}{\Rightarrow} u\alpha\gamma \overset{*}{\Rightarrow} uab\xi\gamma$
$\Leftrightarrow T(A,\ ab) = [\ \ ]p$

(2) $S\overset{*}{\Rightarrow} uA\alpha X\delta \underset{p}{\Rightarrow} u\beta\alpha X\delta \overset{*}{\Rightarrow} ua\alpha X\delta$
$\overset{*}{\Rightarrow} uaX\delta \overset{*}{\Rightarrow} uab\xi\delta$
$\Leftrightarrow T(A,\ ab)\ = [X]p$

(3) $S\overset{*}{\Rightarrow} uA\alpha X\delta \underset{p}{\Rightarrow} u\beta\alpha X\delta \overset{*}{\Rightarrow} uX\delta$
$\overset{*}{\Rightarrow} ua\delta \overset{*}{\Rightarrow} uab\xi$
$\Leftrightarrow T(A,\ ab) = [X]p$

(4) $S\overset{*}{\Rightarrow} uA\alpha X\delta \underset{p}{\Rightarrow} u\beta\alpha X\delta \overset{*}{\Rightarrow} uX\delta$
$\overset{*}{\Rightarrow} uab\xi\delta$
$\Leftrightarrow T(A,\ ab) = [X]p$

(i)と(ii)の大きな違いはXのもつ意味であって、(i)ではAXという並びを意味するのに対し、(ii)では、aまたはabを導出するXを意味している。

4. 解析アルゴリズム

解析表T、入力列M、プッシュダウンリストの初期値$R=S\$\$$を与え、以下の動作を行う。

(1) $M(1)=R(1)$, $M(1)$, $R(1)\in\Sigma$のとき、Rから、$R(1)$、Mから$M(1)$をポップする。
(2) $M(1)\neq R(1)$, $M(1)$, $R(1)\in\Sigma$のとき、入力エラー。
(3) $R(1)\in N$のとき、$A=R(1)$, $a=M(1)$, $b=M(2)$とし、以下の処理を行う。
解析表$T(A,\ ab)$の要素をUとして表し、Uの値によって、次の処理を行う。

1. $|U|=0$のとき、入力エラー。
2. $|U|=1$のとき、$U=\{[\ \ ]p\}$ならば、生成規則pを選択する。$U=\{[X]p\}$ならば、Rからスキャンして、Xがあれば、生成規則pを適用する。なければ、入力エラー。
3. $|U|\geq 2$のとき、$[\ \ ]p\in U$であれば、生成規則pを選択する。$[\ \ ]p\notin U$であれば、$R(2)$からスキャンし、最初に現れるX_iであるものを$U=\{[X_1]p_1, [X_2]p_2, \cdots, [X_n]p_n\}$ $(n\geq 2)$の中から選択し、生成規則p_iを選択する。(最初に現れるX_iを適用するのは、(ii)の表の作成法から明らか。)

また、上記のような解析アルゴリズムが可能であるのは次に示す定理によるものである。

[定理] $|U|\geq 2$であるとき、文法がＬＬ(２)であるとすれば、任意の要素を二つ取り出しても、以下の組み合わせは起こりえない。(証明は省略)

(1) $[\ \ ]p, [\ \ ]q,\ p\neq q$
(2) $[\ \ ]p, [X]q,\ p\neq q$
(3) $[X]p, [X]q,\ p\neq q$

5. 評価

準ＬＬ(２)文法の解析法を僅かに修正し、さらにプッシュダウンリストの先読みを導入することでＬＬ(２)文法の構文解析が可能となることを示した。

6. 今後の課題

実際のプログラミング言語をＬＬ(２)文法に書き換えて応用したとき、プッシュダウンリストの先読みの深さがどの程度行のものになるのか、並びにその深さと解析速度の関係。

参考文献

[1]吉田・竹内：準LL(2)文法に対する構文解析表の作成アルゴリズム,「情報処理学会論文誌」第３１号,第６月号(1990)
[2]吉田・竹内：準LL(2)文法に対する解析表の構造と解析アルゴリズム,「情報処理学会論文誌」第３１号,第９月号(1990)
[3]吉田・竹内：準LL(2)文法に対する構文解析高速化のための解析表の構造,「情報処理学会論文誌」第３３号,第２月号別紙(1992)
[4]井上謙蔵：コンパイラ「プログラム言語処理の基礎」pp77-117　丸善
[5]Aho,A.V.and Ullman,J.D.：*The Theory of Parsing,Translation and Compiling*,Vol.1 pp334-361 Prentice-Hall(1972)
[6]川島・吉田：テーブルを利用したLL(2)文法から準LL(2)文法への書き換えアルゴリズム,「情報処理学会第55回全国大会講演論文集」(1997)