LL(k)-грамматики и трансляции

Сильный LL(1), полный LL(1), LL(k)

В. С. Полозов

Кафедра системого программирования СПбГУ

Теория автоматов и формальных языков

Сильный $LL(1)$-анализатор

Сильный $LL(1)$-анализатор может делать ε-шаги перед определением ошибки.
Например:

$$\begin{array}{lcl} S_s & \rightarrow & a\ A\ b \mid b\ A\ a \\ A & \rightarrow & c\ S \mid \varepsilon \end{array}$$

Управляющая таблица:

	a	b	c	#
S	$a\ A\ b$	$b\ A\ a$		
A	ε	ε	$c\ S$	

И вход $aacabb$

Анализ ошибок $LL(1)$-разбора

Последний шаг перед ошибкой "уничтожил"предсказание, хотя ни $FIRST(b\#)$ ни $FIRST(cS\#)$ не содержат a.

- Полный $LL(1)$-анализатор обладает свойством *немедленного обнаружения ошибки* – ошибка выдается сразу же при анализе ошибочного символа.
- Сильный $LL(1)$-анализатор обладает свойством *максимального корректного префикса* – ошибка выдается при попытке сдвинуться по ошибочному символу (match/shift).

Полный $LL(1)$-разбор

Будем принимать решение по предсказанию $A...\#$ основываясь на $FIRST(A..\#)$, а не на $FIRST(A\ FOLLOW(A))$

- Для входного символа a и предсказания $A...\#$ проверим: $a \in FIRST(A..\#)$?

Проблема: производительность.

Полный $LL(1)$-разбор

Нам могло бы помочь следующее соображение:

$$\boxed{④X③Y②Z①\#}$$

где

- множество символов, допустимых в ① – $FIRST(\#)$
- множество символов, допустимых в ② – $FIRST(Z\#)$
- множество символов, допустимых в ③ – $FIRST(YZ\#)$
- множество символов, допустимых в ④ – $FIRST(XYZ\#)$

Полный $LL(1)$-разбор

Для нашей грамматики

$$
\begin{array}{rcl}
\text{FIRST(S)} & = & \{a, b\} \\
\text{FIRST(A)} & = & \{c, \varepsilon\}
\end{array}
$$

Первое предсказание: $S\#$:

$$
\begin{array}{c|l}
 & aacabb\# \\
\hline\hline
 & \boxed{\{a,b\}}\, S\, \boxed{\{\#\}}\, \#
\end{array}
$$

По допустимому символу a получаем:

$$
\begin{array}{c|l}
 & aacabb\# \\
\hline\hline
S_1 & \boxed{?}\; a\; \boxed{?}\; A\; \boxed{?}\; b\; \boxed{\#}\; \#
\end{array}
$$

Полный $LL(1)$-разбор

Для нашей грамматики

$$\begin{aligned}\mathrm{FIRST(S)} &= \{a, b\}\\ \mathrm{FIRST(A)} &= \{c, \varepsilon\}\end{aligned}$$

Первое предсказание: $S\#$:

$$\begin{array}{l|l} & aacabb\# \\ \hline\hline & \boxed{\{a,b\}}\, S\, \boxed{\{\#\}}\, \# \end{array}$$

По допустимому символу a получаем:

$$\begin{array}{l|l} & aacabb\# \\ \hline\hline S_1 & \boxed{?}\ a\ \boxed{?}\ A\ \boxed{?}\ b\ \boxed{\#}\ \# \end{array}$$

Полный $LL(1)$-разбор

Идея:

- Совместим нетерминал с множеством $FIRST$, которое посчитали за ним в предсказании.
- Начинаем всегда с нетерминала S и множества $\{\#\}$: обозначим как $[S, \{\#\}]$.
- Далее продолжаем по всем правилам и всем возможным предпросмотрам (look-ahead).

Для нашего примера:

- При предпросмотре $a, [S, \{\#\}]$ в правой части $a[A, \{b\}]b$

и т.п

Полный $LL(1)$-разбор

Т.о. получили переписанную грамматику:

$$
\begin{array}{lcl}
[S,\{\#\}] & \rightarrow & a\ [A,\{b\}]\ b \mid b\ [A,\{a\}]\ a \\
[S,\{a\}] & \rightarrow & a\ [A,\{b\}]\ b \mid b\ [A,\{a\}]\ a \\
[S,\{a\}] & \rightarrow & a\ [A,\{b\}]\ b \mid b\ [A,\{a\}]\ a \\
[A,\{a\}] & \rightarrow & c\ [S,\{a\}] \mid \varepsilon \\
[A,\{b\}] & \rightarrow & c\ [S,\{b\}] \mid \varepsilon
\end{array}
$$

И таблица разбора:

	a	b	c	#
[S,{#}]	a [A, {b}] b	b [A, {a}] a		
[S,{a}]	a [A, {b}] b	b [A, {a}] a		
[S,{b}]	a [A, {b}] b	b [A, {a}] a		
[A,{a}]	ε		c [S, {a}]	
[A,{b}]		ε	c [S, {b}]	

Полный $LL(1)$-разбор

Т.о. получили переписанную грамматику:

$$
\begin{array}{lcl}
[S,\{\#\}] & \rightarrow & a\ [A,\{b\}]\ b \mid b\ [A,\{a\}]\ a \\
[S,\{a\}] & \rightarrow & a\ [A,\{b\}]\ b \mid b\ [A,\{a\}]\ a \\
[S,\{a\}] & \rightarrow & a\ [A,\{b\}]\ b \mid b\ [A,\{a\}]\ a \\
[A,\{a\}] & \rightarrow & c\ [S,\{a\}] \mid \varepsilon \\
[A,\{b\}] & \rightarrow & c\ [S,\{b\}] \mid \varepsilon
\end{array}
$$

И таблица разбора:

	a	b	c	#
[S,{#}]	a [A, {b}] b	b [A, {a}] a		
[S,{a}]	a [A, {b}] b	b [A, {a}] a		
[S,{b}]	a [A, {b}] b	b [A, {a}] a		
[A,{a}]	ε		c [S, {a}]	
[A,{b}]		ε	c [S, {b}]	

Разрешение конфликтов в $LL(1)$

Если в грамматике есть конфликт, то можно

- Вернуться к поиску.
- Левая рекурсия – всегда ведет к конфликту. Устранить левую рекурсию
- Левая факторизация: $a*b+a*c = a*(b+c)$.
- Разрешение конфликтов

Разрешение конфликтов в $LL(1)$

Если в грамматике есть конфликт, то можно

- Вернуться к поиску.
- Левая рекурсия – всегда ведет к конфликту.
  Устранить левую рекурсию
- Левая факторизация: $a * b + a * c = a * (b + c)$.
- Разрешение конфликтов

Разрешение конфликтов в $LL(1)$

Если в грамматике есть конфликт, то можно

- Вернуться к поиску.
- Левая рекурсия – всегда ведет к конфликту. Устранить левую рекурсию
- Левая факторизация: $a * b + a * c = a * (b + c)$.
- Разрешение конфликтов

LL(k)-грамматики: $FIRST_k^G$

Есть ли смысл в предпросмотре более, чем на один символ?

Определение

Пусть $G = (V_N, V_T, P, S)$ — контекстно-свободная грамматика. Определим функцию

$$FIRST_k^G(\alpha) = \{w \in V_T^* \mid \text{либо } |w| < k \text{ и } \alpha \overset{*}{\underset{G}{\Rightarrow}} w,$$
$$\text{либо } |w| = k \text{ и } \alpha \overset{*}{\underset{G}{\Rightarrow}} wx$$
$$\text{для некоторой цепочки } x \in V_T^*\}.$$

Здесь $k \geq 0$ — целое, $\alpha \in (V_N \cup V_T)^*$.

LL(k)-грамматики: $FIRST_k^G$

Есть ли смысл в предпросмотре более, чем на один символ?

Определение

Пусть $G = (V_N, V_T, P, S)$ — контекстно-свободная грамматика. Определим функцию

$$FIRST_k^G(\alpha) = \{w \in V_T^* \mid \text{либо } |w| < k \text{ и } \alpha \overset{*}{\underset{G}{\Rightarrow}} w,$$
$$\text{либо } |w| = k \text{ и } \alpha \overset{*}{\underset{G}{\Rightarrow}} wx$$
$$\text{для некоторой цепочки } x \in V_T^*\}.$$

Здесь $k \geq 0$ — целое, $\alpha \in (V_N \cup V_T)^*$.

LL(k)-грамматики

Определение

Пусть $G = (V_N, V_T, P, S)$ — контекстно-свободная грамматика. Говорят, что G есть $LL(k)$-грамматика для некоторого фиксированного k, если для любых двух левосторонних выводов вида

$$1) S \underset{lm}{\overset{*}{\Rightarrow}} wA\alpha \underset{lm}{\Rightarrow} w\beta\alpha \underset{lm}{\overset{*}{\Rightarrow}} wx$$

$$2) S \underset{lm}{\overset{*}{\Rightarrow}} wA\alpha \underset{lm}{\Rightarrow} w\gamma\alpha \underset{lm}{\overset{*}{\Rightarrow}} wy$$

в которых $FIRST_k^G(x) = FIRST_k^G(y)$, имеет место равенство $\beta = \gamma$.

Определение

Говорят, что контекстно-свободная грамматика G есть LL-грамматика, если она $LL(k)$ для некоторого $k \geq 0$.

Сильные $LL(k)$-грамматики

Определение

Контекстно-свободная грамматика $G = (V_N, V_T, P, S)$ называется сильной $LL(k)$-грамматикой, если она удовлетворяет условию:

$$FIRST_k^G(\beta\ FOLLOW_k^G(A)) \cap FIRST_k^G(\gamma\ FOLLOW_k^G(A)) = \emptyset$$

для всех $A \in V_N, \beta, \gamma \in (V_N \cup V_T)^*$, таких, что существуют правила $A \to \beta, A \to \gamma \in P$, $\beta \neq \gamma$.

Лемма

Каждая $LL(1)$ грамматика является сильной.

Сильные $LL(k)$-грамматики

Определение

Контекстно-свободная грамматика $G = (V_N, V_T, P, S)$ называется сильной $LL(k)$-грамматикой, если она удовлетворяет условию:

$$FIRST_k^G(\beta \ FOLLOW_k^G(A)) \cap FIRST_k^G(\gamma \ FOLLOW_k^G(A)) = \emptyset$$

для всех $A \in V_N, \beta, \gamma \in (V_N \cup V_T)^*$, таких, что существуют правила $A \to \beta, A \to \gamma \in P$, $\beta \neq \gamma$.

Лемма

Каждая $LL(1)$ грамматика является сильной.

Свойства $LL(k)$-грамматик

- Существуют $LL(k+1)$ грамматики, которые не $LL(k)$:

$$S_s \rightarrow a^k b | a^k a$$

- Существуют $LL(k+1)$ языки, которые не $LL(k)$:

$$\begin{array}{lcl} S_s & \rightarrow & a\ S\ A \mid \varepsilon \\ A & \rightarrow & a^k\ b\ S \mid c \end{array}$$

- Существуют $LL(k)$ грамматики, которые не сильные $LL(k)$:

$$\begin{array}{lcl} S_s & \rightarrow & a\ A\ a\ a \mid b\ A\ b\ a \\ A & \rightarrow & b \mid \varepsilon \end{array}$$

Свойства $LL(k)$-грамматик

- Существуют $LL(k+1)$ грамматики, которые не $LL(k)$:

$$S_s \to a^k b | a^k a$$

- Существуют $LL(k+1)$ языки, которые не $LL(k)$:

$$\begin{array}{lcl} S_s & \to & a\ S\ A \mid \varepsilon \\ A & \to & a^k\ b\ S \mid c \end{array}$$

- Существуют $LL(k)$ грамматики, которые не сильные $LL(k)$:

$$\begin{array}{lcl} S_s & \to & a\ A\ a\ a \mid b\ A\ b\ a \\ A & \to & b \mid \varepsilon \end{array}$$

Свойства $LL(k)$-грамматик

- Существуют $LL(k+1)$ грамматики, которые не $LL(k)$:

$$S_s \to a^k b | a^k a$$

- Существуют $LL(k+1)$ языки, которые не $LL(k)$:

$$\begin{array}{lcl} S_s & \to & a\ S\ A \mid \varepsilon \\ A & \to & a^k\ b\ S \mid c \end{array}$$

- Существуют $LL(k)$ грамматики, которые не сильные $LL(k)$:

$$\begin{array}{lcl} S_s & \to & a\ A\ a\ a \mid b\ A\ b\ a \\ A & \to & b \mid \varepsilon \end{array}$$

Леворекурсивные и $LL(k)$-грамматики

Theorem

Если $G = (V_N, V_T, P, S)$ – контекстно-свободная грамматика, и G – леворекурсивна, то G — не $LL(k)$-грамматика ни при каком k.

Следствие

Два достаточных признака не LL-грамматик:

- *неоднозначность*
- *леворекурсивность*

Леворекурсивные и $LL(k)$-грамматики

Theorem

Если $G = (V_N, V_T, P, S)$ – контекстно-свободная грамматика, и G – леворекурсивна, то G — не $LL(k)$-грамматика ни при каком k.

Следствие

Два достаточных признака не LL-грамматик:

- *неоднозначность*
- *леворекурсивность*

Вопросы